



Study on the next data frontier: generative AI, regulatory compliance and international dimensions (No. 2024-020)

Final Report

Prepared by ICF, CARSA, VTT and
Timelex
October 2025

EUROPEAN COMMISSION

Directorate-General for Communications Networks, Content and Technology

Directorate G: Data

Unit G1 – Data Policy and Innovation

L-2920 Luxembourg

Contact: CNECT-G1@ec.europa.eu

Final Report

Study on the next data frontier: generative AI, regulatory compliance and international dimensions (No. 2024-020)



carsa



TIMELEX
RESEARCH & INNOVATION

LEGAL NOTICE

This document has been prepared for the European Commission however it reflects the views only of the authors, and the European Commission is not liable for any consequence stemming from the reuse of this publication. More information on the European Union is available on the Internet (<http://www.europa.eu>).

PDF

ISBN 978-92-68-33092-0

doi:10.2759/3847027

KK-01-25-142-EN-N

Luxembourg: Publications Office of the European Union, 2025

© European Union, 2025



The reuse policy of European Commission documents is implemented by the Commission Decision 2011/833/EU of 12 December 2011 on the reuse of Commission documents (OJ L 330, 14.12.2011, p. 39). Except otherwise noted, the reuse of this document is authorised under a Creative Commons Attribution 4.0 International (CC-BY 4.0) licence (<https://creativecommons.org/licenses/by/4.0/>). This means that reuse is allowed provided appropriate credit is given and any changes are indicated.

Table of Contents

Abstract.....	i
Executive Summary	iii
1 Introduction	1
1.1 Objectives of the study	1
1.2 Structure of the report.....	2
2 Task 1: Data needs for innovation and the role of synthetic data	4
2.1 Methodological approach: Task 1a – Data needs for GenAI companies in Europe.....	4
2.2 Findings: Task 1a – Data needs for GenAI companies in Europe	12
2.3 Methodological approach: Task 1b – Examination of the maturity and offering of data synthetisation solutions.....	62
2.4 Findings: Task 1b – Examination of the maturity and offering of data synthetisation solutions	67
3 Task 2: Using automated data reporting mechanisms for demonstrating regulatory compliance	89
3.1 Overall methodological approach	89
3.2 Conceptual framework.....	97
3.3 Agriculture	101
3.4 Automotive.....	128
3.5 Healthcare	152
3.6 Chemicals.....	170
3.7 Energy	191
4 Task 3 International aspects on the data economy.....	212
4.1 Methodological approach to Task 3	212
4.2 Key findings from the country and global analysis	214
4.3 Use case reporting.....	244
4.4 Task 3 policy recommendations	270
5 Conclusions and recommendations.....	279
5.1 Task 1: Data needs for innovation and the role of synthetic data	279
5.2 Task 2: Using automated data reporting mechanisms for demonstrating regulatory compliance	282
5.3 Task 3: International aspects on the data economy	287

List of acronyms

AI – Artificial Intelligence
AI Act – Artificial Intelligence Act
API – Application Programming Interface
CAP – Common Agricultural Policy
CEDS – Common European Data Spaces
CLP – Classification, Labelling and Packaging Regulation
CTR – Clinical Trials Regulation
DG CNECT – Directorate-General for Communications Networks, Content and Technology
EEA – European Economic Area
EED – Energy Efficiency Directive
ERP – Enterprise Resource Planning
ETS – Emissions Trading System
EU – European Union
FMIS – Farm Management Information System
GDPR – General Data Protection Regulation
GenAI – Generative Artificial Intelligence
ICT – Information and Communication Technology
IoT – Internet of Things
IVDR – In Vitro Diagnostic Medical Devices Regulation
KPI – Key Performance Indicator
MDR – Medical Devices Regulation
ML – Machine Learning
NLP – Natural Language Processing
REACH – Registration, Evaluation, Authorisation and Restriction of Chemicals
REMIT – Regulation on Wholesale Energy Market Integrity and Transparency
SME – Small and Medium-sized Enterprise

Abstract

Abstract EN

This study, commissioned by the Directorate-General for Communications Networks, Content and Technology (DG CNECT) of the European Commission, examines three dimensions of the European data economy: data needs for innovation, automated data reporting for regulatory compliance, and international aspects. The analysis draws on desk research, stakeholder interviews, surveys, and comparative studies, supported by academic and policy literature and industry insights.

Task 1 explores Europe's Generative AI ecosystem, which shows strong research output but lags behind the US and China in foundational model development and investment. Access to high-quality, multilingual, and domain-specific datasets is identified as a critical bottleneck. Legal and technical barriers, particularly for SMEs, persist. The report calls for interoperable data spaces, targeted public investment, and clearer regulatory guidance to foster competitiveness and inclusivity.

Task 2 assesses automated data reporting across agriculture, automotive, healthcare, chemicals, and energy. Adoption remains uneven, with manual methods still dominant among SMEs. Barriers include high costs, skill gaps, and fragmented regulatory frameworks. Recommendations focus on interoperable infrastructure, harmonised standards, and digital skills development, alongside public-private partnerships and regulatory flexibility to ensure inclusive transformation.

Task 3 maps and analyses 8 targeted third countries, offering a comparative assessment of data practices, identifying key trends and evaluating the extent of convergence with EU regulations. Task 3 also describes and analyses six use cases identified as potential opportunities for international data sharing. The EU is recommended to pursue a model that achieves a genuine balance between protecting data and fostering innovation. To achieve this, five key pillars should be in focus: Leverage regulatory gravity to export EU data space architecture; Pioneer new governance models to solve trust deficits; Transform friction into partnership through proactive engagement; Prioritise Privacy-Enhancing Technologies to future-proof security cooperation; Scale and structure R&D funding for AI and data spaces

Abstract FR

Cette étude, réalisée à la demande de la Direction générale des réseaux de communication, du contenu et des technologies (DG CNECT) de la Commission européenne, examine trois dimensions du marché européen des données : les besoins de données pour l'innovation, l'automatisation des rapports pour démontrer la conformité réglementaire, et les aspects internationaux. L'analyse s'appuie sur des recherches documentaires, des entretiens avec les parties prenantes, des enquêtes, des études comparatives et scientifiques, ainsi que sur des informations partagées par des représentants de différents secteurs industriels.

La tâche 1 étudie l'écosystème européen de l'IA générative. Ce secteur affiche une production de recherche solide, mais demeure en retrait par rapport aux États-Unis et à la Chine en ce qui concerne le développement de modèles fondamentaux et les investissements. L'accès à des bases de données de haute qualité, en plusieurs langues et spécifiques à certains domaines constitue un obstacle majeur. Des freins juridiques et techniques persistent, notamment pour les petites et moyennes entreprises. Le rapport recommande la création d'espaces de données interopérables, des investissements publics ciblés et des orientations réglementaires plus claires afin de renforcer la compétitivité et l'inclusivité.

La tâche 2 analyse la possibilité d'automatiser la production et la transmission des rapports de conformité réglementaire dans les secteurs de l'agriculture, de l'automobile, de la santé, de l'industrie chimique et de l'énergie. L'adoption des méthodes automatisées demeure inégale, les petites et moyennes entreprises recourant principalement à des procédés manuels. Les freins à cette adoption incluent des coûts élevés, des lacunes en compétences et des cadres réglementaires fragmentés. Les recommandations mettent l'accent sur le développement d'infrastructures interopérables, l'harmonisation des normes et le renforcement des compétences numériques, ainsi que sur la promotion de partenariats public-privé et une flexibilité réglementaire afin de garantir une transformation inclusive.

La tâche 3 répertorie et analyse huit pays tiers. Cette étude permet une comparaison des pratiques en matière de données, en identifiant les principales tendances et en évaluant le degré de convergence avec la réglementation de l'Union européenne. Le chapitre présente également six cas d'usage considérés comme des opportunités pour le partage international des données. Il est recommandé que l'UE adopte un modèle capable de concilier la protection des données avec la promotion de l'innovation. Pour atteindre cet objectif, cinq piliers sont mis en avant : exploiter la gravité réglementaire pour promouvoir l'architecture des espaces de données européens ; concevoir de nouveaux modèles de gouvernance afin de surmonter les déficits de confiance ; transformer les frictions en partenariats grâce à un engagement proactif ; accorder la priorité aux technologies renforçant la confidentialité pour anticiper les enjeux de sécurité ; et structurer ainsi qu'intensifier le financement de la recherche et du développement dans les domaines de l'intelligence artificielle et des espaces de données.

Abstract DE

Diese Studie, in Auftrag gegeben von der Generaldirektion Kommunikationsnetze, Inhalte und Technologien (GD CNECT) der Europäischen Kommission, untersucht drei Dimensionen der europäischen Datenwirtschaft: den Datenbedarf für Innovation, die automatisierte Datenberichterstattung zur Einhaltung regulatorischer Vorgaben sowie internationale Aspekte. Die Analyse stützt sich auf Literatur- und Dokumentenrecherche, Stakeholder-Interviews, Umfragen und vergleichende Studien, ergänzt durch wissenschaftliche und politische Fachliteratur sowie branchenspezifische Erkenntnisse.

Task 1 untersucht das europäische Ökosystem für Generative KI, das zwar eine starke Forschungsleistung aufweist, jedoch bei der Entwicklung und Finanzierung von Basismodellen hinter den USA und China zurückbleibt. Der Zugang zu hochwertigen, mehrsprachigen und domänenspezifischen Datensätzen wird als entscheidender Engpass identifiziert. Rechtliche und technische Hürden, insbesondere für KMU, bestehen fort. Der Bericht fordert interoperable *data spaces*, gezielte öffentliche Investitionen und klarere regulatorische Leitlinien, um Wettbewerbsfähigkeit und Inklusivität zu fördern.

Task 2 analysiert die automatisierte Datenberichterstattung in den Bereichen Landwirtschaft, Automobilindustrie, Gesundheitswesen, Chemie und Energie. Deren Einführung verläuft uneinheitlich; manuelle Verfahren dominieren weiterhin insbesondere bei KMU. Hemmnisse sind hohe Kosten, fehlende Digitalkompetenzen und fragmentierte regulatorische Rahmenbedingungen. Die Studie empfiehlt den Aufbau interoperabler Infrastrukturen, die Harmonisierung von Standards sowie die Entwicklung digitaler Kompetenzen, flankiert durch öffentlich-private Partnerschaften und regulatorische Flexibilität, zur Gewährleistung eines inklusiven Transformationsprozesses.

Task 3 analysiert acht ausgewählte Drittstaaten, durch eine vergleichende Bewertung der Datenpraktiken, wobei wichtige Trends identifiziert und der Grad der Annäherung an EU-Vorschriften bewertet werden. Zudem werden sechs Anwendungsfälle beschrieben und analysiert, welche Chancen für den internationalen Datenaustausch eröffnen. Der EU wird empfohlen, ein Modell zu verfolgen, das ein Gleichgewicht zwischen Datenschutz und

Innovationsförderung herstellt. Fünf zentrale Säulen sind dabei im Fokus: Nutzung des regulatorischen Einflusses zur Verbreitung der EU-Datenraumarchitektur; Entwicklung neuer Governance-Modelle zur Überwindung von Vertrauensdefiziten; Umwandlung von Reibungspunkten in Partnerschaft durch proaktives Engagement; Priorisierung von datenschutzfördernden Technologien zur zukunftssicheren Sicherheitskooperation; sowie Ausbau und Strukturierung der F&E-Finanzierung für KI und *data spaces*.

Executive Summary

Executive Summary EN

This study, commissioned by the Directorate-General for Communications Networks, Content and Technology (DG CNECT) of the European Commission, explores three critical dimensions of the evolving European data economy. Task 1 examines data needs for innovation and the role of synthetic data; Task 2 analyses the use of automated data reporting mechanisms for regulatory compliance across key sectors; and Task 3 addresses international aspects of the data economy. The methodology combines desk research, targeted surveys, stakeholder interviews, and comparative analysis, drawing on academic and policy literature, market studies, and direct consultation with industry and sectoral experts to ensure robust, evidence-based findings.

Task 1: Data needs for innovation and the role of synthetic data

Europe's Generative AI (GenAI) ecosystem is growing but remains behind the United States and China in foundational model development, investment, and market share. The study identifies Europe's strengths in research output and innovation, supported by a vibrant academic and start-up environment. However, most European GenAI companies focus on application layers—such as text, image, and speech tools—rather than building foundational models, which require significant compute resources and financial investment.

Access to high-quality, multilingual, and domain-specific datasets emerges as a critical bottleneck for innovation. Scientific literature and stakeholder feedback highlight that foundational models require vast amounts of unstructured data; while fine-tuning and specialised applications depend on smaller, highly curated datasets. Linguistic and cultural diversity is essential, as the dominance of English in training data risks underperformance in other languages and cultural underrepresentation.

Legal, technical, and practical barriers persist, particularly for SMEs. GDPR compliance, copyright restrictions, and fragmented enforcement across Member States increase complexity and costs. Technical challenges include the scarcity of annotated, multilingual datasets and risks of bias in training data. Stakeholders report difficulties in accessing domain-specific and multimodal datasets, compounded by high licensing fees and reluctance among data holders to share assets due to trust and competitive concerns.

The cost of acquiring, preparing, and managing data is a major operational burden. Licensing fees, annotation services, and compliance costs dominate expenditure for European GenAI firms. Public initiatives such as Common European Data Spaces and GAIA-X are widely anticipated to improve data accessibility, but awareness and participation remain limited, especially among SMEs.

Task 2: Automated data reporting for regulatory compliance

Automated data reporting tools are increasingly relevant for demonstrating regulatory compliance in agriculture, automotive, healthcare, chemicals, and energy. These sectors face complex and evolving EU regulations, including CAP Strategic Plans (agriculture), MDR/IVDR (healthcare), REACH and CLP (chemicals), REMIT and ETS (energy), and

sustainability directives (automotive). Despite the growing availability of digital tools, adoption remains uneven. Standard methods—such as spreadsheets and manual data entry—continue to dominate, particularly among SMEs, due to their low cost and flexibility, even though they are labour-intensive and error-prone.

Advanced tools, including digital logbooks, Farm Management Information Systems (FMIS), ERP systems, and specialised compliance software, are gaining traction among larger organisations. These solutions enable structured data capture, integration with geospatial and sensor data, and improved traceability. However, uptake is constrained by high licensing costs, technical complexity, and interoperability issues with national portals and legacy IT systems. Fully automated solutions, such as blockchain-based traceability platforms, remain experimental and limited to pilot projects.

Barriers to adoption include financial costs, digital skill gaps, and fragmented regulatory frameworks. Smaller operators face disproportionate challenges, risking exclusion from increasingly digitalised compliance regimes. Cost-benefit assessments indicate that advanced digital logbooks offer the best balance of efficiency and affordability, while FMIS and fully automated tools promise transformative benefits but remain viable primarily for large enterprises.

Investment in interoperable infrastructure, harmonised standards, and digital skills is essential. Partnerships and open data ecosystems should enable access to quality datasets, while flexible regulation and SME support ensure inclusive, scalable digital transformation.

Task 3: International aspects on the data economy

The analysis of international approaches to data governance across eight key economies identifies pathways for the EU to shape a more open, secure and rights-respecting global data ecosystem. Convergence with EU regulation varies markedly across the targeted countries. Japan's 2021 APPI and Data Free Flow with Trust concept achieve full alignment with EU standards, while Canada and South Korea hold adequacy decisions under GDPR but still face sectoral fragmentation. Australia and Brazil exhibit partial convergence through pragmatic, business-friendly models. The UK retains EU adequacy post-Brexit yet risks divergence under its Data (Use and Access) Act. The United States follows a market-driven, sector-specific patchwork, and China enforces a state-driven model emphasising localisation and security controls.

Four governance paradigms coexist: rights-driven (EU GDPR), market-driven (US innovation), state-driven (Chinese sovereignty) and consensus-driven (Japan's hybrid model). The EU's "third way" seeks to combine robust individual protections with innovation-friendly, federated data architectures.

Six illustrative use cases reveal practical opportunities and common barriers: 1) Digital Battery Passports link EU, North American and Asian manufacturers under the EU Battery Regulation; 2) Federated Learning for Rare Disease Genomics shows how collaborative AI can advance health research without sharing patient data; 3) Interoperable Safety Data Exchange for Autonomous Vehicles demonstrates the need for neutral "data trustees" to aggregate incident data; 4) Deforestation-Free Supply Chain Traceability under the EU Deforestation Regulation calls for inclusive, low-cost traceability platforms for smallholders; 5) Privacy-Enhancing Passenger Name Record (PNR) exchange offers a rights-respecting security model; and 6) Federated Digital Twins for Aerospace highlights both industrial promise and IP-protection challenges.

To capitalise on these findings, the EU should leverage its regulatory gravity – for example through sectoral adequacy agreements and Gaia-X/IDSA-compliant pilot projects – to export interoperable data spaces; pioneer novel governance models, such as data trustees and privacy-enhancing technologies, to address persistent trust deficits; forge practical partnerships (for instance a "Traceability for Sustainable Trade" initiative with

Brazil) to turn regulatory friction into cooperation; prioritise PETs in security cooperation and establish clear definitions for sensitive and highly sensitive non-personal data; and scale up R&D funding in generative AI, federated learning and cross-border data infrastructure. By aligning robust safeguards with technical and governance frameworks, the EU can solidify its role as the global standard-setter for a human-centric, innovation-driven data economy.

Executive Summary FR (résumé exécutif)

Cette étude, réalisée à la demande de la Direction générale des réseaux de communication, du contenu et des technologies (DG CNECT) de la Commission européenne, explore trois dimensions clés de l'économie européenne des données. La tâche 1 examine les besoins de données pour l'innovation ainsi que le rôle potentiel des données synthétiques. La tâche 2 analyse l'utilisation de méthodes automatisées pour démontrer la conformité réglementaire dans plusieurs secteurs stratégiques. La tâche 3 aborde les aspects internationaux de cette économie. La méthodologie combine des recherches documentaires, des enquêtes ciblées, des entretiens avec les parties prenantes et des analyses comparatives. Elle s'appuie également sur la littérature académique, des études de marché et des consultations directes avec des représentants de différents secteurs, afin de garantir des résultats solides et fondés sur des données probantes.

Tâche 1 : Besoins de données pour l'innovation et rôle des données synthétiques

L'écosystème européen de l'intelligence artificielle générative (GenAI) connaît une croissance soutenue, mais demeure en retrait par rapport aux États-Unis et à la Chine en matière de développement de modèles fondamentaux, d'investissements et de parts de marché. L'étude met en lumière les atouts européens en recherche et en innovation, portés par un environnement académique et entrepreneurial dynamique. Cependant, la majorité des entreprises européennes de GenAI se concentre sur les couches applicatives — telles que les outils de traitement du texte, de l'image et de la parole — plutôt que sur la conception de modèles fondamentaux, qui exigent des ressources de calcul considérables et des investissements financiers importants.

L'accès à des bases de données de haute qualité, multilingues et spécifiques à certains domaines constitue un obstacle majeur à l'innovation. La littérature scientifique et les retours des parties prenantes indiquent que les modèles fondamentaux nécessitent de vastes volumes de données non structurées, tandis que les applications spécialisées reposent sur des ensembles plus restreints et soigneusement organisés. La diversité linguistique et culturelle est essentielle, car la prédominance de l'anglais dans les données utilisées pour développer les modèles risque de mener à une faible performance dans d'autres langues et une représentation culturelle limitée.

Des obstacles juridiques, techniques et opérationnels persistent, en particulier pour les petites et moyennes entreprises. La conformité au RGPD, les restrictions liées au droit d'auteur et l'application fragmentée des règles entre États membres augmentent la complexité et les coûts. Les défis techniques incluent la rareté des bases de données annotées et multilingues, ainsi que les risques de biais dans les données utilisées pour développer les modèles. Les parties prenantes signalent des difficultés d'accès aux bases de données spécifiques à certains domaines. Ces difficultés sont aggravées par des frais de licence élevés et par la réticence des détenteurs de données à les partager, souvent liée à un manque de confiance et à la crainte de perdre un avantage concurrentiel.

Les coûts d'acquisition, de traitement et de gestion des données représentent une charge opérationnelle significative. Les frais de licence, les services d'annotation et les exigences liées à la conformité réglementaire constituent les principales dépenses pour les entreprises européennes spécialisées en GenAI. Les initiatives publiques telles que les espaces européens communs de données (*Common European Data Spaces*) et GAIA-X sont largement perçues comme des leviers pour améliorer l'accessibilité aux données. Toutefois, la sensibilisation à ces dispositifs et le niveau de participation restent limités, en particulier parmi les petites et moyennes entreprises.

Tâche 2 : Déclaration automatisée des données pour la conformité réglementaire

Les outils permettant l'automatisation de la production et de la transmission des rapports sont de plus en plus pertinents pour démontrer la conformité réglementaire dans les secteurs de l'agriculture, de l'automobile, de la santé, de l'industrie chimique et de l'énergie. Ces secteurs sont confrontés à une législation européenne complexe et en constante évolution, notamment les plans stratégiques de la PAC (agriculture), les règlements MDR/IVDR (santé), REACH et CLP (l'industrie chimique), REMIT et ETS (énergie), ainsi que les réglementations sur la durabilité (l'industrie automobile). Malgré la disponibilité croissante d'outils numériques, leur adoption reste inégale. Les méthodes traditionnelles — telles que les tableurs et la saisie manuelle — demeurent largement utilisées, en particulier par les petites et moyennes entreprises, en raison de leur faible coût et de leur flexibilité, bien qu'elles soient chronophages et sujettes aux erreurs.

Des outils avancés, tels que les registres digitaux (*digital logbooks*), les systèmes de gestion et d'information agricole (FMIS), les progiciels de gestion intégrée et les logiciels spécialisés de conformité, gagnent du terrain auprès des grandes organisations. Ces solutions permettent une saisie structurée des données, l'intégration avec des données géospatiales ou générées par des capteurs, ainsi qu'une meilleure traçabilité. Toutefois, leur adoption reste limitée en raison de coûts de licence élevés, de leur complexité technique et de problèmes d'interopérabilité avec les plateformes nationales et avec les systèmes informatiques existants. Quant aux solutions entièrement automatisées, comme les plateformes de traçabilité basées sur la blockchain, elles demeurent expérimentales et confinées à des projets pilotes.

Les obstacles à l'adoption incluent les coûts financiers, les lacunes en compétences numériques et la fragmentation des cadres réglementaires. Les opérateurs de petite taille sont confrontés à des défis disproportionnés et risquent de ne pas pouvoir s'adapter à la numérisation des formalités administratives et des obligations de déclaration. Les analyses coût-bénéfice indiquent que les registres digitaux avancés offrent le meilleur équilibre entre efficacité et accessibilité financière. Les FMIS et les outils entièrement automatisés promettent des bénéfices transformateurs mais restent principalement viables pour les grandes entreprises.

Les investissements dans des infrastructures interopérables, l'harmonisation des normes et le développement des compétences numériques sont essentiels. Des partenariats solides et des écosystèmes de données ouvertes devraient faciliter l'accès à des bases de données de qualité. Parallèlement, une réglementation plus flexible et un soutien accru aux petites et moyennes entreprises sont indispensables pour garantir une transformation numérique inclusive et durable.

Tâche 3 : Aspects internationaux de l'économie des données

L'analyse des approches internationales en matière de gouvernance des données dans huit pays tiers met en évidence des pistes permettant à l'UE de façonner un écosystème

mondial des données plus ouvert, sécurisé et respectueux des droits. Le degré de convergence avec la réglementation européenne varie considérablement selon les pays étudiés. Le Japon, avec sa loi APPI de 2021 et son concept de « flux de données libre et digne de confiance », est pleinement aligné sur les normes de l'UE. Le Canada et la Corée du Sud bénéficient de décisions d'adéquation au titre du RGPD, mais présentent encore des fragmentations sectorielles. L'Australie et le Brésil affichent une convergence partielle, fondée sur des modèles pragmatiques et favorables aux entreprises. Le Royaume-Uni conserve son statut d'adéquation post-Brexit, mais pourrait diverger avec l'adoption du Data (Use and Access) Act. Les États-Unis suivent un modèle sectoriel axé sur le marché, tandis que la Chine applique un modèle étatique centré sur la localisation et le contrôle de la sécurité.

Quatre paradigmes de gouvernance coexistent : le modèle axé sur les droits (RGPD de l'UE), le modèle axé sur le marché (innovation américaine), le modèle axé sur l'État (souveraineté chinoise) et le modèle axé sur le consensus (approche hybride japonaise). La « troisième voie » européenne vise à concilier une protection robuste des individus avec des architectures de données fédérées favorisant l'innovation.

Six cas d'usage emblématiques mettent en lumière des opportunités concrètes ainsi que des obstacles récurrents : les passeports numériques pour batteries (*Digital Battery Passports*), l'apprentissage fédéré appliqué à la génomique des maladies rares (*Federated Learning for Rare Disease Genomics*), le partage de données de sécurité pour les véhicules autonomes (*Interoperable Safety Data Exchange for Autonomous Vehicles*), la traçabilité des chaînes d'approvisionnement exemptes de déforestation (*Deforestation-Free Supply Chain Traceability under the EU Deforestation Regulation*), l'échange de données PNR respectueux de la vie privée (*Privacy-Enhancing Passenger Name Record*), et les jumeaux numériques fédérés dans le secteur aéronautique (*Federated Digital Twins for Aerospace*).

Pour capitaliser sur ces enseignements, l'Union européenne devrait mobiliser sa puissance réglementaire — notamment à travers des décisions d'adéquation et des projets pilotes alignés sur les cadres Gaia-X et IDSA — afin de promouvoir l'exportation d'espaces de données (*data spaces*) interopérables. Elle gagnerait à innover en matière de gouvernance, en explorant des modèles tels que les *data trustees* et les technologies de protection de la vie privée (*privacy-enhancing technologies*); à établir des partenariats concrets, comme une initiative de traçabilité pour un commerce durable avec le Brésil ; à accorder la priorité aux technologies de préservation de la vie privée dans la coopération en matière de sécurité ; et à renforcer le financement de la R&D dans des domaines clés tels que l'IA générative, l'apprentissage fédéré et les infrastructures transfrontalières de données. En alignant des garanties robustes avec des cadres techniques et de gouvernance ambitieux, l'UE peut consolider son rôle de référence mondiale pour une économie des données centrée sur l'humain et propulsée par l'innovation.

Executive Summary DE (Kurzfassung)

Diese Studie, in Auftrag gegeben von der Generaldirektion Kommunikationsnetze, Inhalte und Technologien (GD CNECT) der Europäischen Kommission, untersucht drei zentrale Dimensionen der sich entwickelnden europäischen Datenwirtschaft. Task 1 analysiert den Datenbedarf für Innovation und die Rolle synthetischer Daten; Task 2 befasst sich mit dem Einsatz automatisierter Datenberichterstattungsmechanismen zur regulatorischen Compliance in Schlüsselsektoren; Task 3 behandelt internationale Aspekte der Datenwirtschaft. Die Methodik kombiniert Literatur- und Dokumentenrecherche, gezielte Umfragen, Stakeholder-Interviews und vergleichende Analysen, gestützt auf

wissenschaftliche und politische Fachliteratur, Marktstudien sowie direkte Konsultationen mit Branchen- und Fachexperten, und gewährleistet so eine belastbare Evidenzbasis.

Task 1: Datenbedarf für Innovation und die Rolle synthetischer Daten

Das europäische Ökosystem für generative KI (GenAI) wächst, bleibt jedoch in Bezug auf die Entwicklung von Basismodellen, Investitionen und Marktanteile hinter den Vereinigten Staaten und China zurück. Die Studie sieht die Stärken Europas in den Bereichen Forschungsleistung und Innovation, die durch ein dynamisches akademisches Umfeld und eine lebendige Start-up-Szene unterstützt werden. Die meisten europäischen GenAI-Unternehmen konzentrieren sich jedoch auf Anwendungsebenen – wie Text-, Bild- und Sprachtools – anstelle der Entwicklung eigener Basismodelle, welche erhebliche Rechenressourcen und finanzielle Investitionen erfordern.

Der Zugang zu hochwertigen, mehrsprachigen und domänenspezifischen Datensätzen erweist sich als kritischer Engpass für Innovationen. Wissenschaftliche Literatur und Rückmeldungen von Interessengruppen zeigen, dass Basismodelle große Mengen unstrukturierter Daten erfordern, während Feinabstimmungen und spezialisierte Anwendungen auf kleinere, sorgfältig kuratierte Datensätze angewiesen sind. Sprachliche und kulturelle Vielfalt ist von entscheidender Bedeutung, da die Dominanz der englischen Sprache in Trainingsdaten zu einer unterdurchschnittlichen Leistung in anderen Sprachen sowie zu kultureller Unterrepräsentation führen kann.

Rechtliche, technische und praktische Hindernisse bestehen weiterhin, insbesondere für KMU. Die Einhaltung der DSGVO, urheberrechtliche Beschränkungen und eine uneinheitliche Durchsetzung in den Mitgliedstaaten erhöhen Komplexität und Kosten. Zu den technischen Herausforderungen zählen der Mangel an annotierten, mehrsprachigen Datensätzen und das Risiko von Verzerrungen in den Trainingsdaten. Stakeholder berichten von Schwierigkeiten beim Zugang zu domänenspezifischen und multimodalen Datensätzen, die durch hohe Lizenzgebühren und die Zurückhaltung der Dateninhaber aufgrund von Vertrauens- und Wettbewerbsbedenken noch verstärkt werden.

Die Kosten für die Beschaffung, Aufbereitung und Verwaltung von Daten stellen eine erhebliche operative Belastung dar. Lizenzgebühren, Annotationsdienste und Compliance-Kosten machen den größten Teil der Ausgaben europäischer GenAI-Unternehmen aus. Öffentliche Initiativen wie Common European Data Spaces und GAIA-X sollen die Datenzugänglichkeit verbessern, doch deren Bekanntheit und Teilnahme sind vorerst begrenzt, insbesondere bei KMU.

Task 2: Automatisierte Datenberichterstattung zur Unterstützung regulatorischer Compliance

Automatisierte Tools zur Datenberichterstattung werden immer wichtiger, um die Einhaltung gesetzlicher Vorschriften in den Bereichen Landwirtschaft, Automobilindustrie, Gesundheitswesen, Chemie und Energie nachzuweisen. Diese Branchen sind mit komplexen und sich ständig entwickelnden EU-Vorschriften konfrontiert, darunter die strategischen Pläne der GAP (Landwirtschaft), MDR/IVDR (Gesundheitswesen), REACH und CLP (Chemikalien), REMIT und ETS (Energie) sowie Nachhaltigkeitsrichtlinien (Automobilindustrie). Trotz der zunehmenden Verfügbarkeit digitaler Tools werden diese noch uneinheitlich eingesetzt. Herkömmliche Methoden wie Tabellenkalkulationen und manuelle Dateneingabe dominieren aufgrund ihrer geringen Kosten und Flexibilität weiterhin, insbesondere bei KMU, obschon diese arbeitsintensiv und fehleranfällig sind.

Fortschrittliche Tools wie digitale Logbücher, Farm Management Information Systems (FMIS), ERP-Systeme und spezialisierte Compliance-Software gewinnen bei größeren Organisationen zunehmend an Bedeutung. Diese Lösungen ermöglichen eine strukturierte Datenerfassung, die Integration mit Geodaten und Sensordaten, sowie eine verbesserte Rückverfolgbarkeit. Die Akzeptanz dieser Systeme wird jedoch durch hohe Lizenzkosten, technische Komplexität und Interoperabilitätsprobleme mit nationalen

Portalen und älteren IT-Systemen eingeschränkt. Vollautomatisierte Lösungen, wie Blockchain-basierte Plattformen zur Rückverfolgung, befinden sich noch im Versuchsstadium und sind auf Pilotprojekte beschränkt.

Zu den Hindernissen für eine flächendeckende Einführung zählen der Kostenaufwand, fehlende Digitalkompetenzen und uneinheitliche Rechtsrahmen. Kleinere Marktteilnehmer stehen vor unverhältnismäßig großen Herausforderungen und laufen Gefahr, von den zunehmend digitalisierten Compliance-Systemen ausgeschlossen zu werden. Kosten-Nutzen-Analysen zeigen, dass moderne digitale Logbücher das beste Verhältnis zwischen Effizienz und Erschwinglichkeit bieten, während FMIS und vollautomatisierte Tools zwar transformative Vorteile versprechen, aber vor allem für große Unternehmen rentabel sind.

Investitionen in interoperable Infrastruktur, harmonisierte Standards und digitale Kompetenzen sind unerlässlich. Partnerschaften und offene Datenökosysteme sollten den Zugang zu hochwertigen Datensätzen ermöglichen, während flexible Regulierung und die Unterstützung von KMU eine inklusive und skalierbare digitale Transformation gewährleisten können.

Task 3: Internationale Aspekte der Datenwirtschaft

Die Analyse internationaler Ansätze zur Datenverwaltung in acht wichtigen Volkswirtschaften zeigt Wege auf, wie die EU ein offeneres, sichereres und rechtskonformes globales Datenökosystem gestalten kann. Die Konformität dieser Ansätze mit den Vorschriften in der EU variiert zwischen den untersuchten Ländern erheblich. Japans APPI 2021 und das Konzept *Data Free Flow with Trust* entsprechen vollständig den EU-Standards, während für Kanada und Südkorea zwar DSGVO-Angemessenheitsbeschlüsse vorliegen, jedoch eine sektorale Fragmentierung besteht. Australien und Brasilien weisen mit pragmatischen, wirtschaftsfreundlichen Modellen eine teilweise Konvergenz auf. Das Vereinigte Königreich bewahrt nach dem Brexit DSGVO-Angemessenheit, riskiert jedoch Abweichungen aufgrund seines Gesetzes über die Nutzung und den Zugang zu Daten (*Data (Use and Access) Act*). In den Vereinigten Staaten gibt es einen marktorientierten, sektorspezifischen Flickenteppich, und China setzt auf ein staatlich gelenktes Modell, das den Schwerpunkt auf Lokalisierung und Sicherheitskontrollen legt.

Es existieren vier Governance-Paradigmen nebeneinander: rechtsorientiert (EU-DSGVO), marktorientiert (US-Innovation), staatlich orientiert (Chinesische Souveränität) und konsensorientiert (hybrides Modell Japans). Der „dritte Weg“ der EU zielt darauf ab, einen robusten Schutz des Einzelnen mit innovationsfreundlichen, förderierten Datenarchitekturen zu verbinden.

Sechs Fallstudien zeigen Anwendungspotenziale und häufige Hindernisse auf: 1) *Digital Battery Passports* verbinden Hersteller aus der EU, Nordamerika und Asien im Rahmen der EU-Batterieverordnung; 2) die Fallstudie *Federated Learning for Rare Disease Genomics* zeigt, wie kollaborative KI die Gesundheitsforschung voranbringen kann, ohne Patientendaten weiterzugeben; 3) *Interoperable Safety Data Exchange for Autonomous Vehicles* verdeutlicht die Notwendigkeit neutraler „Datentreuhänder“ zur Aggregation von Unfalldaten; 4) *Deforestation-Free Supply Chain Traceability under the EU Deforestation Regulation* fordert integrative, kostengünstige Rückverfolgbarkeitsplattformen für Kleinbauern; 5) *Privacy-Enhancing Passenger Name Record (PNR)* bietet ein rechtskonformes Sicherheitsmodell; und 6) *Federated Digital Twins for Aerospace* zeigen sowohl das industrielle Potenzial als auch die Herausforderungen beim Schutz des geistigen Eigentums auf.

Um die Erkenntnisse der Studie für sich zu nutzen, sollte die EU ihren regulatorischen Einfluss geltend machen – etwa durch sektorale Angemessenheitsvereinbarungen und Gaia-X/IDSA-konforme Pilotprojekte – um interoperable *data spaces* zu exportieren; neue Governance-Modelle wie *data trustees* und Technologien zum Schutz der Privatsphäre (*privacy-enhancing technologies* – PETs) einführen, um anhaltende Vertrauensdefizite zu

beheben; praktische Partnerschaften schließen (etwa eine Initiative zur Rückverfolgbarkeit für nachhaltigen Handel mit Brasilien), um regulatorische Reibungspunkte durch Kooperation zu ersetzen; PETs in der Sicherheitszusammenarbeit priorisieren und klare Definitionen für sensible und hochsensible nicht-personenbezogene Daten festlegen; und die F&E-Finanzierung in den Bereichen generative KI, föderiertes Lernen und grenzüberschreitende Dateninfrastruktur aufstocken. Durch die Abstimmung robuster Schutzmaßnahmen mit technischen und Governance-Rahmenwerken kann die EU ihre Rolle als globaler Standardsetzer für eine innovationsgetriebene Datenwirtschaft, bei der der Mensch im Mittelpunkt steht, festigen.

1 Introduction

1.1 Objectives of the study

This Final Report presents the results of the study “The Next Data Frontier: Generative AI, Regulatory Compliance and International Dimensions,” commissioned by DG CNECT. The study provides a comprehensive, evidence-based analysis of recent developments in the European data economy, with the aim of supporting informed policy and strategic decision-making. A key characteristic of this study is that it is structured around three core and self-standing objectives, as described below.

1. **Understanding the data needs for European based GenAI companies and exploring the role of synthetic data.** Within this overarching objective, this study aims at:
 - a. Assessing the evolving **landscape of data needs** for generative AI (GenAI) development in Europe, with particular attention to the challenges of accessing high-quality, diverse, and domain-specific datasets.
 - b. Examining **barriers to data access and use**, including legal, technical, financial, and linguistic factors, and analyse the implications of linguistic and cultural biases in available data.
 - c. **Evaluating the role of synthetic data** as a potential solution to data scarcity, privacy, and compliance challenges, considering both opportunities and limitations.
 - d. Identify **potential policy priorities** to improve data availability, quality, and legal certainty for European AI developers, supporting innovation while ensuring compliance with EU regulations.
2. **Assessing the current and potential use of automated data reporting for regulatory compliance**, which entails the following activities:
 - a. Investigate the **potential of automated data reporting mechanisms** to enhance regulatory compliance across key sectors of the EU economy.
 - b. **Map current practices and sectoral readiness** for automation in compliance-related data collection, verification, and reporting.
 - c. Assess the **challenges of adopting advanced technologies** for streamlining compliance, reducing administrative burden, and improving data quality.
 - d. Identify **barriers to adoption and integration**, including technical, organizational, and regulatory factors, and propose solutions to facilitate effective and scalable implementation.
 - e. Sketch potential sector-specific policy priorities and recommendations to enhance the uptake
3. **Examining third-country practices and identify opportunities for global data collaboration** while maintaining compliance with EU laws, which entails the following activities:
 - a. Map and analyse measures that promote data access and measures that protect sensitive and highly sensitive data
 - b. Assess the barriers to cross-border data flows and identify opportunities

- c. Propose actionable recommendations to address barriers and exploit opportunities for international data cooperation

By addressing these objectives, this Final Report delivers a thorough and actionable resource for policymakers, industry, and stakeholders seeking to navigate and shape the future of data-driven innovation in Europe. Each core area was addressed through a tailored methodological approach—drawing on desk research, targeted surveys, stakeholder consultations, and comparative analysis—ensuring that the findings and recommendations are robust, relevant, and aligned with the diverse challenges and opportunities across the European data economy.

1.2 Structure of the report

This Final Report is organised to provide a clear, self-contained account of the study's objectives, methods, and findings. Each of the study's core objectives has been integrated into a dedicated task, with each task addressed in a specific, self-standing section of the report. This structure ensures that readers can engage with each topic independently, while also understanding how the components contribute to the overall analysis. The remainder of this report is organised as follows:

- **Section 2** dives into the first objectives, namely '**Data needs for innovation and the role of synthetic data**', presented as **Task 1**. More specifically, Task 1 has been divided into two parallel sub-tasks: *Task 1a Data needs for GenAI companies in Europe* and *Task 1b Examination of the maturity and offering of data synthetisation solutions*. This conceptual division is reflected in the overall structure of this Section 2:
 - **Section 2.1.** presents the overall methodological approach to Task 1a, including details on the literature review, targeted surveys and expert interviews
 - **Section 2.2.** presents the findings for Task 1a, divided into sub-topics covering all the relevant areas described in the overarching objectives.
 - **Section 2.3.** provides information on the methodological framework for Task 1b by describing the approach to the mapping exercise, the design of company factsheets and survey results.
 - **Section 2.4.** presents the findings of Task 1b by sketching a comprehensive landscape of the current offer of data synthetisation solutions
- **Section 3** presents the underlying methodological approach and the findings for Task 2 addressing the current and potential use of automated data reporting for regulatory compliance. It is organised into the following sub-sections:
 - **Section 3.1.** presents the methodological framework applied to each sectorial analysis, providing specific information for each sectorial fiche.
 - **Section 3.2.** lays down the conceptual framework which defines the objectives and scope of the sectorial analysis.
 - **From Section 3.3 to Section 3.7,** each specific sectorial analysis (Agriculture, Automotive, Healthcare, Chemical, and Energy) is provided.
- **Section 4** presents the methodological approach and the findings related to Task 3, which focuses on the **international dimension** of data sharing and collaboration.

- **Section 4.1.** briefly presents the methodological approach to Task 3, including for example the mapping of measures in the eight targeted countries, the creation of policy fiches, country- and global analyses and use case analyses.
- **Section 4.2** offers a comparative assessment of data practices across eight targeted third countries, identifying key trends and evaluating the extent of convergence with EU regulations. The country-specific findings are consolidated into a global analysis, setting out the main trends shaping data practices today and in the future.
- **Section 4.3** provides the descriptions of six selected use cases representing opportunities for international data sharing.
- **Section 4.4** presents conclusions and policy recommendations related to Task 3.

The report is accompanied by the following annexes:

- Annex 1: Task 1a – Reference list
- Annex 2: Task 1b – Mapping of synthetic solutions output
- Annex 3: Task 2 – Legal analysis
- Annex 4: Task 3 – Policy fiches
- Annex 5: Task 3 – Mapping of definitions related to sensitive and highly sensitive data
- Annex 6: List of interviewees – Task 3 use cases

2 Task 1: Data needs for innovation and the role of synthetic data

Task 1 addresses a twofold objective of the study: to identify the **data-related challenges and opportunities** faced by generative AI (GenAI) companies in Europe and to examine the **maturity and offering of synthetic data solutions** as potential enablers of innovation. The study aims to understand which types of data are needed by GenAI developers, the difficulties they encounter in accessing or using them, and the costs and legal complexities involved. At the same time, the analysis explores how synthetic data can mitigate these barriers by providing alternative, privacy-preserving, and cost-effective options.

For operational reasons, Task 1 has been split into two complementary parts:

- **Task 1a** focuses on the **data needs of EEA-based GenAI companies**, investigating their market context, data requirements, accessibility challenges, and the costs associated with data acquisition. This includes desk research, market analysis, and targeted consultations with companies and stakeholders.
- **Task 1b** examines the **landscape of synthetic data solutions in Europe**, mapping providers, assessing their maturity, and identifying opportunities and challenges linked to their use, including privacy safeguards and future policy considerations.

The subsequent sections are structured accordingly:

- **Section 2.1** presents the overall **methodology for Task 1a** which includes desk research, market analysis and consultation activities.
- **Section 2.2** provides the **findings from Task 1a**, covering the European GenAI market, data needs and requirements, accessibility challenges, costs of data acquisition, data spaces and alternative access models, and potential solutions to support innovation in the EEA.
- **Section 2.3** outlines the **methodological approach for Task 1b**, including the mapping of synthetic data providers and the consultation activities carried out through surveys and interviews.
- **Section 2.4** presents the **findings from Task 1b**, providing an overview of the conceptual framework for mapping providers, the current landscape of synthetic data companies in Europe, illustrative company factsheets, and an analysis of challenges and future outlook.

2.1 Methodological approach: Task 1a – Data needs for GenAI companies in Europe

The methodological approach for Task 1a combines three main components. First, **desk research** was undertaken, with a strong focus on academic and policy literature, to identify and assess the data needs of generative AI companies in Europe. This was complemented by a **market analysis** relying on publicly available sources, including financial statements, annual reports and existing market studies, in order to derive indicative estimates of market trends and indicators. In addition, **consultation activities** were conducted through a survey of GenAI companies and a series of targeted interviews with experts and industry representatives, aimed at validating and enriching the evidence collected through the desk and market research.

2.1.1 Desk research and literature review

As part of Task 1.1a, the literature review was designed to provide a **comprehensive and balanced understanding** of the current landscape, challenges, and potential pathways for generative AI (GenAI) in the European Economic Area (EEA). Building upon the initial scoping and diagnostic, it served as a **foundational evidence base** for subsequent analytical tasks.

The review focused on literature published between **January 2020 and May 2025**, a timeframe chosen to capture the rapid acceleration in AI research and policy debates following the COVID-19 pandemic and the increased political salience of GenAI in Europe.

The search strategy combined **academic, institutional, and industry sources**. Academic literature was identified through databases such as Google Scholar, arXiv, Springer, Elsevier and, for healthcare applications, PubMed. **Grey literature** was accessed through repositories including the OECD iLibrary and the EU Open Data Portal. Particular attention was paid to **reports and working papers** produced by European and international institutions such as the European Commission, the European Data Protection Supervisor (EDPS), the European Data Protection Board (EDPB), the OECD, and the Joint Research Centre (JRC). Complementary perspectives were drawn from **think tanks and consultancies** including Bruegel, IFRI, the European Round Table for Industry (ERT), and the Implement Consulting Group. Finally, industry reports and technical documentation from major players (e.g. Microsoft, Google, Nvidia, OpenAI, Hugging Face) as well as global observatories (e.g. the Stanford AI Index 2025 and the OECD.AI Public Observatory) provided practical insights into investment trends, infrastructure developments, and dataset availability.

Source selection followed a structured, multi-stage process. An initial broad pool was created using targeted keywords such as *“Generative AI,” “Large Language Models,” “data needs,” “data access,” “Europe,”* and *“regulatory challenges.”* This pool was then screened for relevance, giving priority to highly cited academic work and reports from recognised institutions. Strict adherence to the defined publication window was maintained, and the process continued until **thematic saturation** was reached. In total, **more than 60 publications** were retained, ensuring a balanced mix of peer-reviewed studies, policy documents, consultancy analyses, and industry white papers.

To structure the analysis, sources were organised into **six thematic categories**:

1. The overall landscape and trends in GenAI in Europe.
2. The data needs and requirements of GenAI companies in the EEA.
3. Data sources and accessibility challenges.
4. The cost of data acquisition, processing, and compliance.
5. Data spaces and alternative access models.
6. Solutions and policy recommendations to support GenAI innovation in Europe.

Coverage across these categories was generally satisfactory. Topics relating to **trends** and **data needs** were particularly well documented, supported by a broad set of institutional and academic contributions. However, some gaps remain. Evidence on **data accessibility and costs** remains partial, with little quantitative information available at the enterprise or sector level and most concrete examples coming from large US-based actors. Similarly, while there is an extensive conceptual literature on **data spaces** and on **potential solutions and policy instruments**, relatively little empirical work exists on their actual effectiveness, uptake, or user satisfaction. These limitations were addressed, as much as possible, by complementary market analysis and by the consultation activities described under Section 2.1.3.

The distribution of sources across categories and the main evidence gaps are summarised in Table 2.1. The full list of references is provided in **Annex 1**.

Table 1. Overview of literature review sources Task 1a

N.	Category	N.	Coverage and gaps
1	<i>Generative AI in Europe: landscape and trends</i>	27	Strong coverage drawing on OECD, JRC, AppliedAI, and Stanford AI Index. Only minor updates expected as new industry data emerge.
2	<i>Data needs and requirements of EEA-based GenAI companies</i>	23	Well documented through academic and institutional sources. Limited SME-specific insights, but overall requirements are well captured.
3	<i>Data sources and accessibility challenges</i>	31	Broadly covered, including GDPR, copyright, and multilingual data issues. Some quantitative sector-level estimates remain limited, though trends are clear.
4	<i>Cost of data acquisition, processing and compliance</i>	28	Coverage strengthened by recent evidence on licensing costs and compliance. Enterprise-level detail is less visible due to confidentiality, but overall patterns are well established.
5	<i>Data spaces and alternative access models</i>	11	Literature provides detailed conceptual frameworks and policy initiatives. Early evidence on practical effectiveness is emerging and will continue to develop.
6	<i>Solutions and policy recommendations for GenAI in Europe</i>	22	Comprehensive coverage of proposed solutions, policy measures, and pilots. Only limited empirical evaluation of outcomes to date, which is expected given their recent launch.

The literature review constitutes the **backbone of the study findings** presented in Section 2.2. The wide range of sources examined enabled a comprehensive overview of the data needs and related challenges faced by GenAI companies, with particular attention to the European perspective wherever possible. However, as discussed in Section 2.2.1, despite significant growth in recent years, the **European GenAI market still lags behind the US and China**. Consequently, much of the existing literature—especially scientific publications—focuses predominantly on these markets, where more consolidated evidence and data are available. This limitation is also acknowledged within literature itself, with several sources **underscoring the need for dedicated attention to underrepresented languages, markets, and sectors** in order to foster more inclusive AI development. This includes research on improving performance in low-resource languages and addressing biases and cultural underrepresentation in training data^{1,2}. In

¹ Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big?. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

² High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe: https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

this context, the OECD.AI Public Observatory could play a valuable role by analyzing open-source AI training datasets to identify and monitor linguistic imbalances³.

2.1.2 Market analysis and overview

The market analysis was designed to **complement and expand upon** the results of the desk research, thereby providing an evidence-based context for the subsequent findings (Sections 2.2.1 and 2.2.4). Its objectives were threefold:

- To provide overall context for the analysis of data needs and for the policy recommendations forming the backbone of Task 1a.
- To ensure an EU-centred perspective, by detailing concrete facts and figures on the European generative AI (GenAI) market, while also situating these within the wider global landscape to enable meaningful comparisons.
- To generate quantitative and monetary estimates regarding the size of the GenAI market and, in particular, the costs associated with data acquisition and preparation.

At the same time, a number of limitations must be acknowledged. **The vast majority of GenAI companies in Europe are privately held**, which means that they do not systematically publish financial statements, revenue breakdowns, or market disclosures. This lack of transparency significantly narrows the availability of reliable monetary and quantitative information, particularly when compared to more established segments of the digital economy (e.g. cloud services or platform-based industries) where reporting standards are more consolidated.

Table 2. Overview of sources for Market Analysis

Type of sources	Examples	Purpose
Market studies and landscape reports	• JRC Generative AI Outlook Report (2025) • AppliedAI Generative AI in the European Startup Landscape (2024); • EIT Digital Reports	Provide structural overview of GenAI ecosystem, funding, market positioning and EU vs global trends
Annual reports of listed GenAI companies	• LightOn Annual Report 2024 (Euronext Growth Paris); • Teneo.AI Annual Report 2024; • Unith Ltd Annual Report 2024 (ASX & Frankfurt)	Provide rare financial benchmarks on revenues, costs, and strategic positioning of EU-based GenAI players
Specialised news outlets and sector blogs	Financial Times on Mistral AI and Europe's competitiveness; TechCrunch and Sifted on GenAI fundraising; VentureBeat and Forbes on costs of annotation; industry press releases from ScaleAI,	Provide real-time insights into competitive dynamics, funding flows, annotation cost benchmarks, and emerging EU actors

³ Ibidem.

	Labelbox, SuperAnnotate, and Roboflow	

2.1.3 Consultation activities

Consultation activities were designed to gain direct input from relevant stakeholders on key aspects of data needs and challenges for European GenAI companies. The consultation activities were organised into two activities:

1. **A targeted online survey** (from May to July 2025) addressed to about 90 GenAI companies in Europe, including start-ups, large companies and covering all type of AI development.
2. **An interview programme** (August to September 2025) addressed to academic and sectorial experts in the field of generative AI and related topics.

Further detail on the methodological approach is described in the following sub-sections.

2.1.3.1 Targeted survey

This activity underpinned the design, implementation, and analysis of a targeted survey that complemented and expanded the findings from the literature review and gathered structured feedback directly from industry stakeholders to inform evidence-based policymaking.

Survey outreach, structure and response rate.

Respondents were selected through a targeted and referral-based approach, focusing on **companies actively developing or deploying generative AI systems** across the EEA. A total of 91 GenAI (or related) companies have been identified. Special care was taken to include a diverse sample in terms of company size, domain focus (e.g. legal tech, health, creative industries), AI modality (text, image, audio, video), and technical maturity. As a result, the final respondent pool included early-stage startups, scaleups, and more mature AI firms with different operational models and data needs.

The survey was structured around six thematic areas reflecting both operational and strategic data-related issues faced by generative AI companies. Question design followed an iterative process informed by desk research, stakeholder mapping, and expert consultation, ensuring relevance to the study objectives and clarity in wording. A combination of multiple-choice items, matrix rating scales, and open-ended questions was used to generate both quantitative indicators and qualitative insights. An overview of the key topics covered is presented in the table below.

Table 3. Overview of sources for Market Analysis

Survey section	Key objectives
Section 0: Company Profile	Understand the organisational background of respondents, including their role, company size, primary market sector, and the types of generative AI products and modalities they work with. This information establishes the context for interpreting other responses.
Section 1: Data Needs and Requirements	Identify the volume, type, and modality of data required by generative AI firms. Capture specific needs related to annotation, language diversity, and domain specificity to understand how these influence model training and development.
Section 2: Data Sources and Accessibility	Map where companies currently source their data and identify key accessibility issues, including legal, technical, financial, and linguistic barriers. This informs policy actions to improve data availability.

Challenges	
Section 3: Cost of Data Acquisition	Gather insights into the cost structures associated with acquiring high-quality datasets, including licensing, curation, and annotation. Explore how costs vary by company size and use case, and whether they pose a barrier to scaling innovation.
Section 4: Data Spaces and Alternative Models	Assess awareness, participation, and satisfaction with European data spaces or other access mechanisms. Explore expectations regarding their functionality, governance, and ability to support cross-border and cross-sector AI development.
Section 5: Solutions and Policy Recommendations	Collect stakeholder input on desired regulatory changes, public investment priorities, and support mechanisms to enable more effective data access and use for generative AI development.
Section 6: Future Outlook	Understand the evolving needs and anticipated challenges of generative AI firms over the next 3–5 years, including emerging data types, infrastructure requirements, and perceived risks or opportunities in the data ecosystem.

Survey implementation

The survey was launched on 20 May 2025 via the online platform Forsta. It was initially distributed to 95 targeted contacts, with recipients encouraged to further disseminate the link within their networks to broaden outreach. A general reminder was issued on 3 June 2025, followed by the identification and inclusion of an additional 50 relevant contacts. An additional reminder was sent on 16.06.2025 followed by another final reminder on 4.07.2025. The survey was eventually closed on 15.07.2025 collecting a total 5 completed replies.

Disclaimer: low response rate and limited representativeness of the survey

From a methodological perspective, it is important to **acknowledge the limitations** of the targeted survey carried out under this study. The survey achieved only five fully completed responses, significantly below expectations, which limits its representativeness and reduces its overall scientific added value. While the survey was carefully designed, distributed to over 140 targeted contacts, and accompanied by reminders, the final response rate remained low.

Several factors can explain this outcome. First, the **European generative AI ecosystem remains relatively young and fragmented**, with only a limited number of active companies at different stages of maturity. Unlike more established sectors, there are few strong stakeholder associations or industry groupings capable of amplifying survey outreach or mobilising participation on a larger scale.

Second, **the subject matter of the survey touches on highly sensitive and commercially confidential topics**, especially regarding data access strategies, cost structures, and financial expenditures. Many companies are reluctant to disclose such information, even in anonymised form, due to competitive pressures and the perception that data-related costs or sourcing practices constitute a core part of their business model.

Third, **many GenAI firms are small, early-stage startups or research-driven spin-offs**. These organisations often face severe resource constraints and may lack the time or dedicated staff to respond to detailed surveys, particularly when the questions require technical or managerial input. In addition, survey fatigue is a known issue in the fast-moving AI ecosystem, with companies regularly approached by researchers, consultancies, and policymakers for input.

Despite these limitations, **the survey still provides interesting indicative insights into**

the data needs and challenges of generative AI companies in Europe. The responses did cover all of the key thematic areas defined in the survey design, including data needs, data sources and accessibility, costs, data spaces, and policy perspectives. While the sample size prevents representative conclusions, the evidence collected still provides useful insights into each of the main research topics and contributes to triangulating findings from the literature review and interviews.

In this sense, while the findings cannot be considered representative of the entire GenAI ecosystem, they nonetheless complement the literature review and stakeholder interviews by offering direct perspectives from industry actors, helping to illustrate both the diversity of experiences and the recurring themes that shape the European GenAI data landscape.

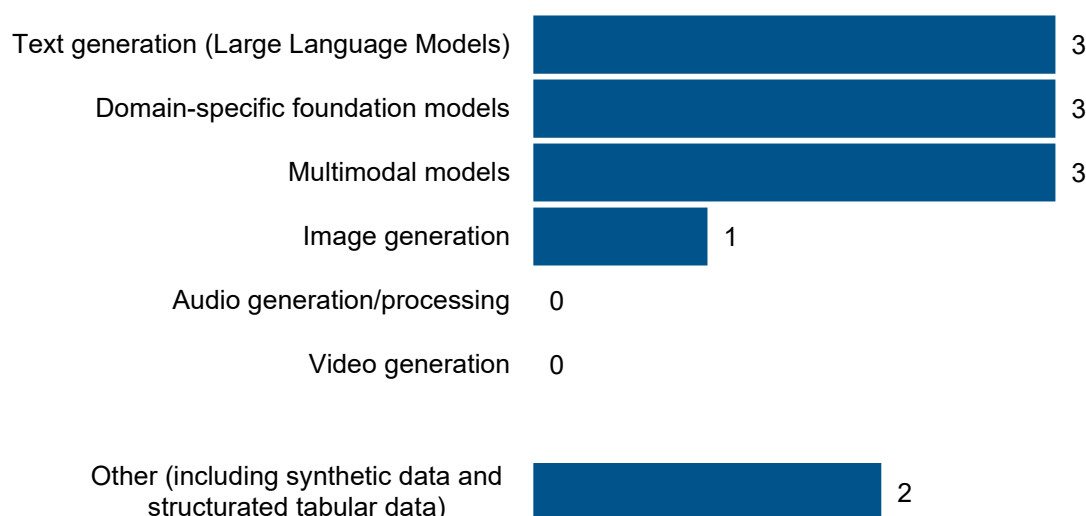
Survey sample

The representativeness of the survey sample – albeit limited by the overall number of responses received – provides balanced input from different type of GenAI

- Company sizes ranged from micro-enterprises (1–9 employees) to SMEs (10–49 employees) and larger firms (49–250 employees).
- Respondent profiles included 2 C-level executives, 1 data scientist/engineer, and 2 who preferred not to disclose their role.
- All participating companies were founded between 2018 and 2022.
- Geographically, respondents were based in Germany (2), France (1), and Spain (1), though all are likely active across the EU.
- Business domains covered all relevant sectors identified in Q4 (Healthcare/Financial services/Legal services/Manufacturing/Creative /Public sector).

In terms of the type of AI solutions, the sample display a balance representation of the key models, as shown in the figure below:

Figure 1. Survey Q3 AI solution types - What type of generative AI solutions does your company develop?



All key model types are represented in the sample, with particular emphasis on **text generation (LLMs)** and **domain-specific foundation models**, which the literature

identifies as the most common approaches among European GenAI firms. Some modalities, such as audio or video generation, are not represented, but these are not yet widespread in the European landscape. Interestingly, the sample also includes work on **structured tabular data**, highlighting that non-traditional generative approaches are also represented in the survey sample.

2.1.4 Targeted interviews

To complement the inputs collected through the survey and to gain an additional perspective, four experts' interviews were organized between August and September 2025.

Interviews focused on collecting inputs from academic experts and industry representative to enrich and complement the findings from the literature review and the survey. In particular, interviews focused on gaining qualitative insights on the overarching challenges faced by GenAI companies in Europe as well as sketching long-term outlook and potential policy approaches. An overview of the topics discussed is provided below:

1. **Data as evidence & human judgment:** Data alone are not evidence; interpretation, context, and human judgment are essential to turn data into , support public infrastructure, and build skills pipelines
2. **Access to qualitative and sensitive data:** Qualitative data such as text, images, and conversations are most valuable for generative AI, but are difficult to access and curate, especially for under-represented groups.
3. **Synthetic data & privacy risks:** Synthetic data can help but are not a complete solution to privacy risks, as AI systems still rely on real user data, raising consent and fairness issues.
4. **GDPR and innovation:** GDPR is seen as a long-term benefit for trustworthy innovation if applied with case-specific governance, but overly restrictive interpretations can block useful applications; there is little evidence that GDPR actually hinders innovation.
5. **Data governance & policy recommendations:** Effective data governance should balance openness with privacy, using tiered access and privacy-preserving analytics, and policies should support data curation, public infrastructure, and skills development.
6. **Fair value & user compensation:** There is a need to consider fair compensation for users whose data fuel AI and digital services, ensuring benefits are shared more equitably.

Interviews questionnaires were tailored according to the interviewee profile of the respondents. Key inputs from interviews have been anonymised and integrated in the findings presented under Section 2.2.

Table 4. Overview of the interview programme (Task 1a)

Type of stakeholders	Topics discussed in the interview
Academic expert	Data access challenges; regulatory impacts; EU data initiatives; strategies for data scarcity; future trends and risks in generative AI.
Academic expert	Data quality and coverage; multimodal and multilingual data; European specialization in AI; regulatory frameworks; methodologies for robustness and compliance.
Academic expert	Data needs assessment; synthetic data applications; privacy and utility considerations; market narratives; policy and regulatory context.
Academic expert	Data privacy regulations and considerations in relation to generative AI and synthetic data.

2.2 Findings: Task 1a – Data needs for GenAI companies in Europe

This section presents the findings for Task 1a ‘**Data needs for GenAI companies in Europe**’. It is based on evidence and insights from desk research, survey and interviews already triangulated and presented in an integrated fashion. This following subsection are structured as follows:

- Section 2.2.1 presents an overview of GenAI in Europe providing both market and technological context
- Section 2.2.2 builds a conceptual framework for data needs, highlighting the different components and characteristics of data used for GenAI.
- Section 2.2.3 dives into the challenges explains their nature and implications
- Section 2.2.4 focuses on cost of data acquisition and processing, using real market data as proxies.
- Section 2.2.5 and 2.2.6 concludes by highlighting the role of Data Spaces and sketching potential policy recommendations.

2.2.1 The GenAI market in Europe: landscape and trends

2.2.1.1 Definition and overview

According to a recent JRC publication⁴, Generative Artificial Intelligence (GenAI) refers to a class of artificial intelligence that focuses on the creation of new content, including text, images, video, music, or code. Unlike traditional AI models that perform predictive or analytical tasks, GenAI learns patterns from large datasets and generates original outputs that closely resemble human-created content. As defined by the Joint Research Centre (JRC), GenAI is considered a transformative technology with potential applications across diverse sectors including healthcare, education, and scientific research. Other definitions highlight GenAI’s distinguishing feature—its ability to synthesize novel content. According to Feuerriegel et al. (2024)⁵, and the University of Turin Law School & Nexa Center for Internet & Society (2025)⁶, GenAI systems are capable of generating outputs that can be indistinguishable from those created by humans.

The global generative AI landscape is dominated by China and the United States. According to the JRC Generative AI Outlook Report (2025), China accounts for 60% of global GenAI “players”, followed by the United States with 12% and the European Union

⁴ ABENDROTH DIAS, K., ARIAS CABARCOS, P., BACCO, F.M., BASSANI, E., BERTOLETTI, A. et al., Generative AI Outlook Report - Exploring the Intersection of Technology, Society and Policy, NAVAJAS CAWOOD, E., VESPE, M., KOTSEV, A. and VAN BAVEL, R. (editors), Publications Office of the European Union, Luxembourg, 2025, https://data.europa.eu/doi/10.2760/1109679_JRC142598.

⁵ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. Bus Inf Syst Eng 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

⁶ University of Turin Law School & Nexa Center for Internet & Society. (2025, May 12). Development of Generative Artificial Intelligence from a Copyright Perspective. European Union Intellectual Property Office. https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf

with 7%. However, this headline figure requires interpretation. The term “player” in this context refers to any organisation engaged in GenAI-related activities, including research, innovation (such as patenting), and business or investment activities. This broad definition encompasses universities, research institutes, start-ups, and companies at various stages of maturity—not just large-scale commercial actors.

Despite China’s dominance in the number of GenAI players, the United States far outpaces China and the EU in terms of actual investment and commercial impact. In 2024, the US invested \$109.1 billion in AI, compared to \$9.3 billion by China and \$4.5 billion by the UK. From 2019 to 2023, US-based GenAI companies raised \$23.8 billion, accounting for 75% of global GenAI investments, with OpenAI alone raising \$12.3 billion. Across the full GenAI value chain, US companies secured over 80–85% of global funding

In terms of model development, 73% of foundational AI models originate from the US, 15% from China, and a small share from Europe (EIT Digital, 2024). In 2024, the US produced 40 notable AI models, China produced 15, and Europe produced 3 (Stanford Institute for Human-Centered AI, 2025). Patent activity shows similar trends: in 2023, US inventors accounted for 34% of global AI patent applications, Chinese inventors for 22%, and EU27 inventors for 11% (EIT Digital, 2024). According to Stanford data, the EU’s share of global AI patent applications was 2.8% in 2023, compared to China’s 69.7% and the US’s 14.2% (Stanford Institute for Human-Centered AI, 2025).

On the other hand, the European Union holds a strong position in generative AI research and innovation⁷. It ranks second globally in the number of academic publications related to GenAI, reflecting the strength of its research institutions and academic ecosystem (JRC, 2025, Figure 2). According to findings by Renda et al. (2025), Europe surpasses the United States in scientific publications on GenAI, placing second worldwide behind China.

2.2.1.2 Market features and key players

The generative AI startup ecosystem in Europe is expanding but remains relatively modest compared to global leaders. According to the AppliedAI Institute⁸ (2024), approximately 10.6%—equivalent to 669 companies—of the 6,300 AI startups identified across the EU are focused on generative AI. These startups are geographically concentrated, with the highest shares in Germany (19.9%), France (17.5%), the Netherlands (10.9%), and Sweden (8.2%). Notable GenAI companies in Europe include France’s Mistral AI, Germany’s Aleph Alpha, and the UK’s Synthesia. Mistral, valued at \$6 billion, has received significant venture funding and is developing proprietary large language models (LLMs)⁹. Aleph Alpha is focused on secure, explainable AI, while Synthesia specializes in synthetic video production. Finland’s Silo AI is also active through its ‘Poro’ project, aimed at developing LLMs for low-resource languages. Most GenAI startups in Europe concentrate on application layers such as text, image, and speech tools (55%), while 30% develop infrastructure and tooling. Only about 15% focus

⁷ ABENDROTH DIAS, K., ARIAS CABARCOS, P., BACCO, F.M., BASSANI, E., BERTOLETTI, A. et al., Generative AI Outlook Report - Exploring the Intersection of Technology, Society and Policy, NAVAJAS CAWOOD, E., VESPE, M., KOTSEV, A. and VAN BAVEL, R. (editors), Publications Office of the European Union, Luxembourg, 2025, <https://data.europa.eu/doi/10.2760/1109679, JRC142598>.

⁸ AppliedAI, Generative AI in the European Startup Landscape (2024);

⁹ Financial Times, Europe’s Mistral benefits from search for artificial intelligence alternatives: <https://www.ft.com/content/65f79839-d637-48a7-a0f2-3fab8952b315> (consulted in August 2025)

on building foundational models like LLMs, which require considerable compute and financial resources¹⁰.

According to the JRC, the European market for Generalist Conversational AI (GCAI)¹¹—a key application area of GenAI—is marked by high concentration. A small number of platforms dominate user engagement across the EU. Data analysed by the JRC revealed that top five players, predominantly from the US and China, dominate the market, accounting for 82% of monthly active users (MAU). ChatGPT by OpenAI clearly dominates the EU market by concentrating more than 72% of the market share amongst the amongst the top 5 apps terms of MAU and downloads. The key competitors are interface-only solution¹²s like ChatOn (US) and DeepSeek (Chinese). Interestingly, none of the tech giants (e.g. Google and Microsoft) is included in the top 5 GenAI companies in the EU in terms of MAU and downloads of applications.

Offer of GenAI services

While still highly concentrated, the **web-based market**¹³ for GenAI services in the EU shows a slightly different distribution compared to mobile applications. According to the JRC (2025), the **top five websites** account for **91.5% of total market share**. Among them, **ChatGPT by OpenAI** holds the dominant position with **72%**, followed by **Google Gemini (15%)** and **Microsoft Copilot (4.5%)**. The JRC attributes the stronger performance of Google and Microsoft in the online segment to the **integration of their AI tools into widely used platforms**, such as operating systems and web services, as well as their established **brand recognition**.

As outlined above, no European GenAI companies currently match the scale or market influence of global leaders, most of which are based in the United States. Mistral, based in France and considered the most prominent EU-based GenAI company, still lags significantly behind its American counterparts. According to a recent analysis by Menlo Ventures, Mistral holds a market share of approximately 5% in the French enterprise AI market—less than half the share held by Google or Meta, and well behind OpenAI.

Demand for GenAI services

Demand for generative AI in Europe remains relatively limited, with adoption concentrated primarily among large enterprises. According to the JRC, key sectors currently driving GenAI uptake include finance, education, healthcare, creative industries, and cybersecurity—areas that have already undergone significant digital transformation. These sectors possess the necessary infrastructure and resources to integrate GenAI technologies. In contrast, small and medium-sized enterprises (SMEs)—which represent

¹⁰ AppliedAI, Generative AI in the European Startup Landscape (2024);

¹¹ ABENDROTH DIAS, K., ARIAS CABARCOS, P., BACCO, F.M., BASSANI, E., BERTOLETTI, A. et al., Generative AI Outlook Report - Exploring the Intersection of Technology, Society and Policy, NAVAJAS CAWOOD, E., VESPE, M., KOTSEV, A. and VAN BAVEL, R. (editors), Publications Office of the European Union, Luxembourg, 2025, <https://data.europa.eu/doi/10.2760/1109679>, JRC142598.

¹² Interface-only” solutions refer to GenAI applications or platforms that provide a user-facing interface (such as a chatbot or assistant app) but do not develop or operate their own proprietary foundational language models.

¹³ The “web-based market” for GenAI services refers to user access and engagement with GenAI tools via web browsers—i.e., through websites—rather than through dedicated mobile applications or desktop software. In the JRC’s analysis, market share by “websites” is measured by the proportion of unique visitors or usage on the top GenAI-related web domains (such as chat.openai.com for ChatGPT, gemini.google.com for Gemini, etc.) within the EU.

99% of all businesses in the EU—often lack the digital maturity required for meaningful adoption.

While some SMEs are beginning to explore GenAI applications, the overall uptake remains low. Support networks like the **European Digital Innovation Hubs (EDIHs)** play a crucial role in enabling this transition. According to the JRC analysis, current GenAI use cases are highly specific, often limited to improving business processes or enhancing product features. EDIH-led assessments indicate that SMEs typically begin experimenting with AI only after achieving a moderate level of digital transformation. Key enablers such as data management capabilities, digital infrastructure, and skilled personnel are often underdeveloped. Without these foundational elements, widespread adoption of GenAI across Europe's SME sector remains a significant challenge.

2.2.1.3 Ownership structure and investments

The GenAI ecosystem in Europe is predominantly privately held, with most companies funded through venture capital and strategic investors¹⁴. **Mistral**, one of the most prominent European players, is currently valued at **\$6 billion** and remains privately owned. The company has publicly stated that it is not pursuing an IPO at this stage (Financial Times, 2025). Other key European GenAI firms—including **Aleph Alpha** (Germany), **DeepL** (Germany), **Poolside** (France), —are similarly privately held and have received backing from a mix of venture capital firms and corporate investors. In terms of investment trends, **European GenAI startups raised €2.6 billion in 2023**, up from **€700 million in 2022**. However, more than **70% of this capital** was concentrated among a handful of firms—particularly **Mistral**, **Aleph Alpha**, and **Synthesia**. Most companies remain at **seed or Series A stages**, indicating a relatively early phase of development with limited scalability. This investment profile also reflects a broader absence of **publicly listed GenAI companies in Europe**. As of 2025, Lighton – a Paris based start-up – is the only 'pure' European GenAI company publicly listed.

A key event in the European GenAI ecosystem was the acquisition of Finnish artificial intelligence startup Silo AI by Advanced Micro Devices (AMD) in August 2024. The all-cash transaction was valued at approximately \$665 million. This strategic move aims to bolster AMD's AI capabilities and enhance its competitiveness against industry leader Nvidia.

Ownership dynamics also reflect broader international engagement. According to the JRC (2025), 12% of EU-based GenAI companies are foreign-owned, with 49% controlled by U.S. entities, followed by Japan (13%), the United Kingdom (11%), Switzerland (7%), and China (5%). These patterns highlight the increasing role of non-EU investors and operators in shaping the regional GenAI landscape.

2.2.2 Conceptual framework underpinning data needs

2.2.2.1 Overview of the different types of GenAI technical landscape and modalities

The GenAI technical landscape in Europe, and globally, is characterized by a diverse range of solutions and approaches. These solutions span multiple modalities:

- **Text generation:** This is a core application, including LLMs for tasks like drafting reports, summarizing articles, coding assistance, customer service chatbots, and

¹⁴Financial Times, Has Europe's great hope for AI missed its moment?
<https://www.ft.com/content/fa8bad75-dc55-47d9-9eb4-79ac94e54d82> (access in August 2025).

automated narrative generation^{15 16 17}. Examples include ChatGPT, Google Bard, and Claude¹⁸.

- **Image generation:** Solutions range from text-to-image generators like DALL-E and Stable Diffusion to photo editing and digital asset creation for gaming^{19, 20, 21}.
- **Audio generation:** This includes speech synthesis, music generation (e.g., MusicLM), and AI voice agents for customer service.^{22, 23, 24}
- **Video generation:** This involves creating synthetic videos (e.g., OpenAI's Sora, Google's Veo), often for simulation or content creation.^{25, 26}
- **Multimodal solutions:** These models can process and generate content across various data types, such as text, images, audio, and video, enabling complex applications like visual question answering or integrated industrial AI assistants [28, 46]. Examples include GPT-4o and Gemini.^{27, 28}

¹⁵ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%20C%20Implementing%20and%20Developing%20Chapter.pdf

¹⁶ Fonteneau, F., & Anvar, E. (2025, May 2). Exploring the safe and responsible use of generative AI across the data lifecycle: Early insights from the OECD. OECD. https://unece.org/sites/default/files/2025-05/GenAI2025_S2_OECD_FONTENEAU_D.pdf

¹⁷ Tammes-Peters, H. (2025). Guidelines on using generative artificial intelligence to support and assist work in public administration. <https://urn.fi/URN:ISBN:978-952-367-854-5>

¹⁸ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

¹⁹ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

²⁰ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. Bus Inf Syst Eng 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

²¹ Le Picard, H. (2025, April). European Startups and Generative AI: Overcoming Big Tech Dominance. Ifri. https://www.ifri.org/sites/default/files/2025-05/ifri_le_picard_european_startups_genai_2025.pdf

²² Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward

²³ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. Bus Inf Syst Eng 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

²⁴ Le Picard, H. (2025, April). European Startups and Generative AI: Overcoming Big Tech Dominance. Ifri. https://www.ifri.org/sites/default/files/2025-05/ifri_le_picard_european_startups_genai_2025.pdf

²⁵ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

²⁶ Manduchi, L., Pandey, K., Meister, C., Bamler, R., Cotterell, R., Däubener, S., ... & Fortuin, V. (2024). On the challenges and opportunities in generative ai. arXiv preprint arXiv:2403.00025.

²⁷ Stanford Institute for Human-Centered AI. (2025, April 7). The AI Index Report 2025. <https://hai.stanford.edu/ai-index/2025-ai-index-report>

²⁸ Ruokonen, M. (2025). APPLYING AI IN KEY EUROPEAN INDUSTRIES.

Key model types and training approaches underpinning GenAI are diverse and continuously evolving:

- **Large Language Models (LLMs)** are a prominent subset of GenAI, specifically designed to process and generate human-like text. They are built on deep learning architectures, most notably **transformers**, and are trained on massive text corpora.^{29 30 31} Examples include GPT-3, GPT-4, BERT, and Llama.^{32 33 34} LLMs are typically **pre-trained** on vast quantities of primarily unlabeled data using self-supervised learning, where models learn from unannotated data by predicting parts of the input (e.g., masked language modeling in BERT).^{35 36}
- **Diffusion models** are a newer class of generative models, particularly effective for generating high-quality images and are also being explored for audio and cross-modal tasks.^{37 38} Examples include Stable Diffusion and DALL-E.^{39 40}
- **Multimodal models** extend generative capabilities by processing and generating content across multiple data types simultaneously, such as text, images, audio,

²⁹ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

³⁰ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

³¹ Myers, D., Mohawesh, R., Chellaboina, V.I. et al. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Comput* 27, 1–26 (2024). <https://doi.org/10.1007/s10586-023-04203-7>

³² Microsoft Corporation. (2024, April). GDPR and Generative AI: A Guide for the Public Sector. <https://wwps.microsoft.com/wp-content/uploads/2024/04/GDPR-and-Generative-AI-A-Guide-for-the-Public-Sector-FINAL.pdf>

³³ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

³⁴ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. *arXiv preprint arXiv:2403.04190*.

³⁵ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

³⁶ Myers, D., Mohawesh, R., Chellaboina, V.I. et al. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Comput* 27, 1–26 (2024). <https://doi.org/10.1007/s10586-023-04203-7>

³⁷ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. *Bus Inf Syst Eng* 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

³⁸ Manduchi, L., Pandey, K., Meister, C., Bamler, R., Cotterell, R., Däubener, S., ... & Fortuin, V. (2024). On the challenges and opportunities in generative ai. *arXiv preprint arXiv:2403.00025*.

³⁹ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. *Bus Inf Syst Eng* 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

⁴⁰ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. *TechCrunch*. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

and video.^{41 42 43}. Examples include GPT-4 (accepts text and image inputs to generate text outputs) and Google Gemini (processes text, pictures, video, and audio).^{44 45}.

- **Foundation models** are large, general-purpose AI systems trained on broad data at scale that can be adapted or fine-tuned for a wide array of downstream tasks^{46, 47, 48, 49}. They exhibit properties such as **emergence** (implicitly induced behaviors, like GPT models creating calendar entries without explicit training) and **homogenization** (where a wide range of systems can be powered by a single, consolidated model).^{50 51}
- **Fine-tuning** is a common and efficient approach where a pre-trained foundation model is further trained on smaller, task-specific, or domain-specific data to specialize it.^{52 53 54}. This contrasts with **training from scratch**, which is resource-

⁴¹ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. Bus Inf Syst Eng 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

⁴² Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

⁴³ Manduchi, L., Pandey, K., Meister, C., Bamler, R., Cotterell, R., Däubener, S., ... & Fortuin, V. (2024). On the challenges and opportunities in generative ai. arXiv preprint arXiv:2403.00025.

⁴⁴ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. Bus Inf Syst Eng 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

⁴⁵ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

⁴⁶ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

⁴⁷ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

⁴⁸ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258

⁴⁹ Microsoft Corporation. (2024, April). GDPR and Generative AI: A Guide for the Public Sector. <https://wwps.microsoft.com/wp-content/uploads/2024/04/GDPR-and-Generative-AI-A-Guide-for-the-Public-Sector-FINAL.pdf>

⁵⁰ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. Bus Inf Syst Eng 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

⁵¹ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

⁵² European Data Protection Supervisor. (2024, June 3). Generative AI and the EUDPR: First EDPS Orientations for ensuring data protection compliance when using Generative AI systems. https://www.edps.europa.eu/system/files/2024-05/24-05-29_genai_orientations_en_0.pdf

⁵³ Chen, J., & Mueller, J. (2024). Automated data curation for robust language model fine-tuning. arXiv preprint arXiv:2403.12776.

⁵⁴ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

intensive and less common.^{55, 56}. Techniques like **Reinforcement Learning from Human Feedback (RLHF)** are used to align model outputs with human preferences and values.^{57, 58}.

- Other key techniques include **Retrieval-Augmented Generation (RAG)**, which combines data retrieval and generation to improve accuracy and relevance by fetching information from external sources.^{59, 60} and **adversarial training**, where two different deep-learning models are pitted against each other to improve output quality.⁶¹

Most European GenAI companies primarily **fine-tune existing foundation models** rather than training them from scratch,^{62, 63}. This approach is favored due to the high computational costs and extensive data requirements associated with training large models from the ground up.^{64, 65}. Fine-tuning allows adaptation to specific tasks, industries,

⁵⁵ European Data Protection Supervisor. (2024, June 3). Generative AI and the EUDPR: First EDPS Orientations for ensuring data protection compliance when using Generative AI systems. https://www.edps.europa.eu/system/files/2024-05/24-05-29_genai_orientations_en_0.pdf

⁵⁶ Chen, J., & Mueller, J. (2024). Automated data curation for robust language model fine-tuning. arXiv preprint arXiv:2403.12776.

⁵⁷ European Data Protection Supervisor. (2024, June 3). Generative AI and the EUDPR: First EDPS Orientations for ensuring data protection compliance when using Generative AI systems. https://www.edps.europa.eu/system/files/2024-05/24-05-29_genai_orientations_en_0.pdf

⁵⁸ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. Bus Inf Syst Eng 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

⁵⁹ Tammes-Peters, H. (2025). Guidelines on using generative artificial intelligence to support and assist work in public administration. <https://urn.fi/URN:ISBN:978-952-367-854-5>

⁶⁰ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

⁶¹ Deloitte. (2024, October 7). Technology Spotlight — Accounting for the Development of Generative AI Software Products. <https://dart.deloitte.com/USDART/home/publications/deloitte/industry/technology/accounting-generative-ai-software-products>

⁶² Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

⁶³ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

⁶⁴ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

⁶⁵ Deloitte. (2024, October 7). Technology Spotlight — Accounting for the Development of Generative AI Software Products. <https://dart.deloitte.com/USDART/home/publications/deloitte/industry/technology/accounting-generative-ai-software-products>

or linguistic contexts with less data and compute.^{66,67} For example, BloombergGPT is fine-tuned on financial language documents, and Codestral on programming languages.⁶⁸

Linguistic capabilities are a key focus in Europe, given its diverse languages. Models are increasingly optimized for multiple European languages, and initiatives aim to increase the availability of European language data to uphold linguistic diversity.^{69,70} However, the dominance of English in training data (57.3% of open-source datasets on Hugging Face are in English) means that performance often lags for non-English languages, and low-resource languages remain underrepresented.^{71,72} Non-English prompt engineering faces hurdles due to data scarcity, linguistic diversity, and bias/cultural nuance.⁷³ European efforts like Mistral AI in France and Poro (Silo AI) in Finland are developing LLMs tailored for European languages and values.^{74,75} For instance, models like Poro and the later Viking family, which employ cross-lingual training techniques to overcome data scarcity, have achieved state-of-the-art performance for low-resource languages such as Finnish on the FIN-bench benchmark, substantially outperforming general-purpose multilingual

⁶⁶ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

⁶⁷ Implement Consulting Group. (2024, September 30). The economic opportunity of AI in the EU.
<https://cms.implementconsultinggroup.com/media/uploads/articles/2024/The-economic-opportunity-of-generative-AI-in-the-EU/The-economic-opportunity-of-AI-in-the-EU.pdf>

⁶⁸ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

⁶⁹ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

⁷⁰ Google. (2023, November 14). Google's EU AI Opportunity Agenda.
https://static.googleusercontent.com/media/publicpolicy.google/en//resources/eu_ai_opportunity_agenda_en.pdf

⁷¹ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

⁷² Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

⁷³ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

⁷⁴ Le Picard, H. (2025, April). European Startups and Generative AI: Overcoming Big Tech Dominance. Ifri. https://www.ifri.org/sites/default/files/2025-05/ifri_le_picard_european_startups_genai_2025.pdf

⁷⁵ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

and monolingual models in that language⁷⁶. According to one interviewee, the amount of data available for European languages is much lower than for English, presenting both a challenge and an opportunity for specialization. Indeed, that large language models can pivot to other languages due to conceptual representations in latent space, but less-resourced languages require data augmentation and stronger methodologies. Another expert noted that while major languages like French, German, and Italian benefit from reliable automatic translation, smaller minority languages such as Greek and Romanian face significant challenges, with most data infrastructures lacking integrated translation capabilities.

The technical landscape is also marked by a significant **dependency on US-based infrastructure and cloud providers** for compute resources.^{77, 78} U.S. tech giants control 68% of the global cloud market, and Nvidia holds 90% of the GPU market.⁷⁹ While there are European efforts to build local capacity, such as OVHcloud and AI Factories, the overall landscape is influenced by the vertical integration strategies of global tech giants.^{80, 81} The trend toward **open-source models** is gaining traction, with Meta's Llama and DeepSeek's r1 offering alternatives to closed models, potentially playing to EU strengths in collaborative development.^{82, 83} There is also a growing use of **synthetic data** and advanced techniques (e.g., Mixture of Experts, multi-layer attention) to reduce compute and data requirements, as exemplified by DeepSeek.⁸⁴ The EU is also investing in regulatory frameworks (AI Act) and standardisation to ensure trustworthy, ethical, and inclusive GenAI development.⁸⁵

Albeit not fully representative and based on a limited sample, survey results seem to confirm the literature review findings. Four out of five firms stated **that they rely on a combination of model adaptation and limited training from scratch**, while only one firm reported training models exclusively from scratch. Most of respondents also reported a moderate to heavy dependence on external foundation model providers, somehow

⁷⁶ Luukkonen, Risto, et al. "Poro 34b and the blessing of multilinguality." *arXiv preprint arXiv:2404.01856* (2024).

⁷⁷ Le Picard, H. (2025, April). European Startups and Generative AI: Overcoming Big Tech Dominance. Ifri. https://www.ifri.org/sites/default/files/2025-05/ifri_le_picard_european_startups_genai_2025.pdf

⁷⁸ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

⁷⁹ Le Picard, H. (2025, April). European Startups and Generative AI: Overcoming Big Tech Dominance. Ifri. https://www.ifri.org/sites/default/files/2025-05/ifri_le_picard_european_startups_genai_2025.pdf

⁸⁰ Le Picard, H. (2025, April). European Startups and Generative AI: Overcoming Big Tech Dominance. Ifri. https://www.ifri.org/sites/default/files/2025-05/ifri_le_picard_european_startups_genai_2025.pdf

⁸¹ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

⁸² EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

⁸³ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

⁸⁴ Ibidem

⁸⁵ Ibidem

confirming literature review findings which highlighted a dependency on US-based infrastructures.

2.2.2.2 Type of data required

The development of GenAI models necessitates access to **high-quality, extensive, and relevant datasets**.^{86, 87}. Foundation models, which form the core of many GenAI systems, are trained on massive amounts of often unstructured and unlabelled data, including text, images, video, and audio.^{88, 89, 90}.

For **fine-tuning approaches** and specialized applications, which are the most common in the EU GenAI landscape, the data requirements become more precise. This often involves smaller, **domain-specific datasets** that are highly curated and often **annotated**.^{91, 92}. For instance, medical AI models require anonymized patient records, imaging data, and genomics information.⁹³ In industrial settings, real-time sensor readings, maintenance logs, and production metrics are needed.⁹⁴ For official statistics, data needs to be structured according to shared codifications and semantic standards (like SDMX), enriched with clear definitions, contextual metadata, and documentation of data transformations⁹⁵. Annotation levels vary, from basic labelling to complex, context-aware tagging, especially for supervised learning and bias mitigation.⁹⁶

The **importance of multilingual and cultural data** is paramount, particularly in the European context. Given the EU's 24 official languages and diverse cultural landscape, there is a strong need for models to be trained and fine-tuned on data that reflects this

⁸⁶ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

⁸⁷ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

⁸⁸ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

⁸⁹ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

⁹⁰ European Data Protection Supervisor. (2024, June 3). Generative AI and the EUDPR: First EDPS Orientations for ensuring data protection compliance when using Generative AI systems. https://www.edps.europa.eu/system/files/2024-05/24-05-29_genai_orientations_en_0.pdf

⁹¹ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

⁹² Ruokonen, M. (2025). APPLYING AI IN KEY EUROPEAN INDUSTRIES

⁹³ Ibidem

⁹⁴ Ibidem

⁹⁵ Fonteneau, F., & Anvar, E. (2025, May 2). Exploring the safe and responsible use of generative AI across the data lifecycle: Early insights from the OECD. OECD. https://unece.org/sites/default/files/2025-05/GenAI2025_S2_OECD_FONTENEAU_D.pdf

⁹⁶ Piccirilo, F. (2025). Ethical Application of Generative AI in the EU: extending Deloitte's Trustworthy AI Framework for modern AI capabilities whilst adhering to the EU AI Act.

diversity.^{97, 98} The current dominance of English-centric training data (57.3% of open-source datasets on Hugging Face are in English) is a significant challenge, leading to weaker performance in other languages and risking underrepresenting cultural perspectives.^{99, 100} This necessitates greater investment in multilingual data collection and improved training strategies for AI models.¹⁰¹

Survey results reinforce the finding that **domain-specific datasets are considered essential for the development of generative AI in the EEA**¹⁰². Respondents consistently emphasized that access to data reflecting the specific characteristics of their sector or application area is more valuable than generic corpora. While **text data remains a fundamental** input across use cases, companies stressed that **structured tabular data and specialised sectoral databases** are indispensable for achieving accuracy, compliance, and practical relevance. By contrast, modalities such as images, audio, or video were regarded as of secondary importance at this stage, confirming that the most pressing need lies in ensuring availability of high-quality, domain-grounded datasets.

2.2.2.3 Volume and characteristics

The **scale of data required** for modern AI models is immense, often ranging from tens of gigabytes to exabytes for foundation model training.^{103, 104} For example, GPT-3 was trained on 570GB of text data, and Meta's unified storage system for AI training data can reach an exabyte scale (1 billion gigabytes)^{105, 106}. LLM training datasets double in size

⁹⁷ Ruokonen, M. (2025). APPLYING AI IN KEY EUROPEAN INDUSTRIES.

⁹⁸ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

⁹⁹ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

¹⁰⁰ Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

¹⁰¹ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

¹⁰² See Survey Question 6.

¹⁰³ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹⁰⁴ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

¹⁰⁵ Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

approximately every eight months, with recent models trained on trillions of tokens¹⁰⁷ (e.g., Llama 3.3 on 15T tokens).¹⁰⁸ This trend towards ever-larger datasets is driven by the fact that model performance generally increases with the volume, variety, and quality of data used for training.¹⁰⁹

However, sheer volume is not a guarantee of quality. The **desired data qualities** include accuracy, completeness, consistency, reliability, validity, and timeliness.^{110,111} Models operate on a "garbage in, garbage out" principle, meaning inaccurate or biased data will lead to flawed outputs^{112,113}. **Diversity** in datasets is crucial to prevent models from perpetuating societal biases and to ensure they can generalize across different contexts and populations.^{114,115} This includes demographic, linguistic, and cultural diversity.^{116,117}

¹⁰⁶ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

¹⁰⁷ A token is the fundamental unit of text—typically a word, part of a word (subword), or punctuation mark—that a language model processes and generates.

¹⁰⁸ Stanford Institute for Human-Centered AI. (2025, April 7). The AI Index Report 2025. <https://hai.stanford.edu/ai-index/2025-ai-index-report>

¹⁰⁹ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

¹¹⁰ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

¹¹¹ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

¹¹² Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹¹³ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. TechCrunch. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

¹¹⁴ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

¹¹⁵ Organisation for Economic Co-operation and Development. (2024, June 26). AI, data governance and privacy: Synergies and areas of international co-operation. OECD Publishing, Paris. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf

¹¹⁶ Ibidem

¹¹⁷ De Wilde, P., Arora, P., Buarque de Lima Neto, F., Chin, Y., Thinyane, M., Stinckwich, S., ... & Marwala, T. (2024). Recommendations on the Use of Synthetic Data to Train AI Models.

The **role of annotated and synthetic data** is increasingly significant. While pre-training often uses unannotated data, fine-tuning for specific tasks requires accurately labeled data.^{118, 119} **Synthetic data**, artificially generated to simulate real data, is gaining traction as a solution to data scarcity, privacy concerns, and the need for specific scenarios (e.g., in healthcare or autonomous vehicles).^{120, 121, 122} Synthetic data can augment existing datasets, create balanced and comprehensive sets, and even enable near-perfect labelling.^{123, 124} However, the quality and realism of synthetic data are critical, as poorly generated data can lead to inaccurate models, propagate biases, or even risk re-identification.^{125, 126, 127}

The **European "data deficit"** is a major concern. Strong privacy (GDPR) and copyright protections in the EU, while beneficial for privacy, can limit access to the vast datasets needed for training.^{128, 129} This, combined with a fragmented data infrastructure and a perceived conservatism in data sharing within European industry, creates a challenging environment for European GenAI development.^{130, 131, 132} There are also concerns about

¹¹⁸ Chen, J., & Mueller, J. (2024). Automated data curation for robust language model fine-tuning. arXiv preprint arXiv:2403.12776.

¹¹⁹ Myers, D., Mohawesh, R., Chellaboina, V.I. et al. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Comput* 27, 1–26 (2024). <https://doi.org/10.1007/s10586-023-04203-7>

¹²⁰ De Wilde, P., Arora, P., Buarque de Lima Neto, F., Chin, Y., Thinyane, M., Stinckwich, S., ... & Marwala, T. (2024). Recommendations on the Use of Synthetic Data to Train AI Models.

¹²¹ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. arXiv preprint arXiv:2403.04190.

¹²² Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

¹²³ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

¹²⁴ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. arXiv preprint arXiv:2403.04190.

¹²⁵ De Wilde, P., Arora, P., Buarque de Lima Neto, F., Chin, Y., Thinyane, M., Stinckwich, S., ... & Marwala, T. (2024). Recommendations on the Use of Synthetic Data to Train AI Models.

¹²⁶ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. arXiv preprint arXiv:2403.04190.

¹²⁷ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

¹²⁸ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

¹²⁹ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

¹³⁰ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

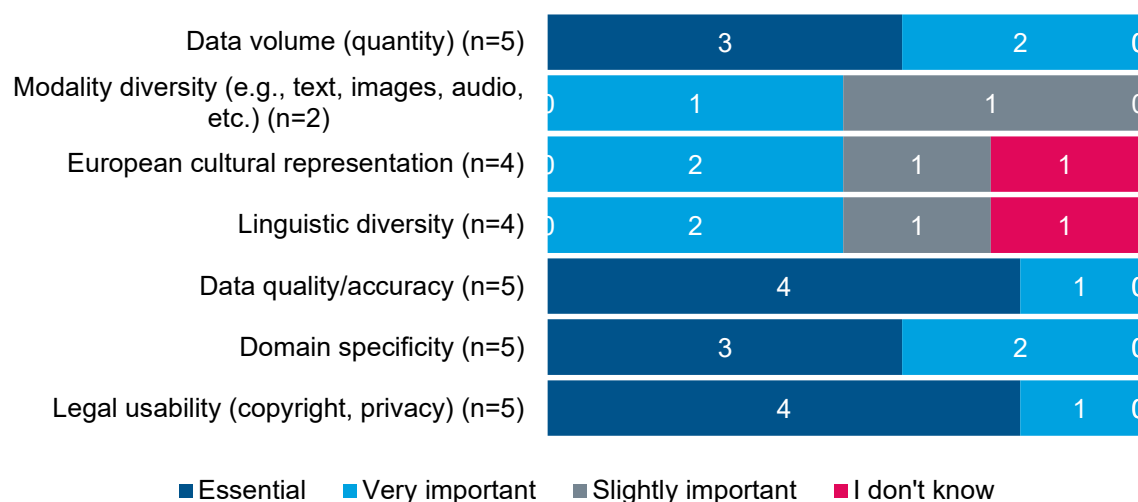
¹³¹ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

¹³² ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

running out of high-quality human-generated language data globally, with some estimates suggesting a potential exhaustion by 2024, which could lead to "model collapse" if future models are recursively trained on AI-generated content.^{133, 134}

As shown in the Figure below, survey evidence reinforces the dual **importance of scale and quality in GenAI development**.

Figure 2. Survey Q8 Results – Rate the importance of following data characteristics for GenAI models



All five respondents rated **data quality/accuracy** and **legal usability (copyright, privacy)** as either *essential* or *very important*, making these the most consistently prioritised characteristics. **Data volume** was also universally valued, with three respondents calling it *essential* and two *very important*. In addition, **domain specificity** was highlighted by all companies, split between *essential* and *very important* ratings, showing that access to sector-specific content is regarded as indispensable.

Other characteristics, while somewhat less unanimously prioritised, were still considered important by a majority. **Linguistic diversity** and **European cultural representation** were most often seen as *very important*, though one or two respondents were uncertain about their relevance. **Modality diversity** (e.g., text, images, audio, video) was rated lower overall, with only two respondents providing a score, suggesting that for many companies text and tabular data remain the most immediate concern. Taken together, the survey results underline that **large-scale, high-quality, legally usable and domain-grounded data** are seen as the foundations of successful GenAI development, with cultural, linguistic, and modality diversity recognised as valuable but not yet universally prioritised.

Input from an interview highlighted the importance of quality of data, besides their quantity and volume. Indeed, according to the interviewee, rigorous data checking and curation are essential. A notable example is AlphaFold, a system designed to predict protein

¹³³ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹³⁴ Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755–759.
<https://doi.org/10.1038/s41586-024-07566-y>

structures. Its success was largely attributed to the availability of high-quality, experimentally verified datasets from the Protein Data Bank. These curated datasets enabled the model to achieve near-experimental accuracy in its predictions, demonstrating the critical role of data integrity in AI outcomes.

Moreover, there is a critical distinction between data quantity and data coverage. While large datasets are often emphasized, coverage refers to the inclusion of all relevant scenarios, particularly edge and corner cases. Insufficient coverage can lead to biased and unreliable machine learning models, as they may fail to generalize across diverse real-world conditions.

2.2.2.4 Annotation and elaboration methods

Data processing techniques are crucial for preparing raw data for model training. For **text**, this involves cleaning, removing unwanted elements, and converting it into an appropriate format, such as tokens for language models.^{135, 136} For fine-tuning and specific tasks, **annotation** is a time-consuming and costly manual process, particularly for syntactic and semantically related jobs.^{137, 138} This can involve creating validated question-answer pairs, metadata, and semantic tagging.¹³⁹ **Human feedback**, specifically **Reinforcement Learning with Human Feedback (RLHF)**, is used to align model outputs with human preferences and values, which involves humans providing demonstration data and ranking outputs.^{140, 141}

Multilingual processing is critical for European GenAI. To address the scarcity of high-quality data for less-resourced languages, techniques like **data augmentation** (translating or paraphrasing existing material) and **transfer learning** are used.^{142, 143} The challenge of

¹³⁵ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹³⁶ Myers, D., Mohawesh, R., Chellaboina, V.I. et al. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Comput* 27, 1–26 (2024). <https://doi.org/10.1007/s10586-023-04203-7>

¹³⁷ Ibidem

¹³⁸ Chen, J., & Mueller, J. (2024). Automated data curation for robust language model fine-tuning. *arXiv preprint arXiv:2403.12776*.

¹³⁹ Fonteneau, F., & Anvar, E. (2025, May 2). Exploring the safe and responsible use of generative AI across the data lifecycle: Early insights from the OECD. OECD. https://unece.org/sites/default/files/2025-05/GenAI2025_S2_OECD_FONTENEAU_D.pdf

¹⁴⁰ European Data Protection Supervisor. (2024, June 3). Generative AI and the EUDPR: First EDPS Orientations for ensuring data protection compliance when using Generative AI systems. https://www.edps.europa.eu/system/files/2024-05/24-05-29_genai_orientations_en_0.pdf

¹⁴¹ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. *Bus Inf Syst Eng* 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

¹⁴² High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%20Implementing%20and%20Developing%20Chapter.pdf

¹⁴³ Myers, D., Mohawesh, R., Chellaboina, V.I. et al. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Comput* 27, 1–26 (2024). <https://doi.org/10.1007/s10586-023-04203-7>

processing multilingual data is compounded by the fact that over 90% of the world's languages have little to no support in language technology, and multilingual models often perform unevenly across languages.¹⁴⁴

For **image generation**, models are trained on large collections of images, often paired with descriptive text (captions or tags). Annotation involves linking images to their captions, and for specialized tasks, detailed labelling like object detection or segmentation may be required^{145, 146}. **Audio generation** models rely on large datasets of audio recordings, often with transcriptions and speaker labels¹⁴⁷. **Video generation** demands extensive video datasets, possibly with associated metadata or transcripts, and is complex due to high dimensionality and the need for frame-level annotation.^{148, 149}

Synthetic and multimodal data preparation involves generating artificial data that mimics real-world properties. This can include synthetic medical images, simulated driving data, or text generated by LLMs to augment existing datasets.^{150, 151} For multimodal models, datasets must integrate and align multiple data types (e.g., image-caption pairs, video with audio and text).¹⁵² The quality of synthetic data is paramount, as it must be diverse, correct, and natural to be effective, often requiring advanced prompt engineering and verbalizer techniques for generation.¹⁵³

Survey evidence¹⁵⁴ provides further context. Most respondents reported that their projects require only **moderate levels of annotation**, with very detailed annotation needs being the exception rather than the rule. This aligns with the dominant reliance on **fine-**

¹⁴⁴ Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency (pp. 610-623).

¹⁴⁵ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹⁴⁶ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. TechCrunch. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

¹⁴⁷ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

¹⁴⁸ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹⁴⁹ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

¹⁵⁰ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

¹⁵¹ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. arXiv preprint arXiv:2403.04190.

¹⁵² Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

¹⁵³ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. arXiv preprint arXiv:2403.04190.

¹⁵⁴ See Survey Question 8 and 9.

tuning foundation models rather than training them from scratch: four out of five firms reported following this approach, which reduces the need for extensive manual labeling and large-scale raw data preparation. In practice, this suggests that European developers are prioritising **targeted data preparation and annotation for domain-specific adaptation**, rather than investing in the full data engineering pipelines needed for large-scale model training.

2.2.2.5 Conceptual mapping of data needs

The conceptual framework for data needs in GenAI is driven by the hierarchical nature of model development and the specific requirements of different application stages.

At the **foundation model** level, the primary need is for **massive, diverse, and often unlabelled data**^{155, 156}. This data, sourced largely from the internet, includes raw text, images, audio, and video, aiming to provide a broad understanding of patterns and relationships across various modalities.^{157, 158} The goal is to build general capabilities that can be transferred to many downstream tasks.¹⁵⁹

When these foundation models are adapted for **specific tasks or domains (vertical AI)**, the data requirements shift towards **smaller, highly curated, and often annotated datasets**.^{160, 161} These datasets must be relevant, accurate, and specific to the target application. For example, legal AI models require specialized legal documents, while healthcare AI needs annotated patient data and imaging.¹⁶² The processing and preparation of this data are resource-intensive, involving cleaning, structuring, and human annotation, with particular attention to data quality dimensions and metrics relevant to the AI Act for high-risk systems.^{163, 164}

¹⁵⁵ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

¹⁵⁶ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

¹⁵⁷ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹⁵⁸ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. Bus Inf Syst Eng 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

¹⁵⁹ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

¹⁶⁰ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

¹⁶¹ Ruokonen, M. (2025). APPLYING AI IN KEY EUROPEAN INDUSTRIES.

¹⁶² Ruokonen, M. (2025). APPLYING AI IN KEY EUROPEAN INDUSTRIES.

¹⁶³ Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

¹⁶⁴ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing.

The European context introduces a critical demand for **multilingual and culturally representative data**. This is essential to ensure models are effective across the EU's diverse linguistic landscape and to prevent the perpetuation of biases from English-centric datasets.^{165, 166}

Synthetic data plays a crucial role in bridging data gaps, addressing privacy concerns, and enabling the creation of data for rare or sensitive scenarios.^{167, 168} It allows for the augmentation of real datasets and the generation of new, labelled data, especially where real data is scarce or restricted.¹⁶⁹

In essence, the data needs evolve from **general, massive, and raw** for foundational training to **specific, high-quality, annotated, and often synthetic** for fine-tuning and specialized applications. This entire process is underpinned by the critical need for **legal usability**, ensuring compliance with data protection (GDPR) and intellectual property regulations (copyright).^{170, 171} The data lifecycle, from collection to processing and deployment, must be coordinated between data spaces and AI Factories, with clear policies on security and data management.¹⁷²

Input from interviews further confirm and expand the results of the literature review. One academic expert highlighted that overriding objective for AI development is the quality and curation of data, whether collected or synthetically generated. They interviewee stressed the importance – especially for EU based companies – to ensure an extensive coverage of all relevant cases, including edge and corner case. Without sufficient coverage, machine learning models risk being unreliable and biased. Data collected in different regions will have varying coverage and inherent biases, and outsourcing data labelling can further affect quality.

The table overleaf synthesises the key findings on data needs for different GenAI solutions, focusing on the relationship between model types, data characteristics, and processing techniques. It provides a conceptual framework for understanding the ecosystem, drawing on quantitative and qualitative data from the report.

https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹⁶⁵ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

¹⁶⁶ Ruokonen, M. (2025). APPLYING AI IN KEY EUROPEAN INDUSTRIES.

¹⁶⁷ De Wilde, P., Arora, P., Buarque de Lima Neto, F., Chin, Y., Thinyane, M., Stinckwich, S., ... & Marwala, T. (2024). Recommendations on the Use of Synthetic Data to Train AI Models.

¹⁶⁸ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. arXiv preprint arXiv:2403.04190.

¹⁶⁹ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

¹⁷⁰ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

¹⁷¹ Piccirilo, F. (2025). Ethical Application of Generative AI in the EU: extending Deloitte's Trustworthy AI Framework for modern AI capabilities whilst adhering to the EU AI Act.

¹⁷² Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

Table 5. Conceptual framework for data needs

Type of GenAI solutions	Type of Model used	Data format	Volumes	Quality requirements	Scope for synthetic data	Annotation / Elaboration Techniques
Text Generation (LLM)	- Pre-trained foundation models - Adapted Models (Instruction-Tuned, Fine-Tuned & RLHF)	- PDF documents - Word documents (DOCX) - Text files (TXT) - Web pages (HTML) - Structured data in XML - Structured data in JSON	EXTREMELY HIGH Billions to trillions of tokens (exponentially expanding data sets)	HIGH Must be accurate, human-verified, and diverse to prevent bias and ensure generalizability. Multilingual and culturally relevant data is crucial for European contexts to support less common language	HIGH Potential to vast text corpora for pre-training or fine-tuning, particularly in low-resource languages and sensitive domains with limited real data	MODERATELY COMPLEX Annotation (sentiment, intent, named entities, context labels); Multilingual processing (data augmentation, transfer learning, multitask learning); Human Feedback Integration (RLHF).
Image Generation	- Diffusion Models - Multimodal models.	- JPEG - PNG - GIF - BMP - TIF.	HIGH Total stock estimated at 300 trillion tokens.	VERY HIGH High-quality, watermark-free images and accurate, detailed descriptive text (captions or tags) for training to achieve optimal output quality.	HIGH Especially useful when real image data is scarce (e.g., rare medical conditions) or sensitive (e.g., patient imaging) to augment datasets or generate new samples.	VERY COMPLEX Linking images to descriptive text (captions or tags); Image cleaning and filtering; Synthetic image generation for specific tasks (e.g., medical training).
Audio Generation/Processing	- Diffusion Models - Multimodal	- MP3 - WAV,	HIGH (unspecified)	HIGH Clear and high-fidelity audio. Accurate	MEDIUM Potentially used to augment datasets with a	VERY COMPLEX

	models.	- FLAC - OGG.	Large, multilingual audio datasets. No specific industry average volume is provided in the available sources.	transcription and speaker identification are essential for both model training and output quality.	variety of accents or linguistic variations and for privacy-preserving applications.	Frame-level labelling for actions, objects, scenarios; Temporal synchronization with other modalities; Video compression and quality optimization.
Video Generation	Multimodal models; Specialized generative models.	- MP4 - MOV - AVI, - WMV.	HIGH Total stock estimated at 1,350 trillion tokens. Requires massive volumes due to high dimensionality.	VERY HIGH High-resolution, clear motion, and accurate temporal annotations (e.g., frame-level labeling) are crucial for realistic generation and reliable analysis in applications like autonomous driving.	VERY HIGH Highly relevant for generating realistic scenarios for simulations (e.g., autonomous vehicles, digital twins) where real-world data collection is dangerous, expensive, or insufficient.	COMPLEX Frame-level labeling for actions, objects, scenarios; Temporal synchronization with other modalities; Video compression and quality optimization.
Multimodal & Domain-Specific Models	Multimodal models; Specialized foundation models	Mixed formats (e.g., JPEG, PDF, CSV, SQL databases, JSON).	WuDao 2.0 was trained on 4.9 TB of multimodal data. Specific domain data volumes are often smaller and more challenging to acquire.	VERY HIGH high-quality, curated, and diverse data across all modalities. Integration of structured and unstructured data is a key quality challenge. Annotation requires a high degree of domain expertise.	HIGH Crucial for applications where real-world data is sparse, heterogeneous, or sensitive, such as medical diagnostics, by generating augmented datasets or filling in missing values.	COMPLEX Cross-modal labeling (e.g., image-caption pairs), synchronization of data streams, fusion layers, shared embedding spaces, and expert-driven annotation for specialized domains.

2.2.3 Data sources and accessibility challenges

2.2.3.1 Overview of the available sources and current reliance

Generative AI development in Europe relies on a heterogeneous set of data sources, each with specific strengths and constraints:

Publicly available open datasets provide a low-cost entry point for training and fine-tuning, such as Wikipedia, Common Crawl, ImageNet, and EuroParl.^{173, 174, 175} They remain indispensable for experimentation and early development but often suffer from issues of quality, documentation, and legal provenance.

Web-scraped data constitutes the backbone of most foundation models, including GPT-3, GPT-4, and Llama.^{176, 177} While it provides vast scale, its use raises copyright, GDPR, and reputational concerns. Platforms increasingly restrict automated scraping, and under the EU Copyright Directive, rightsholders can opt out of text-and-data mining.¹⁷⁸

Commercially licensed datasets have become a critical source for larger players, with transactions involving Shutterstock, Associated Press, and Reddit often cited as examples.¹⁷⁹ Licensing provides legal certainty but is costly, creating barriers for SMEs and academic researchers.

Partnership-based data access is promoted through European initiatives such as the Common European Data Spaces, Catena-X, and ALT-EDIC.¹⁸⁰ These arrangements promise high-quality, domain-specific data, but uptake has so far been constrained by

¹⁷³ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

¹⁷⁴ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹⁷⁵ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

¹⁷⁶ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

¹⁷⁷ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

¹⁷⁸ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

¹⁷⁹ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. TechCrunch. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

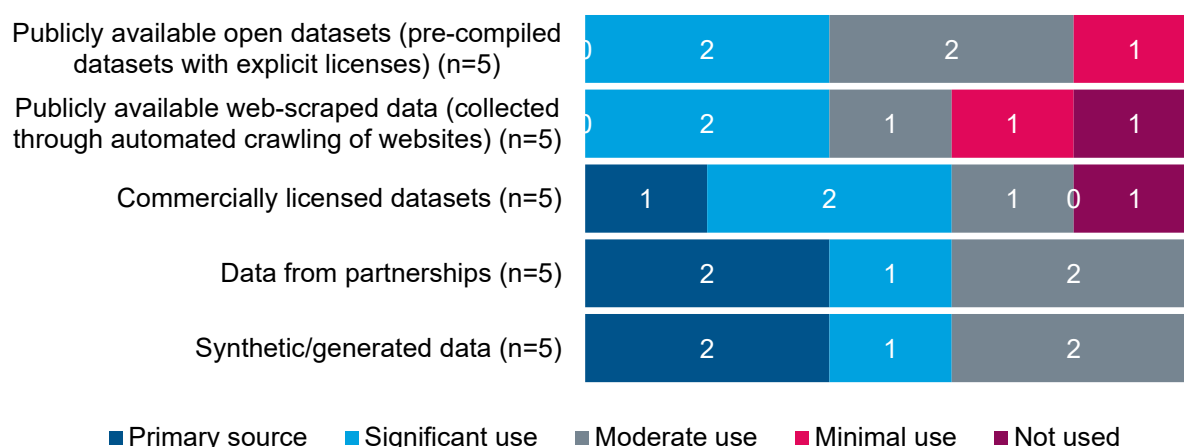
¹⁸⁰ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

trust deficits, coordination challenges, and transaction costs. To streamline access, initiatives like Catena-X are implementing standardized data-sharing contracts and have demonstrated commercial value by enabling suppliers and manufacturers (e.g., Ford, Koller Kunststofftechnik) to share Product Carbon Footprint (PCF) data, leading to measured reductions in calculated PCF values and efficiency gains.¹⁸¹

Synthetic data is an emerging complement, applied in areas where sensitive or scarce data is required, such as healthcare or autonomous vehicles.^{182, 183} While it helps mitigate privacy concerns and rebalance datasets, questions remain about realism, bias, and acceptance in regulated contexts.

The survey evidence must be interpreted cautiously, given the very limited sample size (n=5), which means it cannot be taken as representative of the broader European GenAI ecosystem. Nevertheless, the responses provide **useful indicative insights**, as shown in the figure below.

Figure 3. Survey Q13: How would you characterise your reliance on the following data sources for AI training?



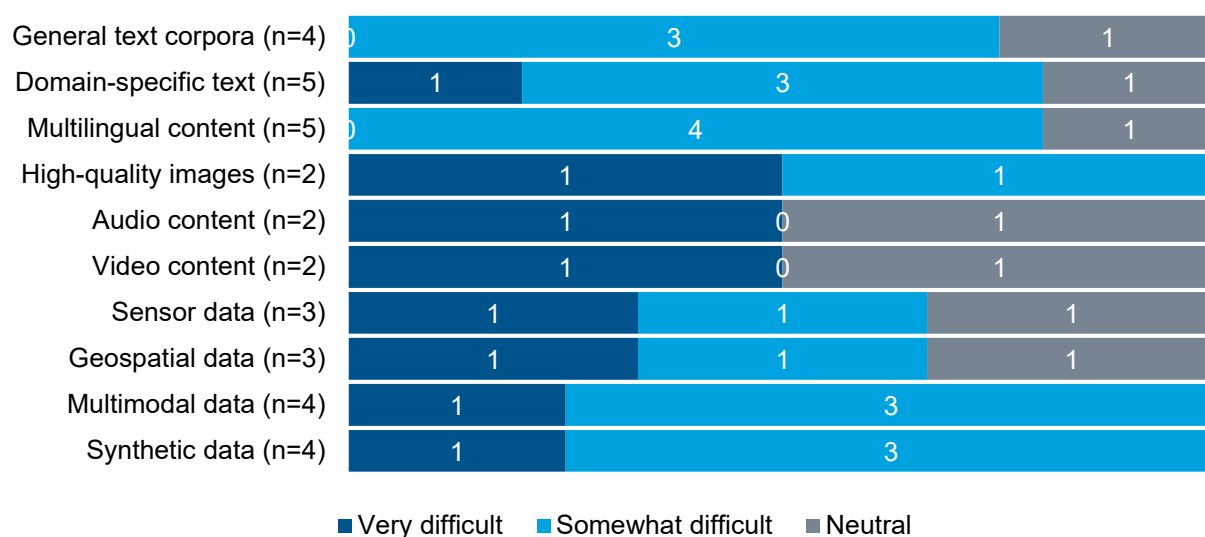
In addition to reliance patterns, the survey also explored **which types of data are most difficult to obtain (Q15a)**. Results indicate that **difficulties are widespread across almost all categories of data**. A majority of respondents rated access to **domain-specific text, multilingual content, multimodal datasets, and synthetic data** as either *very* or *somewhat difficult*. Similar challenges were reported for **sensor and geospatial data**, where most respondents again noted high levels of difficulty. Even data types that might be considered more widely available, such as **general text corpora or high-quality images**, were not described as straightforward to obtain: while these attracted more *neutral* responses than other categories, only a minority of respondents suggested they were unproblematic.

¹⁸¹ Catena-X. (2025). *Success Stories*. Available at: <https://catena-x.net/ecosystem/success-stories/>. (Details the PCF use case with Ford/Flex and the benefit of using standardized data-sharing policies.)

¹⁸² Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

¹⁸³ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. arXiv preprint arXiv:2403.04190.

Figure 4. Survey Q15a: Survey Q13: How would you characterise your reliance on the following data sources for AI training?



2.2.3.2 Legal and regulatory barriers

The regulatory landscape in Europe poses significant hurdles to accessing and using data for GenAI development. The **GDPR (General Data Protection Regulation)** is a central barrier, imposing strict rules on the collection, processing, and storage of personal data, which is often found in large internet-scale datasets.^{184, 185} Its core principles of lawfulness, fairness, transparency, data minimization, and purpose limitation are often difficult to reconcile with the mass collection and processing of personal data for AI training.^{186, 187} Obtaining explicit consent for web-scraped personal data is often impossible, and relying on exceptions like "disproportionate effort" is narrowly interpreted by authorities.¹⁸⁸ The GDPR's technology-neutral nature means its provisions apply regardless of the AI technology used, as long as personal data is processed.¹⁸⁹

Insights from expert interviews highlight that **GDPR does not inherently hinder innovation**. Rather, it can promote high-quality, sustainable, and responsible data

¹⁸⁴ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

¹⁸⁵ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

¹⁸⁶ Organisation for Economic Co-operation and Development. (2024, June 26). AI, data governance and privacy: Synergies and areas of international co-operation. OECD Publishing, Paris. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf

¹⁸⁷ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

¹⁸⁸ Ibidem

¹⁸⁹ Confederation of European Data Protection Organisations AI and Data Working Group. (2023, June 12). AI and Personal Data: A Guide for DPOs – Frequently Asked Questions. <https://cedpo.eu/wp-content/uploads/CEDPO-AI-and-Data-FAQ-12-June-2023.pdf>

ecosystems, with most challenges arising from misinterpretation and insufficient strategic engagement with the law. Many **organisations misunderstand GDPR by seeking top-down guidance without developing tailored strategies for data governance**, often resulting in overly cautious or superficial compliance approaches. Complaints about GDPR often reflect short-term thinking, whereas organisations that meaningfully engage with governance principles achieve more reliable and sustainable innovation outcomes. Furthermore, bypassing governance and relying on opaque AI systems, such as large language models from major companies, exposes organisations to both legal liabilities and the risk of unreliable knowledge.

Copyright and TDM (Text and Data Mining) issues are another major concern. While the EU's CDSM Directive provides TDM exceptions, rights holders can opt-out, requiring AI developers to obtain licenses for opted-out content.^{190, 191} This has led to a nascent licensing market for high-quality datasets, but with a great deal of legal uncertainty and ongoing litigation (e.g., the LAION case in Germany)^{192, 193, 194}. The lack of a single, standardized opt-out mechanism and the varying interpretations of fair use across jurisdictions (EU, US, Japan) create a complex and fragmented legal environment.^{195, 196}. The EU AI Act now requires providers of general-purpose AI models to have policies to comply with copyright law, respect TDM opt-outs, and publish sufficiently detailed summaries of training data.^{197, 198}

¹⁹⁰ University of Turin Law School & Nexa Center for Internet & Society. (2025, May 12). Development of Generative Artificial Intelligence from a Copyright Perspective. European Union Intellectual Property Office. https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf

¹⁹¹ European Copyright Society. (2025, January 15). European Copyright Society Opinion on Copyright and Generative AI. https://europeancopyrightsociety.org/wp-content/uploads/2025/02/ecs_opinion_genai_january2025.pdf

¹⁹² University of Turin Law School & Nexa Center for Internet & Society. (2025, May 12). Development of Generative Artificial Intelligence from a Copyright Perspective. European Union Intellectual Property Office. https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf

¹⁹³ European Copyright Society. (2025, January 15). European Copyright Society Opinion on Copyright and Generative AI. https://europeancopyrightsociety.org/wp-content/uploads/2025/02/ecs_opinion_genai_january2025.pdf

¹⁹⁴ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. TechCrunch. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

¹⁹⁵ University of Turin Law School & Nexa Center for Internet & Society. (2025, May 12). Development of Generative Artificial Intelligence from a Copyright Perspective. European Union Intellectual Property Office. https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf

¹⁹⁶ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

¹⁹⁷ University of Turin Law School & Nexa Center for Internet & Society. (2025, May 12). Development of Generative Artificial Intelligence from a Copyright Perspective. European Union Intellectual Property Office. <https://euipo.europa.eu/tunnel->

Competitive disadvantages for SMEs arise from this complex regulatory environment. Smaller players with limited legal and financial resources struggle to navigate the intricate web of compliance, licensing, and intellectual property (IP) issues.^{199, 200} The high compliance costs and legal uncertainty, combined with fragmented national approaches, hinder cross-border scaling and innovation for European startups.^{201, 202} The high compliance costs and legal uncertainty, combined with fragmented national approaches, hinder cross-border scaling and innovation for European startups.^{203, 204} 26, 47]. The EU AI Act, while aiming to foster trustworthy AI, is perceived by some startups as overly complex and potentially stifling innovation due to its broad definitions and stringent requirements.²⁰⁵

The regulatory landscape is also characterized by **fragmented enforcement** and varying interpretations across Member States. National data protection authorities may take different approaches to AI-related data processing, leading to inconsistencies and added complexity for companies operating across borders.^{206, 207} The overlap between AI regulations (like the AI Act) and existing privacy regulations (like the GDPR) can be a source of confusion and unclarity regarding compliance, especially for SMEs.²⁰⁸

[web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf](web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf)

¹⁹⁸ European Copyright Society. (2025, January 15). European Copyright Society Opinion on Copyright and Generative AI. https://europeancopyrightsociety.org/wp-content/uploads/2025/02/ecs_opinion_genai_january2025.pdf

¹⁹⁹ Implement Consulting Group. (2024, September 30). The economic opportunity of AI in the EU. <https://cms.implementconsultinggroup.com/media/uploads/articles/2024/The-economic-opportunity-of-generative-AI-in-the-EU/The-economic-opportunity-of-AI-in-the-EU.pdf>

²⁰⁰ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

²⁰¹ Implement Consulting Group. (2024, September 30). The economic opportunity of AI in the EU. <https://cms.implementconsultinggroup.com/media/uploads/articles/2024/The-economic-opportunity-of-generative-AI-in-the-EU/The-economic-opportunity-of-AI-in-the-EU.pdf>

²⁰² ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

²⁰³ Implement Consulting Group. (2024, September 30). The economic opportunity of AI in the EU. <https://cms.implementconsultinggroup.com/media/uploads/articles/2024/The-economic-opportunity-of-generative-AI-in-the-EU/The-economic-opportunity-of-AI-in-the-EU.pdf>

²⁰⁴ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

²⁰⁵ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

²⁰⁶ Confederation of European Data Protection Organisations AI and Data Working Group. (2023, June 12). AI and Personal Data: A Guide for DPOs – Frequently Asked Questions. <https://cedpo.eu/wp-content/uploads/CEDPO-AI-and-Data-FAQ-12-June-2023.pdf>

²⁰⁷ Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *International Data Privacy Law*, 7(2), 76–99. <https://doi.org/10.1093/idpl/ix005>

²⁰⁸ Organisation for Economic Co-operation and Development. (2024, June 26). AI, data governance and privacy: Synergies and areas of international co-operation. OECD Publishing,

Furthermore, the increasing adoption of barriers to cross-border data flows globally raises concerns about data accessibility for AI development.²⁰⁹

2.2.3.3 Technical barriers

The development of GenAI in Europe is hindered by several technical limitations related to data. A prevalent issue is the **lack of annotated and multilingual data**. High-quality, labeled datasets are scarce, especially for less commonly spoken European languages, impacting the performance of multilingual models.^{210, 211, 212} Over 90% of the world's languages have little to no support in language technology, and multilingual models often perform unevenly across languages, highlighting the scarcity of high-quality multilingual data.²¹³

Bias and cultural underrepresentation are significant technical risks. Large datasets scraped from the internet often overrepresent hegemonic viewpoints, encoding biases that can harm marginalized populations and perpetuate stereotypes (e.g., gender, racial, religious)^{214, 215}. The English-centric nature of most training data leads to "linguistic colonialism," where non-English languages and cultural nuances are overlooked or misrepresented, resulting in weaker performance and ethical concerns.^{216, 217}

The **risk of model collapse** is a growing technical concern. If future models are predominantly trained on data generated by their predecessors, they may lose information about the original data distribution, leading to degraded performance and reduced

Paris. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf

²⁰⁹ Organisation for Economic Co-operation and Development. (2024, June 26). AI, data governance and privacy: Synergies and areas of international co-operation. OECD Publishing, Paris. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf

²¹⁰ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

²¹¹ Myers, D., Mohawesh, R., Chellaboina, V.I. et al. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Comput* 27, 1–26 (2024). <https://doi.org/10.1007/s10586-023-04203-7>

²¹² Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

²¹³ Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

²¹⁴ Ibidem

²¹⁵ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. *Bus Inf Syst Eng* 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

²¹⁶ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%2C%20Implementing%20and%20Developing%20Chapter.pdf

²¹⁷ Manduchi, L., Pandey, K., Meister, C., Bamler, R., Cotterell, R., Däubener, S., ... & Fortuin, V. (2024). On the challenges and opportunities in generative ai. *arXiv preprint arXiv:2403.00025*.

diversity in outputs.²¹⁸ This structural risk is increasingly understood to be manageable, by strictly controlling the ratio of synthetic to real data and applying expert-level data governance. This "data pollution" highlights the need for access to real, human-produced data to sustain long-term learning and prevent the model from mis-perceiving the underlying learning task.²¹⁹

Furthermore, there is a significant **fragmentation of data standards**. Datasets often lack clear licensing information, provenance, and consistent metadata, making them difficult to reuse and integrate efficiently.^{220, 221, 222} For example, audits have found that over 70% of datasets on popular hosting sites lack adequate license information or are miscategorized.^{223, 224} This absence of interoperable standards and high-quality metadata hinders automation and complicates dataset curation, quality assessment, and legal compliance within data spaces.^{225, 226} The "black-box" nature of state-of-the-art AI models also makes it difficult to verify outputs and ensure reliability²²⁷.

2.2.3.4 Practical access and partnership barriers

Beyond regulatory and technical challenges, practical barriers significantly limit data accessibility for GenAI companies. **High licensing costs** for acquiring large, high-quality datasets are a major impediment, particularly for startups and academic

²¹⁸ Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). AI models collapse when trained on recursively generated data. *Nature*, 631(8022), 755–759. <https://doi.org/10.1038/s41586-024-07566-y>

²¹⁹ Ibidem

²²⁰ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

²²¹ Li, J., Zhang, Y., Xu, K., Wang, T., & Wang, H. (2024, July). Understanding the Challenges of Data Management in the AI Application Development. In 2024 IEEE International Conference on Joint Cloud Computing (JCC) (pp. 68-75). IEEE.

²²² Publications Office of the European Union. (2023, November 23). European data spaces and the role of data.europa.eu. <https://op.europa.eu/en/publication-detail/-/publication/70d01867-8ce1-11ee-8aa6-01aa75ed71a1/language-en>

²²³ Stanford Institute for Human-Centered AI. (2025, April 7). The AI Index Report 2025. <https://hai.stanford.edu/ai-index/2025-ai-index-report>

²²⁴ University of Turin Law School & Nexa Center for Internet & Society. (2025, May 12). Development of Generative Artificial Intelligence from a Copyright Perspective. European Union Intellectual Property Office. https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf

²²⁵ Publications Office of the European Union. (2023, November 23). European data spaces and the role of data.europa.eu. <https://op.europa.eu/en/publication-detail/-/publication/70d01867-8ce1-11ee-8aa6-01aa75ed71a1/language-en>

²²⁶ Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

²²⁷ Feuerriegel, S., Hartmann, J., Janiesch, C. et al. Generative AI. *Bus Inf Syst Eng* 66, 111–126 (2024). <https://doi.org/10.1007/s12599-023-00834-7>

researchers.^{228, 229} OpenAI, for instance, has reportedly spent hundreds of millions of dollars on licensing content from news publishers and stock media libraries, a budget far beyond the reach of smaller players.²³⁰ The global market for AI training datasets is currently valued at **\$2.6–3.2 billion** in 2024. Analysts project strong growth over the next decade, with forecasts ranging from **\$8.6 billion by 2030** to approximately **\$17 billion by 2032**. Other long-range estimates suggest the market could exceed **\$16 billion by 2034**.

231232233234235

This growth is being driven by rising demand for large, high-quality datasets to train generative AI and other machine learning systems. As competition intensifies, platforms and data providers have gained increasing leverage to charge higher prices for access to proprietary content.

Access is also hampered by **complex negotiations** and a general reluctance among data holders to share their valuable assets. Private data, often housed securely within company cloud environments, is typically not accessible for external model training due to confidentiality, privacy regulations, and competitive concerns.^{236, 237} Even public data, when scraped from the web, raises legal and ethical concerns, and platform operators often implement anti-scraping technologies or terms of service restrictions.²³⁸ The

²²⁸ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. TechCrunch. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

²²⁹ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

²³⁰ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. TechCrunch. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

²³¹ MarketsandMarkets. *AI Training Dataset Market by Type, Vertical, and Region – Global Forecast to 2029*. 2024. Available at: <https://www.marketsandmarkets.com/Market-Reports/ai-training-dataset-market-153819655.html>

²³² Grand View Research. *AI Training Dataset Market Size, Share & Trends Analysis Report, 2024–2030*. 2024. Available at: <https://www.grandviewresearch.com/industry-analysis/ai-training-dataset-market>

²³³ Fortune Business Insights. *AI Training Dataset Market Size, Share & Growth Forecast [2024–2032]*. 2024. Available at: <https://www.fortunebusinessinsights.com/ai-training-dataset-market-109241>

²³⁴ Global Market Insights. *AI Training Dataset Market Size, Industry Analysis, 2024–2034*. 2024. Available at: <https://www.gminsights.com/industry-analysis/ai-training-dataset-market>

²³⁵ Reuters. *Inside Big Tech’s Underground Race to Buy AI Training Data*. April 2024. Available at: <https://www.reuters.com/technology/inside-big-techs-underground-race-buy-ai-training-data-2024-04-05>

²³⁶ Ruokonen, M. (2025). APPLYING AI IN KEY EUROPEAN INDUSTRIES.

²³⁷ Microsoft Corporation. (2024, April). GDPR and Generative AI: A Guide for the Public Sector. <https://wwps.microsoft.com/wp-content/uploads/2024/04/GDPR-and-Generative-AI-A-Guide-for-the-Public-Sector-FINAL.pdf>

²³⁸ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

absence of standardized legal and technical frameworks for data sharing increases transaction costs and friction in partnerships.²³⁹

This friction is rooted in a fundamental lack of **trust in collaborations**. Organizations are hesitant to share sensitive or proprietary data due to fears of misuse, loss of control, or regulatory non-compliance.^{240, 241} The opacity of many AI models, which do not fully disclose their training data sources, further erodes trust and makes it difficult for data owners to verify compliance.^{242, 243} This climate of distrust can lead to a "data winter," where data holders become increasingly resistant to sharing, limiting the amount of data available for innovation.²⁴⁴

Open data limitations are also a factor. While open-source datasets exist (e.g., MNIST, ImageNet, Open Images, Common Crawl), they may be of lower quality, lack necessary metadata, or have unclear/conflicting licensing terms, creating legal risks and discouraging use by risk-averse developers.^{245, 246} The high computational cost and complex engineering requirements for training foundation models further restrict accessibility for most academic researchers and smaller organizations.²⁴⁷

2.2.3.5 Comparative assessment of challenges and barriers

The literature review highlights a broad spectrum of challenges facing European GenAI developers in accessing and using data. These range from legal and regulatory barriers (copyright, GDPR, contractual restrictions) to technical constraints (dataset quality,

²³⁹ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

²⁴⁰ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

²⁴¹ Organisation for Economic Co-operation and Development. (2024, June 26). AI, data governance and privacy: Synergies and areas of international co-operation. OECD Publishing, Paris. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf

²⁴² Piccirilo, F. (2025). Ethical Application of Generative AI in the EU: extending Deloitte's Trustworthy AI Framework for modern AI capabilities whilst adhering to the EU AI Act.

²⁴³ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

²⁴⁴ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

²⁴⁵ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. TechCrunch. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

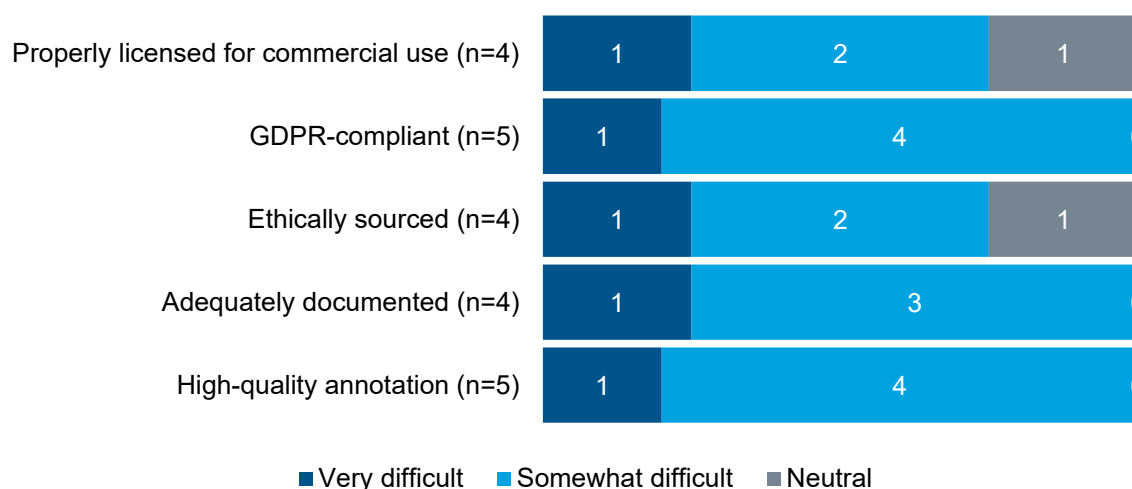
²⁴⁶ Organisation for Economic Co-operation and Development. (2025, February 9). Intellectual property issues in artificial intelligence trained on scraped data. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/02/intellectual-property-issues-in-artificial-intelligence-trained-on-scraped-data_a07f010b/d5241a23-en.pdf

²⁴⁷ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

representativeness, multilingual coverage) and practical obstacles (transaction costs, trust deficits in partnerships, concentration of resources among a few global players).

As shown in the Figure below, survey responses highlight that legal and economic barriers are the most frequently cited obstacles to data access.

Figure 5. Survey Question 16b - Rate the difficulty of accessing data that meets the following requirements



Survey responses to Q15b highlight a broad set of **barriers to making data usable for GenAI training**. Respondents most frequently identified **legal and compliance constraints**—notably GDPR and copyright—as significant difficulties. **Data quality and documentation issues** were also widely reported, with concerns about missing metadata, limited annotations, and lack of interoperability standards. In addition, respondents pointed to **fragmentation and silos** across sectors, which reduce the availability of rich, well-structured datasets that can be directly integrated into model pipelines.

The responses to open-ended questions provide further colour²⁴⁸. One company stressed that GDPR obligations—such as strict anonymisation, consent frameworks, and usage limitations—add substantial overhead to data collection and **slow down AI training cycles**. While these requirements were seen as necessary to safeguard ethical and responsible AI, the respondent underlined that they also **restrict access to rich human-centred datasets**, particularly those needed to simulate collaboration or design user-facing systems.

Another respondent, active in design decision support, explained that even when data is available, it often lacks the **metadata depth** needed for their applications. General-purpose corpora such as ShapeNet or COCO were accessible, but inadequate for systemic or sustainable design tasks. The absence of metadata on **material performance, environmental impact, or decision trade-offs**, combined with **proprietary silos** in automotive and industrial design, forced the company to rely on **manual curation, simulation, and hybrid augmentation strategies**. This case underscores that barriers to usability are not limited to compliance costs but extend to the **fundamental absence of domain-specific, richly annotated, and interoperable datasets**.

²⁴⁸ Respondents were given the possibility to further elaborate on their replies. Their inputs have been summarized and anonymized.

2.2.4 Cost of data acquisition, processing and compliance

This section explores the different type of costs related to data, namely acquisition, processing and compliance. While monetary and financial data for European companies are limited, overall cost drivers and factors are described below. Input from the market analysis (Section 2.2.4.5) add real quantitative examples and proxies extracted from the European GenAI context.

2.2.4.1 Licensing and external access costs

The **rising costs of acquiring datasets** are a significant financial burden for GenAI developers. The market for AI training data is projected to grow substantially, leading data brokers and platforms to charge premium prices for access²⁴⁹. OpenAI's expenditure of hundreds of millions of dollars on content licensing from news publishers and stock media libraries exemplifies the scale of these costs, which are largely out of reach for academic groups, nonprofits, and startups.²⁵⁰ For example, Shutterstock has inked deals with AI vendors ranging from \$25 million to \$50 million, and Reddit claims to have made hundreds of millions from licensing data²⁵¹. These high licensing fees contribute to the centralization of AI development among a few wealthy tech companies.^{252, 253} The uncertainty around pricing models for proprietary AI services also complicates long-term financial planning.²⁵⁴

Survey results add little clarity to this issue. When asked to estimate the **share of budgets dedicated to data licensing (Q21)**, respondents provided **highly scattered figures**, ranging from minimal shares to over 30 per cent, with no discernible pattern. This distribution suggests that companies may lack reliable tracking mechanisms or are reluctant to disclose precise figures. Similarly, when asked about the **absolute costs of licensing datasets (Q23)**, the majority of respondents answered “*prefer not to say*”, confirming that data licensing costs are considered a **highly sensitive and confidential issue**. Taken together, these responses highlight that there are currently **no confirmed figures or even credible ballpark estimates** of data licensing expenditures among EU-based GenAI companies.

As an additional research activity, an in-depth analysis of the financial statements of **LightOn**, the only identifiable European GenAI company for which such data is accessible, is provided in Section 2.2.4.5.

²⁴⁹ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. TechCrunch. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

²⁵⁰ Ibidem

²⁵¹ Ibidem

²⁵² Ibidem

²⁵³ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

²⁵⁴ High-Level Group for the Modernization of Official Statistics. (2025, March 17). Generative AI For Official Statistics. Using, Implementing and Developing Chapter. United Nations Economic Commission for Europe. https://unece.org/sites/default/files/2025-03/Gen%20AI%20For%20Official%20Statistics_Using%20Implementing%20and%20Developing%20Chapter.pdf

2.2.4.2 Internal data preparation costs

Beyond external licensing, **internal data preparation costs** are substantial. The resource intensity of **annotation and cleaning** is a major factor, as raw data needs to be meticulously processed, formatted (e.g., into tokens), and often labelled to be usable for model training.^{255, 256} These costs are highly variable, depending on factors such as the specific model training approach (e.g., fine-tuning versus training from scratch), the required level of annotation detail (e.g., basic labelling versus complex semantic tagging), and the data modality. This manual process is time-consuming and expensive, particularly for syntactic and semantically related jobs, making large-scale labelled datasets a huge problem for most NLP tasks.²⁵⁷ According to Gartner, poor data quality costs organizations an average of \$12.9 million annually, emphasizing the importance of proper data preparation.^{258, 259} This includes data collection, cleansing (standardizing formats, correcting misspellings, removing duplicates), and transformation (normalizing data scales, encoding categorical variables).^{260, 261} These internal costs, including labour and computational resources for data storage and cleaning, can be significant, especially for projects that require fine-tuning with specific, high-quality, and labelled datasets.^{262, 263}

2.2.4.3 Synthetic data: cost of access and use²⁶⁴

Synthetic data offers a compelling promise in addressing data scarcity and privacy concerns. It can be more cost-effective to generate than to collect real-world data, especially for applications that require costly collection, such as clinical trials or market

²⁵⁵ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing.
https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

²⁵⁶ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

²⁵⁷ Myers, D., Mohawesh, R., Chellaboina, V.I. et al. Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. *Cluster Comput* 27, 1–26 (2024).
<https://doi.org/10.1007/s10586-023-04203-7>

²⁵⁸ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

²⁵⁹ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

²⁶⁰ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

²⁶¹ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

²⁶² Deloitte. (2024, October 7). Technology Spotlight — Accounting for the Development of Generative AI Software Products.
<https://dart.deloitte.com/USDART/home/publications/deloitte/industry/technology/accounting-generative-ai-software-products>

²⁶³ Chen, J., & Mueller, J. (2024). Automated data curation for robust language model fine-tuning. arXiv preprint arXiv:2403.12776.

²⁶⁴ Please note that an in-depth analysis of the synthetic data landscape in Europe is provided in Section 2.4.

research.^{265, 266, 267} Synthetic data can also facilitate the creation of larger, more diverse datasets from a smaller amount of real data, and may allow for near-perfect labelling, thereby improving data quality and reducing manual annotation efforts.^{268, 269, 270} It provides significant advantages in terms of privacy, allowing organizations to train AI models without exposing sensitive or private information.^{271, 272, 273}

However, synthetic data comes with its own challenges and price. While it can be more cost-effective than real-world data collection, the **computational and environmental costs** of generating high-quality synthetic data can still be substantial.²⁷⁴ Furthermore, synthetic data does not necessarily equate to anonymous data, and there remains a **risk of re-identification** or inadvertent revelation of underlying training data elements. If these revealed real data elements originated from licensed datasets, their accidental exposure constitutes a potential breach of the original licensing agreements.^{275, 276} This highlights the need for careful management and robust ethical guidelines in its generation and use.²⁷⁷ Poorly generated synthetic data can lead to inaccurate and unreliable AI models, and biases present in original data can be propagated or amplified.²⁷⁸

As pointed out by one interviewee, **synthetic data is often positioned as a privacy-preserving solution**, particularly for sensitive datasets subject to regulatory constraints

²⁶⁵ De Wilde, P., Arora, P., Buarque de Lima Neto, F., Chin, Y., Thinyane, M., Stinckwich, S., ... & Marwala, T. (2024). Recommendations on the Use of Synthetic Data to Train AI Models.

²⁶⁶ Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., ... & Roseland, J. (2019). Towards federated learning at scale: System design. *Proceedings of machine learning and systems*, 1, 374-388.

²⁶⁷ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

²⁶⁸ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

²⁶⁹ European Union Intellectual Property Office. (2025). The development of generative artificial intelligence from a copyright perspective. EUIPO.

²⁷⁰ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

²⁷¹ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward

²⁷² European Union Intellectual Property Office. (2025). The development of generative artificial intelligence from a copyright perspective. EUIPO.

²⁷³ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

²⁷⁴ De Wilde, P., Arora, P., Buarque de Lima Neto, F., Chin, Y., Thinyane, M., Stinckwich, S., ... & Marwala, T. (2024). Recommendations on the Use of Synthetic Data to Train AI Models.

²⁷⁵ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

²⁷⁶ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. arXiv preprint arXiv:2403.04190.

²⁷⁷ Ibidem

²⁷⁸ De Wilde, P., Arora, P., Buarque de Lima Neto, F., Chin, Y., Thinyane, M., Stinckwich, S., ... & Marwala, T. (2024). Recommendations on the Use of Synthetic Data to Train AI Models.

such as the General Data Protection Regulation (GDPR). While synthetic data can mimic the structure and statistical properties of original datasets, its effectiveness as a privacy safeguard remains contested. If synthetic data retains the unique and informative characteristics of the source data, it remains vulnerable to traditional privacy attacks, thereby undermining its protective function. Conversely, when synthetic data is rigorously tested and shown to be robust against privacy risks, it often lacks the utility and fidelity that made the original data valuable for downstream applications.

2.2.4.4 Regulatory costs

The need to comply with a complex and fragmented regulatory framework adds another layer of cost. The **GDPR** (General Data Protection Regulation) and the **AI Act** are the two main pillars shaping this landscape.^{279, 280} The GDPR imposes significant compliance costs on GenAI developers, including expenses for data protection impact assessments (DPIAs), risk management systems, and technical documentation.^{281, 282} The AI Act introduces additional obligations for high-risk and general-purpose AI systems, such as requirements for data governance, transparency, and human oversight, which incur further costs for providers and deployers.^{283, 284}

Compliance with **copyright** and licensing requirements can be high, with the need to either license content or navigate complex opt-out mechanisms.^{285, 286} The uncertainty of the legality of some data usages raises risks for prospective developers who may need to invest in understanding the stance of different jurisdictions.²⁸⁷ This regulatory complexity and fragmentation across national rules, along with the lack of clarity on future rules, create uncertainty and can increase legal and administrative costs, disproportionately

²⁷⁹ Confederation of European Data Protection Organisations AI and Data Working Group. (2023, June 12). AI and Personal Data: A Guide for DPOs – Frequently Asked Questions. <https://cedpo.eu/wp-content/uploads/CEDPO-AI-and-Data-FAQ-12-June-2023.pdf>

²⁸⁰ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

²⁸¹ Confederation of European Data Protection Organisations AI and Data Working Group. (2023, June 12). AI and Personal Data: A Guide for DPOs – Frequently Asked Questions. <https://cedpo.eu/wp-content/uploads/CEDPO-AI-and-Data-FAQ-12-June-2023.pdf>

²⁸² Piccirilo, F. (2025). Ethical Application of Generative AI in the EU: extending Deloitte's Trustworthy AI Framework for modern AI capabilities whilst adhering to the EU AI Act.

²⁸³ Ibidem

²⁸⁴ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

²⁸⁵ European Copyright Society. (2025, January 15). European Copyright Society Opinion on Copyright and Generative AI. https://europeancopyrightsociety.org/wp-content/uploads/2025/02/ecs_opinion_genai_january2025.pdf

²⁸⁶ University of Turin Law School & Nexa Center for Internet & Society. (2025, May 12). Development of Generative Artificial Intelligence from a Copyright Perspective. European Union Intellectual Property Office. https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf

²⁸⁷ Organisation for Economic Co-operation and Development. (2024, May 24). Artificial intelligence, data and competition. OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/05/artificial-intelligence-data-and-competition_9d0ac766/e7e88884-en.pdf

affecting smaller companies with limited resources^{288, 289}. Only larger companies may have the financial means and incentive to bear these costs.²⁹⁰

2.2.4.5 Quantitative estimation from market data

To compensate for the limited quantitative information available in existing EU-based literature and policy reports, this section draws on financial data from public filings and market-level analysis. However, the data landscape remains highly fragmented. The vast majority of generative AI companies in Europe are privately held, and many are still in early development stages—such as seed or pre-IPO phases—where financial disclosures are minimal and often treated as commercially sensitive.

Due to the absence of consolidated or standardized cost reporting, especially concerning data acquisition and processing, this section leverages available information from a small sample of public or semi-public sources. These include annual reports from LightOn (France) and Teneo.ai (Sweden), as well as commercial pricing benchmarks from leading annotation and labeling service providers. While not exhaustive, the insights offered here aim to provide a preliminary ballpark estimate and serve as a reference point for understanding operational cost structures in the European GenAI ecosystem.

Input from Annual Reports: LightOn

LightOn is a French generative AI startup headquartered in Paris, specializing in the development of foundation models and customizable GenAI solutions for enterprise and public sector clients. The company offers proprietary large language models (LLMs) optimized for deployment in secure, on-premise or hybrid environments, with a strong emphasis on data sovereignty and privacy. LightOn's technology stack includes:

- **Auriga, a foundational LLM family designed for private deployment.**
- **A specialized LLM-as-a-service platform for regulated industries.**
- **Tailored GenAI tools focused on search, summarization, and conversational interfaces.**

The company positions itself as a European alternative to U.S.-based LLM providers, aiming to deliver scalable, privacy-first GenAI infrastructure suited for high-compliance environments such as government, defense, and healthcare. Like most developers of foundational models, LightOn is likely to rely in part on publicly available or licensed third-party datasets to train its language models, though specific sources are not publicly disclosed.

As of 2024, LightOn is also the only publicly listed GenAI scale-up in Europe, having joined Euronext Growth Paris. According to its reporting²⁹¹, the listing enables the

²⁸⁸ Implement Consulting Group. (2024, September 30). The economic opportunity of AI in the EU. <https://cms.implementconsultinggroup.com/media/uploads/articles/2024/The-economic-opportunity-of-generative-AI-in-the-EU/The-economic-opportunity-of-AI-in-the-EU.pdf>

²⁸⁹ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

²⁹⁰ Implement Consulting Group. (2024, September 30). The economic opportunity of AI in the EU. <https://cms.implementconsultinggroup.com/media/uploads/articles/2024/The-economic-opportunity-of-generative-AI-in-the-EU/The-economic-opportunity-of-AI-in-the-EU.pdf>

²⁹¹ LightOn, Yearly Report (2024): [https://cdn.prod.website-files.com/6668dab2882cdb0ec42fe47e/67fe9fa08db26a4a2fb10ab6_LightOn%20-%20RA%202024%20\(v.15.04.25\).pdf](https://cdn.prod.website-files.com/6668dab2882cdb0ec42fe47e/67fe9fa08db26a4a2fb10ab6_LightOn%20-%20RA%202024%20(v.15.04.25).pdf)

company to attract broader investor interest while reinforcing its strategic focus on sovereign AI infrastructure and transparency in regulated sectors.

In 2024, LightOn underwent a strategic transition that had a significant impact on its financial performance. Revenue fell sharply to **€1.13 million**, down from **€7.97 million** in 2023. Despite an increase in other operating income to **€1.57 million**, the company's operating costs rose substantially. Personnel expenses nearly doubled, reaching **€3.86 million**, and external charges increased to **€3.24 million**.

As a result, the company recorded a negative **operating result of €(5.06) million**, compared to a positive result of **€3.68 million** in 2023. The **net result** also turned negative at **€(4.81) million**, reflecting the combined effects of declining revenue and rising operational expenses.

Given these exceptional circumstances, the assessment of cost structure and the identification of possible cost related to data acquisition, processing and compliance is carried out on both fiscal year 2024 and 2023. The table below presents a comparison of the results. These are summarized in the table below, which shows no significant variance in terms cost structure.

Table 6. Extract from LightOn Yearly Report (2024)

Cost items		Amount 2024 (,000 EUR)	% of Total 2024 Operating Costs	Amount 2023 (,000 EUR)	% of Total 2023 Operating Costs
Staff cost	Salaries, social security contributions, and benefits related to expanded R&D and technical staff.	3860	49.7	2016	42%
Purchase & external charges	Expenses for cloud services, software licenses, consulting, and outsourced R&D activities.	3239	41.8	2147	45%
Taxes	Local business taxes and administrative fees.	2	0.03	28	5%
Depreciations and provisions	Amortization of assets, capitalized development, and provisions for risks and liabilities.	656	8.5	548	11%

Source: LightOn (2024)

As illustrated in the figure, **Staff Costs** and **Purchases & External Charges** together account for approximately **90% of LightOn's total operating costs** in 2024. However, the **Annual Report does not provide a detailed cost breakdown**, limiting visibility into the specific expenditures associated with core GenAI operations such as data acquisition, processing, and compliance.

Several observations can be made based on the available financial statements:

- The cost of **third-party licensed datasets** is **not explicitly disclosed**. These expenses are likely embedded within the broader *Purchases & External Charges* category, which also includes **cloud and compute infrastructure**, software tools, and general **subcontracted services**.
- **Staff Costs** encompass internal labour, including data-related tasks such as **processing, annotation, and curation**. However, the report does not indicate what share of these costs is directly attributable to data-centric activities.
- **R&D expenditures** are not itemized as standalone operational costs. Instead, they are **capitalized as intangible assets**, which makes it difficult to assess how much investment is directed toward data training, model tuning, or related efforts.

This lack of cost granularity may reflect a strategic decision to withhold commercially sensitive information. However, based on the structure of LightOn's balance sheet, it is reasonable to infer that external purchases — potentially including licensed datasets — represent a cost category comparable in scale to internal staffing expenses related to data processing and model training.

This interpretation is further supported when benchmarking LightOn against Teneo.ai²⁹², one of the few publicly listed European companies operating in the AI space (listed on the Stockholm Stock Exchange). Unlike LightOn, Teneo.ai does not develop its own foundational models, and is therefore less reliant on third-party data acquisition. This distinction is reflected in its cost structure: in 2024, staff costs accounted for 60% of total operating expenses, while external purchases represented only 30%.

The divergence may also be attributed to differences in business models. Teneo.ai focuses on delivering conversational AI platforms and enterprise dialogue automation tools, relying on scripting and rule-based NLU frameworks rather than generative model training. As a result, its operational cost profile is more service-oriented, with fewer infrastructure or data licensing needs compared to a foundational model developer like LightOn.

In conclusion, producing a clear monetary estimate of the costs associated with data acquisition, processing, and management in the European GenAI sector remains extremely challenging. This difficulty stems from several structural and informational barriers. First, there is a **lack of aggregated and standardized financial data** specific to GenAI operations, particularly concerning datasets and infrastructure. Second, many GenAI companies—especially those involved in foundational model development—treat **data-related expenditures as commercially sensitive**, often embedding them in broad accounting categories such as “external purchases” or “R&D.” Third, **benchmarking across firms is hindered by the fact that only a small number of European GenAI companies are publicly listed**, limiting the availability of comparable financial disclosures. Finally, the **relative novelty of the sector in Europe**, coupled with its still-developing ecosystem, means that cost structures are highly variable and evolving, further complicating efforts to draw generalizable insights.

AI training data providers cost ranges

While commercially licensed datasets are addressed in a separate section, a useful proxy for estimating the costs associated with data acquisition and preparation is the market pricing of annotation and labelling services. These platforms form a critical operational layer in the GenAI development pipeline, particularly for companies building or fine-tuning foundational models. They support essential processes such as supervised learning, reinforcement learning with human feedback (RLHF), and domain-specific data curation. Annotation and labelling tasks are increasingly outsourced to specialized providers offering scalable solutions for data tagging, classification, entity recognition, transcription, and alignment. In Europe, the market is served by a combination of local and global actors, many of whom offer GDPR-compliant services and localized workforce integration. The table below summarizes key platforms, based on industry research and vendor documentation.

²⁹² Teneo.ai, Yearly Report (2024): <https://www.teneo.ai/investors/financial-information/financial-reports>

Table 7. Overview of data providers cost range²⁹³

Provider	EU Presence / HQ	Service Area	Estimated Pricing
Unitlab Annotate	\$99/month (Active Plan)	10K source images + 10K auto-labelling	AI-driven annotation, human-in-the-loop, dataset management
Roboflow Annotate	\$65/month (Basic Plan)	30 credits (10K stored, 10K uploaded, 2.7K AI labels)	Dataset management, preprocessing tools, model training
ScaleAI (Self-Serve)	Pay-as-you-go	10K free uploads + 1K data labels	End-to-end service, 3D LiDAR, model-assisted curation
SuperAnnotate	Custom (Pro Plan)	Configurable	Vendor/data management, curation, analytics, integrations
DataLoop	Tailored pricing	Custom limits	Versioning, pipeline integration, unstructured data support
LabelBox	\$16.67/month + \$1,000 one-time	Pay-as-you-go (LabelBox Units - LBU)	Collaboration, ML pipeline integration, intuitive interface

While the comparison table provides indicative pricing for various data annotation platforms, it is important to note that actual pricing structures are often far more complex and opaque. Most vendors operate on custom pricing models or use flexible credit-based systems that depend heavily on project specifications, volume, annotation type, turnaround time, and service-level agreements.

Even platforms that publish basic monthly rates (e.g., Unitlab or Roboflow) tend to reserve more advanced features, large-scale projects, or enterprise-level support for negotiated contracts. Providers like ScaleAI, SuperAnnotate, and DataLoop do not disclose pricing publicly and require direct engagement for quotations, making it difficult to obtain consistent or comparable benchmarks.

2.2.4.6 Comparative cost assessment and limitations

The evidence collected for this study shows that the **cost of data acquisition, processing, and compliance is a multi-dimensional challenge** for European GenAI developers. These costs can be grouped into four main categories:

- **Licensing and external access costs**, which have risen sharply as large-scale data brokers and platforms charge premium fees for content. These are exemplified by OpenAI's deals with news publishers and stock media companies worth hundreds of millions of dollars, or Shutterstock's \$25–50 million licensing contracts with GenAI vendors.²⁹⁴

²⁹³ Authors elaboration on data collected from BasicAI.blog: <https://www.basic.ai/>

²⁹⁴ Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. TechCrunch. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

- **Internal data preparation costs**, covering annotation, cleaning, and formatting, which are time- and labour-intensive. Gartner estimates that poor data quality costs organisations an average of \$12.9 million annually,²⁹⁵ ²⁹⁶ underscoring the scale of these burdens.
- **Synthetic data generation costs**, which can be lower than collecting real-world datasets but nonetheless require significant compute resources and careful quality control.
- **Regulatory compliance costs**, associated with GDPR and the forthcoming AI Act, which demand investment in anonymisation, risk management, documentation, and copyright licensing.

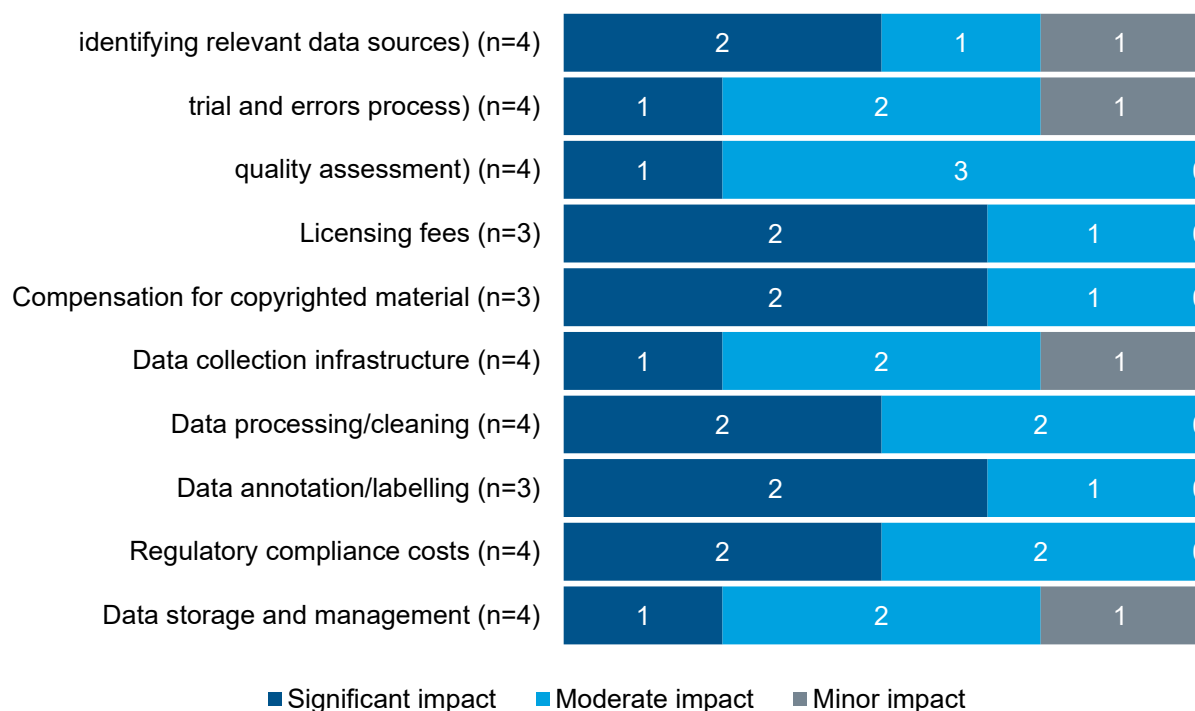
Financial reporting from European GenAI firms sheds partial light on these components. For example, LightOn's 2024 financial statements show that **staff costs (49.7%) and external purchases (41.8%)** dominate its operating expenses, together accounting for nearly 90% of total costs. While these categories undoubtedly include data-related expenditures—such as licensing and annotation—they are aggregated with other items such as cloud infrastructure, software licences, and subcontracted services, preventing precise attribution. This lack of granularity illustrates why quantitative evidence remains sparse: cost structures are commercially sensitive and embedded within broad accounting lines.

Survey evidence provides additional insights, though with limitations. When asked to rate the **most burdensome costs (Q22)**, respondents most frequently pointed to **licensing fees, internal data preparation, and compliance overheads** as the categories with the greatest impact. These findings echo the literature's identification of rising licensing costs and intensive manual processing as key challenges. Qualitative feedback reinforces this view: one respondent emphasised that **GDPR compliance slows down data collection and training cycles**, adding to costs through anonymisation, consent frameworks, and usage restrictions. Another highlighted that **manual curation and annotation of sector-specific data** remain unavoidable, particularly in industries where metadata is scarce or siloed.

²⁹⁵ Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258.

²⁹⁶ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. Global Partnership on AI. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

Figure 6. Survey Responses Q22 - Rate the impact of the following cost factors on your data acquisition/processing budget



Taken together, the desk and survey evidence show that **the cost of data acquisition, processing, and management is not only a major operational burden, but also a clear competitive dimension of the GenAI market**. Cost structures are inherently shaped by each company's business model—for example, whether they license external data at scale, rely on proprietary synthetic generation, or invest in in-house annotation teams. At the same time, because these costs are a competitive factor, they are typically treated as **confidential and commercially sensitive**, making it extremely difficult to provide accurate estimates.

2.2.5 Data spaces and alternative access models

This section provides an overview of stakeholders' awareness and participation in EU policies and initiatives in the field of AI and data. In addition, it draws on insights from the literature review and stakeholder consultation to highlight expectations regarding Data Spaces and their potential contribution to strengthening the uptake of generative AI in Europe. The policies and initiatives considered are the following:

- Common European Data Spaces (CEDs) & GAIA-X Infrastructure:** Data Spaces are a core concept for realizing a single European data market. The foundational, conceptual, and technical work to bring these Data Spaces to life has been developed – for instance - under the GAIA-X collaboration framework. As a result of this initiative, several operational Data Spaces are now being built. Additionally, the European Commission supports the emergence of "Common European Data Spaces" in specific domains financially, working in close collaboration with the relevant stakeholders in the GAIA-X ecosystem. These infrastructures aim to facilitate trusted data sharing across the EU, providing a federated architecture, technical standards, and governance frameworks to allow organizations to pool, exchange, and re-use high-quality datasets while reducing dependence on non-EU providers. Together, GAIA-X and the CEDs initiative form the technical backbone for Europe's vision of a secure and interoperable data economy, creating the conditions

under which the principles of the Data Governance Act can be implemented in practice.

- **AI Factories**, a more recent initiative launched under the European Commission's AI package (2023), which aims to create large-scale hubs combining compute, data, and expertise. They are designed to serve as open infrastructures for startups and SMEs, offering access to high-performance computing, curated data, and testing facilities to accelerate the development of trustworthy GenAI applications. By linking compute and data resources within the broader Data Space ecosystem, AI Factories are expected to help operationalize data availability for AI development, ensuring that European innovators can build and test GenAI models under trusted and sovereign conditions.
- **The Data Governance Act (DGA)**, adopted in 2022 provides the foundational legal framework for establishing the trust and mechanisms necessary for B2B data sharing within the CEDS. It introduces the concept of "data intermediation services" as neutral organizations that act as crucial infrastructure to connect data supply (from individuals/organizations) with data demand (e.g., from AI developers). The DGA's rules govern these specific actors to ensure they function as trustworthy organizers of data sharing. It also sets requirements for the reuse of non-open public sector data under certain conditions and establishes rules for data altruism to encourage voluntary data sharing by individuals and organizations. In this sense, the DGA provides the governance and regulatory layer that complements the technical infrastructure of the Common European Data Spaces, ensuring that data exchange in the EU occurs under clear, transparent, and trustworthy conditions.

2.2.5.1 Awareness and participation level

The literature review provided limited insights concerning the current awareness and participation level of EEA-based GenAI companies in EU Data Spaces. According to the consulted sources, awareness and participation levels in data spaces, particularly among SMEs, remain a significant challenge. Despite acknowledging their potential, participation and access to common data spaces remain limited, especially amongst SMEs. European startups report difficulties in accessing cutting-edge AI technologies and cite the EU's lack of good data and clear rules as significant obstacles.²⁹⁷ Many SMEs lack the necessary knowledge, skills, and funds to effectively engage with and benefit from data-sharing initiatives.^{298, 299} The fragmented European landscape, with its linguistic and cultural barriers, also limits the ability of startups to distribute their services at scale, further impacting their motivation and capacity to participate in cross-border data spaces.³⁰⁰

²⁹⁷ Implement Consulting Group. (2024, September 30). The economic opportunity of AI in the EU. <https://cms.implementconsultinggroup.com/media/uploads/articles/2024/The-economic-opportunity-of-generative-AI-in-the-EU/The-economic-opportunity-of-AI-in-the-EU.pdf>

²⁹⁸ Google. (2023, November 14). Google's EU AI Opportunity Agenda. https://static.googleusercontent.com/media/publicpolicy.google/en/resources/eu_ai_opportunity_agenda_en.pdf

²⁹⁹ Implement Consulting Group. (2024, September 30). The economic opportunity of AI in the EU. <https://cms.implementconsultinggroup.com/media/uploads/articles/2024/The-economic-opportunity-of-generative-AI-in-the-EU/The-economic-opportunity-of-AI-in-the-EU.pdf>

³⁰⁰ Le Picard, H. (2025, April). European Startups and Generative AI: Overcoming Big Tech Dominance. Ifri. https://www.ifri.org/sites/default/files/2025-05/ifri_le_picard_european_startups_genai_2025.pdf

There is a perceived conservatism within the European industry regarding data sharing, with one startup stating, "The European industry is too conservative when it comes to sharing data. High quality data is our only chance to achieve long term competitive advantage"³⁰¹. This reluctance, combined with bureaucratic effort and uncertainty around accessing resources like EuroHPC JU's supercomputers, creates further barriers for SMEs.³⁰²

Survey evidence broadly confirms this limited engagement. When asked about familiarity with EU data and AI policy initiatives³⁰³, most respondents reported only slight or moderate familiarity. For instance, in relation to **Common European Data Spaces**, three respondents were slightly familiar, one was moderately familiar, and one reported high familiarity. Similar patterns were found for **GAIA-X**, where two respondents were slightly familiar, one moderately familiar, and one not at all familiar. For the **AI Factories initiative**, two respondents were moderately familiar, one slightly familiar, and two not at all familiar. Even for the **Data Governance Act**, two respondents indicated slight familiarity and two moderate familiarity.

2.2.5.2 Expectations from Data Spaces

Data spaces are widely anticipated to address many of the current data needs and accessibility challenges in the GenAI ecosystem. Key **anticipated benefits** include:

- **Quality data:** Data spaces are expected to provide access to high-quality, curated, and diverse datasets that are essential for training robust and accurate GenAI models.^{304, 305} This includes ensuring data is "AI-ready" through semantic standards and metadata enrichment, and guaranteeing data quality throughout its lifecycle.^{306, 307}
- **Multilingual coverage:** Initiatives within data spaces, such as the Alliance for Language Technologies European Digital Infrastructure Consortium (ALT-EDIC), aim to centralize access to language resources, thereby increasing the availability of European language data and supporting the development of multilingual LLMs.^{308, 309}

³⁰¹ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

³⁰² Ibidem

³⁰³ See Question 27 Results

³⁰⁴ Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

³⁰⁵ Publications Office of the European Union. (2023, November 23). European data spaces and the role of data.europa.eu. <https://op.europa.eu/en/publication-detail/-/publication/70d01867-8ce1-11ee-8aa6-01aa75ed71a1/language-en>

³⁰⁶ Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

³⁰⁷ Fonteneau, F., & Anvar, E. (2025, May 2). Exploring the safe and responsible use of generative AI across the data lifecycle: Early insights from the OECD. OECD. https://unece.org/sites/default/files/2025-05/GenAI2025_S2_OECD_FONTENEAU_D.pdf

³⁰⁸ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

Governance and legal certainty: Data spaces are designed to provide interoperable frameworks with common governance principles and enabling services to facilitate trusted and secure data transactions.^{310, 311} They aim to overcome legal and technical barriers to data sharing by combining necessary tools and infrastructures and addressing issues of trust and compliance with regulations like GDPR, the Data Governance Act (DGA), and the AI Act.^{312, 313} The DGA, for instance, regulates data intermediaries to ensure they function as trustworthy organizers of data sharing.³¹⁴

- **Innovation support:** By unlocking currently inaccessible industrial data and fostering collaboration, data spaces are expected to offer a significant advantage to AI innovators and enhance the mission of AI Factories³¹⁵. They can enable the pooling of data to create larger, more diverse datasets, accelerating innovation and supporting the emergence of joint European projects.^{316, 317, 318} This includes providing secure, standardized environments for data sharing, which can lead to more sophisticated AI algorithms and comprehensive solutions for various sectors, such as energy management.³¹⁹ Data spaces can also provide the flexibility

³⁰⁹ Google. (2023, November 14). Google's EU AI Opportunity Agenda. https://static.googleusercontent.com/media/publicpolicy.google/en/resources/eu_ai_opportunity_agenda_en.pdf

³¹⁰ Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

³¹¹ Publications Office of the European Union. (2023, November 23). European data spaces and the role of data.europa.eu. <https://op.europa.eu/en/publication-detail/-/publication/70d01867-8ce1-11ee-8aa6-01aa75ed71a1/language-en>

³¹² Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

³¹³ Publications Office of the European Union. (2023, November 23). European data spaces and the role of data.europa.eu. <https://op.europa.eu/en/publication-detail/-/publication/70d01867-8ce1-11ee-8aa6-01aa75ed71a1/language-en>

³¹⁴ Publications Office of the European Union. (2023, November 23). European data spaces and the role of data.europa.eu. <https://op.europa.eu/en/publication-detail/-/publication/70d01867-8ce1-11ee-8aa6-01aa75ed71a1/language-en>

³¹⁵ Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

³¹⁶ Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

³¹⁷ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

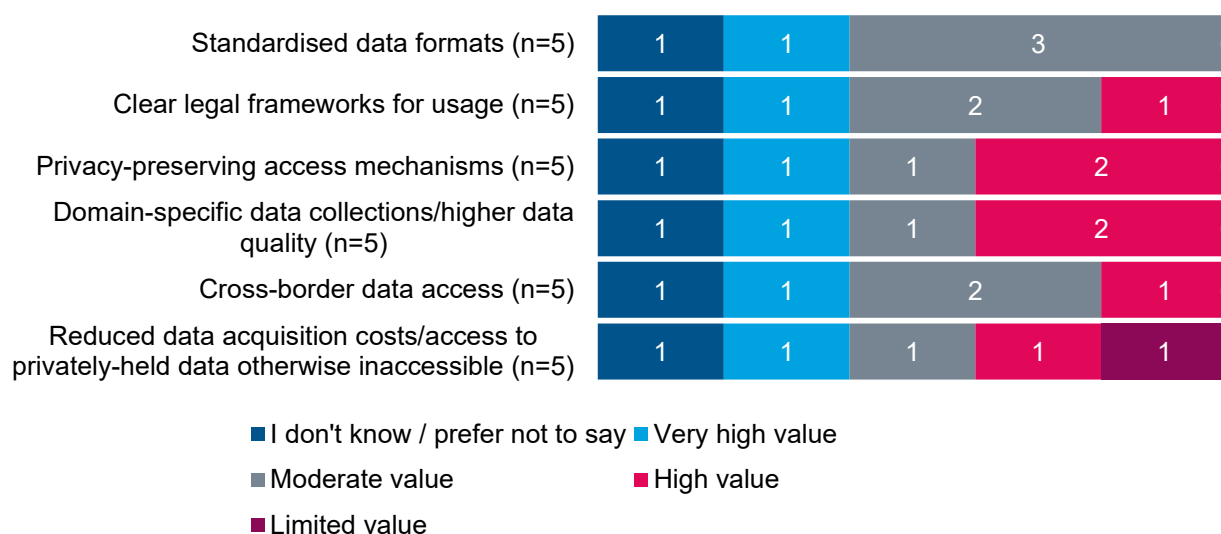
³¹⁸ Ruokonen, M. (2025). APPLYING AI IN KEY EUROPEAN INDUSTRIES.

³¹⁹ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward.

needed to prepare data to the level required by AI practitioners, complementing data with specific data models and metadata descriptions for ML/AI.³²⁰

Survey evidence offers additional insights. Respondents emphasised that **domain-specific datasets and privacy-sensitive data** are perceived as the areas where data spaces could deliver the greatest value. **Cross-border access** was also considered important, though to a lesser extent, while **standardised data formats** were rated as only moderately valuable. These findings suggest that industry expectations are focused less on general infrastructure and more on the provision of specialised, sensitive, and sector-specific resources.

Figure 7. Rate the potential value of the following Common European Data Spaces features for your AI development



Qualitative input further expands on these expectations. One respondent underlined that, to better serve generative AI companies—particularly startups—Common European Data Spaces should enable **massive-scale, high-quality data ingestion and automated curation**, with cross-sector semantic interoperability supported by harmonised ontologies and AI-optimised discovery tools. They called for **robust AI-centric governance frameworks**, including clear IP licensing, granular access controls, and mandatory data provenance, combined with less bureaucracy to streamline access and usage. The respondent also highlighted the importance of **tight integration with compute resources**, for instance through *data-near-compute* paradigms and support for real-time data streams. Finally, they stressed the need for an attractive ecosystem, including **incentives for data contribution, AI sandbox environments, and targeted mechanisms to help startups scale up**.

2.2.6 Potential solutions and policy recommendations

2.2.6.1 Data Access Solutions

To address the critical data access challenges, a multi-faceted approach is required. **Partnerships** between EU and non-EU companies can play a role in bridging data and

³²⁰ Román, D. A. (2025, February 27). AI factories and the data challenge: access, acquisition and usage of data. Connection to data spaces. BDVA. <https://bdva.eu/blog/ai-factories-and-the-data-challenge/>

compute gaps, although their immediate availability may vary.³²¹ Industry consortia, such as Catena-X which is a pioneering example of a data space developed under the GAIA-X framework in the automotive sector, can also help overcome shared, sector-wide obstacles to data sharing.³²²

Federated learning offers a privacy-preserving solution by enabling models to be trained on decentralized data (e.g., on mobile phones) without the raw data ever leaving the device, thereby addressing fundamental problems of privacy, ownership, and data locality.^{323, 324, 325} This approach enhances security, protects IP, and minimizes breach risks while allowing for real-time learning from decentralized sources.^{326, 327}

The increasing use of **synthetic data** is a promising avenue. Synthetic data can scale up smaller datasets and generate robust AI models, particularly when real data is sensitive, scarce, or biased.^{328, 329} It allows for the creation of larger, more diverse datasets for training and refining AI models across various applications, from healthcare to education.³³⁰ Synthetic data can also reduce costs involved at all stages of the data value chain by limiting the need for excessive data collection, cleaning, and storage.³³¹

Promoting **open-source** models and datasets is crucial for democratizing access to resources and fostering innovation. Efforts like EleutherAI and Hugging Face's FineWeb aim to create massive, publicly available datasets for training GenAI models.³³²

³²¹ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

³²² Ibidem

³²³ Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., ... & Roselander, J. (2019). Towards federated learning at scale: System design. *Proceedings of machine learning and systems*, 1, 374-388.

³²⁴ Publications Office of the European Union. (2023, November 23). European data spaces and the role of data.europa.eu. <https://op.europa.eu/en/publication-detail/-/publication/70d01867-8ce1-11ee-8aa6-01aa75ed71a1/language-en>

³²⁵ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. *Global Partnership on AI*. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

³²⁶ Publications Office of the European Union. (2023, November 23). European data spaces and the role of data.europa.eu. <https://op.europa.eu/en/publication-detail/-/publication/70d01867-8ce1-11ee-8aa6-01aa75ed71a1/language-en>

³²⁷ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. *Global Partnership on AI*. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

³²⁸ De Wilde, P., Arora, P., Buarque de Lima Neto, F., Chin, Y., Thinyane, M., Stinckwich, S., ... & Marwala, T. (2024). Recommendations on the Use of Synthetic Data to Train AI Models.

³²⁹ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

³³⁰ Guo, X., & Chen, Y. (2024). Generative ai for synthetic data generation: Methods, challenges and the future. *arXiv preprint arXiv:2403.04190*.

³³¹ Confederation of European Data Protection Organisations. (2023, October 26). Generative AI: The Data Protection Implications. <https://cedpo.eu/generative-ai-the-data-protection-implications/>

³³² Wiggers, K. (2024, June 1). AI training data has a price tag that only Big Tech can afford. *TechCrunch*. <https://techcrunch.com/2024/06/01/ai-training-data-has-a-price-tag-that-only-big-tech-can-afford/>

Additionally, **EU funding initiatives** like the AI Factories and the AI innovation package are designed to provide financial support and access to supercomputing resources for startups and SMEs, lowering entry barriers and accelerating the development of a European AI ecosystem.^{333, 334}

2.2.6.2 Policy measures and instruments

Effective policy measures are essential to foster a thriving GenAI ecosystem in Europe. **Public investment in datasets** is critical, particularly for high-quality, multilingual, and sector-specific data, to fill the gaps left by a fragmented private sector^{335, 336}. The U.S. Department of Commerce's guidelines on Generative AI and Open Data provide actionable recommendations for optimizing public data assets for AI use.^{337, 338} Europe must strengthen its role in catalysing wider data availability by developing Open Data for model training and interoperability and encouraging data sharing between the private and public sectors across Member States.³³⁹

Fostering the development of **open-source EU-scale models** can reduce Europe's dependency on foreign providers and ensure that models reflect European values and linguistic diversity.^{340, 341} Initiatives like the Alliance for Language Technologies EDIC are stepping in to provide centralized access to language resources for European LLMs.³⁴²

³³³ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

³³⁴ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

³³⁵ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

³³⁶ Hutchinson, P., Goll, F., & Mügge, C. (2024, February 27). Generative AI in the European Startup Landscape 2024. appliedAI Institute for Europe. <https://www.appliedai-institute.de/en/hub/2024-generative-ai-study>

³³⁷ Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, 99, 101896

³³⁸ Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024, May). From co-generated data to generative AI: New rights and governance models in digital ecosystems. *Global Partnership on AI*. <https://wp.oecd.ai/app/uploads/2024/12/From-co-generated-data-to-generative-AI-1.pdf>

³³⁹ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

³⁴⁰ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

³⁴¹ Balland, P. A., Grabova, O., Marcus, J. S., Praas, R., & Renda, A. Generative Ai and Foundation Models in the Eu: Uptake, Opportunities, Challenges, and a Way Forward

³⁴² EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

Policymakers must establish **clearer compliance pathways** for GenAI developers. This involves providing swift and clarifying guidance on the implementation of the AI Act and related legislation, and ensuring regulatory consistency in areas like data, privacy, and copyright across the EU.³⁴³ It is important that secondary legislation and codes of practice do not expand the requirements of the AI Act and instead focus on making compliance easier to achieve.³⁴⁴

Finally, **certification and incentives for data sharing** are vital. In this context, 'certification' refers to developing formal, independently verified seals or labels (such as the Europrivacy or GAIA-X labels) to prove that a service, a data set, or a data-sharing process meets all EU requirements for trust, security, and data sovereignty (including IP protection). Regulatory sandboxes and test beds can also provide safe environments for experimentation and validation of AI systems, promoting responsible innovation while ensuring compliance.^{345, 346} Institutional support, such as from national IP offices, can provide technical support for implementing federated databases of TDM reservations and raise public awareness of copyright issues.³⁴⁷

2.2.6.3 Stakeholders view on priorities and policy options

These policy recommendations identified by the literature review were also discussed and complemented by stakeholders' consultation³⁴⁸ aimed at validating their relevance and priority.

Overall policy priorities

As shown in the Figure below, survey evidence shows that **stakeholder views on policy priorities are fragmented**, making it difficult to identify a single clear direction. Nevertheless, some patterns emerge. **Solutions related to data sharing and acquisition were consistently perceived as the most urgent**, particularly measures that facilitate access to high-quality and sector-specific datasets. Direct financial support was also highlighted: **three out of five respondents considered financial incentives for data sharing to be extremely effective**, underlining the importance of public funding mechanisms to bridge current gaps. A **clearer legal framework for data acquisition** was similarly valued, with respondents emphasising the need for regulatory clarity and

³⁴³ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

³⁴⁴ ERT (European Round Table for Industry). (2025, March). Harnessing Generative AI for Europe's Industrial Future. https://ert.eu/wp-content/uploads/2025/03/ERT-Harnessing-GenAI-for-Europe-Industrial-Future_March-2025.pdf

³⁴⁵ Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). Connecting the dots in trustworthy Artificial Intelligence: From AI principles, ethics, and key requirements to responsible AI systems and regulation. *Information Fusion*, 99, 101896.

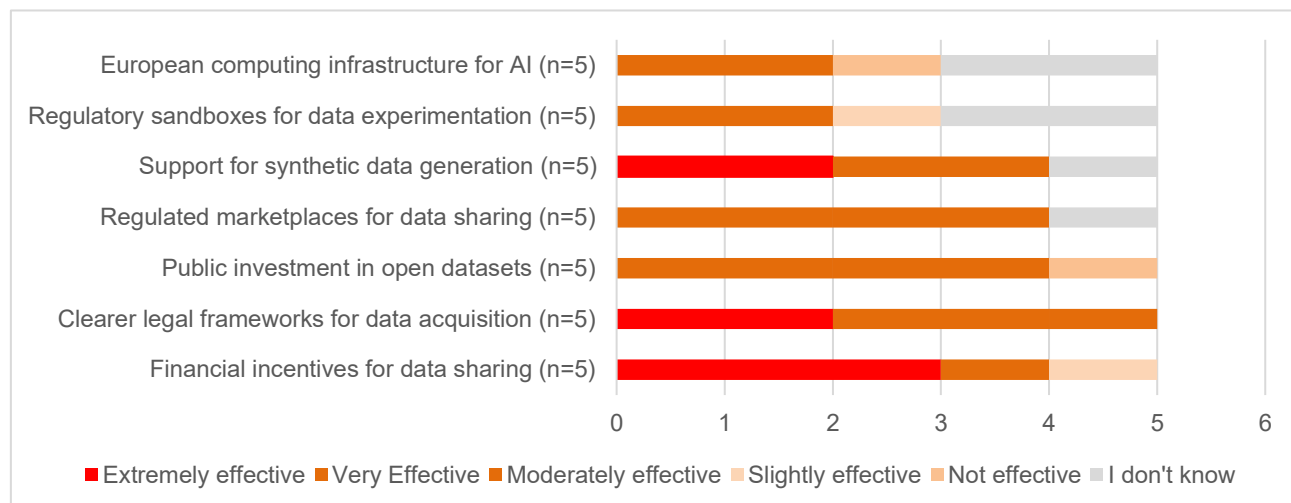
³⁴⁶ EIT Digital. (2024). Generative AI: Europe's Quest for Regulation and Industry Leadership. https://www.eitdigital.eu/fileadmin/2024/ecosystem/downloads/EIT-Digital_AiNed_Generative-AI-Report-web.pdf

³⁴⁷ University of Turin Law School & Nexa Center for Internet & Society. (2025, May 12). Development of Generative Artificial Intelligence from a Copyright Perspective. European Union Intellectual Property Office. https://euipo.europa.eu/tunnel-web/secure/webdav/guest/document_library/observatory/documents/reports/2025_GenAI_from_copyright_perspective/2025_GenAI_from_copyright_perspective_FullR_en.pdf

³⁴⁸ Policy priorities and recommendations suggested by stakeholder cannot be considered representative, given the limitations of the consultation activities explained in Section 2.1.3

simplification. At the same time, **all other proposed policy measures were considered important**, even if none commanded overwhelming consensus. This suggests that stakeholders see value across a wide range of interventions, but **establishing strong priorities remains difficult in a still-fragmented policy debate**.

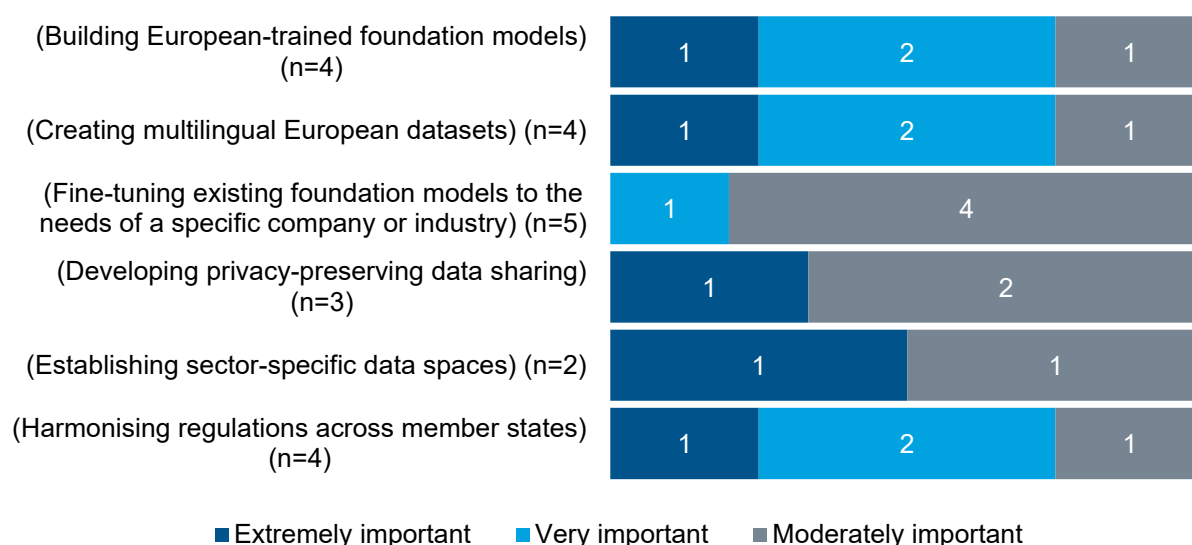
Figure 8. Survey Results Q33 - Rate the potential effectiveness of these policy measures for addressing AI data challenges



A similarly, fragmented scenario emerges when assessing key competitiveness factors according to consulted stakeholders. **Building European-trained foundation models, creating multilingual datasets, and harmonising regulations** were most frequently rated as highly important, reflecting the perceived role of sovereignty, linguistic coverage, and legal consistency in boosting Europe's competitiveness. **Fine-tuning models for sector-specific needs** was seen as relevant but mostly rated as moderately important, while **privacy-preserving data sharing** and **sector-specific data spaces** were mentioned less often, though still recognised as part of the competitiveness mix.

Overall, the findings suggest that stakeholders see **a broad portfolio of factors contributing to competitiveness**, with particular emphasis on models, data, and regulatory harmonisation, but without consensus on a single decisive priority.

Figure 9. Survey Results Q34 - Rate the importance of policy priorities



Qualitative feedback from the survey provides additional perspective on how policy and governance frameworks could better support GenAI development in Europe. One respondent stressed that to effectively support startups, data spaces and related initiatives **must evolve beyond current designs to enable large-scale, high-quality data ingestion and ensure semantic interoperability across sectors**. They called for streamlined governance with less bureaucracy, alongside targeted support for startups, such as secure environments for AI experimentation. Above all, the respondent highlighted the need for regulatory changes, including clearer copyright rules for AI training, adaptation of GDPR principles, stronger data governance under the AI and Data Acts with clear accountability frameworks, and greater use of regulatory sandboxes to enable responsible access to data

Regulatory balance between privacy, copyright and innovation

Balancing data accessibility for innovation with the need to protect privacy remains one of the most sensitive and contested issues in the European GenAI landscape. **Stakeholders expressed a wide range of views, with no clear consensus emerging**. Some emphasised the need for stronger safeguards, while others argued for greater flexibility to enable innovation. What does emerge consistently is the **recognition that privacy must remain a priority**. Several stakeholders noted that technologies such as **synthetic data now make it possible to protect personal information** while still delivering usable training datasets. Europe's stronger focus on privacy, compared with North America, was seen by some as a **potential competitive advantage**, enabling the continent to differentiate its AI ecosystem while preserving public trust.

Expert interviews reinforced this perspective by reframing **GDPR not as a hindrance, but as an asset**. Misinterpretation of the law was described as a common problem, leading companies to adopt overly cautious compliance strategies. In reality, proper engagement with GDPR can foster high-quality, sustainable, and responsible data ecosystems, creating the conditions for more reliable knowledge production and long-term innovation. The short-term burdens of compliance were contrasted with the long-term benefits of robust data governance, suggesting that Europe's legal framework may ultimately strengthen its AI competitiveness if applied strategically.

When it comes to copyright, opinions were less divided. **Stakeholders generally leaned towards approaches that favour innovation over a strict interpretation of copyright rules**. While there was recognition of the importance of protecting rights-holders, the prevailing view was that uncertainty and fragmentation across Member States currently act as barriers, particularly for smaller firms. What companies most need are clearer frameworks, standardised licensing models, and practical tools—such as transparent registries for opt-outs—that make it easier to access copyrighted data for training while respecting creators' rights.

This inherent privacy-utility trade-off is a fundamental limitation that synthetic data does not resolve, despite prevailing narratives in the commercial and research domains. As one expert noted in a recent interview, *"Synthetic data should primarily be used for testing and development, where only the structure and format are needed, not the full informational content. For tasks requiring statistical fidelity, privacy regulations should still apply."* This perspective reinforces the view that synthetic data is best suited for environments where representational accuracy is secondary to format compatibility, and where privacy concerns can be mitigated by limiting informational depth.

A final remark from another interviewee highlighted that – if properly leveraged - **GDPR serves as an asset for innovation**, encouraging robust data governance and reliable knowledge production essential for long-term success and sustainability.

2.3 Methodological approach: Task 1b – Examination of the maturity and offering of data synthetisation solutions

The methodological approach for Task 1b consists of two main components. First, desk research was carried out to develop a comprehensive mapping of synthetic data vendors operating in the EEA. This included collecting both basic company details, such as year of establishment, and more specific information on the types of synthetic data produced and their underlying sources. Second, online consultations—through a survey and targeted interviews—were conducted to validate the mapping and gather additional insights not available in the public domain. These complementary methods resulted in the creation of detailed company factsheets, presenting the most relevant information on synthetic data providers in a structured and accessible format.

2.3.1 Methodological approach: mapping exercise

The mapping of synthetic data providers was conducted using a structured and policy-relevant desk research methodology, designed to support the broader objectives of the study on the next data frontier. The aim was to present the current landscape of offering synthetic data solutions as a commercial offer.

Scoping and inclusion criteria

Providers were selected based on their active engagement in synthetic data generation, their relevance to key (and often highly regulated) sectors such as healthcare, finance, public administration, and AI development, and their alignment with privacy-enhancing technologies (PETs) and GDPR compliance. The mapping focused on vendors offering synthetic data solutions either independently or in augmentation of real-world datasets, with particular attention to whether the data involved personal or commercially sensitive information. This ensured that the analysis captured both the technical diversity and regulatory implications of synthetic data practices across the EEA.

Data collection and source strategy

The data collection strategy was two-fold. Firstly, systematic desk research, employing targeted keyword searches and structured review of publicly available materials. Boolean operators and domain-specific terminology were used to refine search queries, including terms such as “synthetic data platform GDPR”, “privacy-preserving analytics healthcare”, and “generative models for data synthesis”. Sources included vendor websites, technical documentation, product brochures, white papers, press releases, and publicly available regulatory filings.

Secondly, company level data collection aimed at gaining specific information was gathered through online consultation activities in the form of a survey and online interviews.

Vendor profiling and comparative analysis

Each vendor was profiled using a standardised template that captured a comprehensive set of attributes relevant to synthetic data generation and privacy compliance. These included the vendor’s geographic location, year of establishment, core synthetic data methodologies (e.g. GANs, VAEs, rule-based synthesis), domains of activity, types of services offered, and the nature of their data dependencies. Particular attention was paid to whether the provider could generate synthetic data independently or whether their platforms required input from real-world datasets for augmentation. The profiles also documented the vendors’ commercial maturity, including indicators such as product availability, client base, and integration with machine learning of analytics pipelines. GDPR compliance was assessed based on publicly stated practices, including the use of

anonymisation techniques, differential privacy, privacy-by-design engineering, and alignment with European and international data protection frameworks.

The mapping also included illustrative examples of use cases, detailing the source data (where applicable), the parties involved, and the specific problems addressed through synthetic data. Where available, indicators on customer adoptions and application volume were included to assess the commercial maturity and scalability of the solutions.

Validation and quality assurance

To ensure the credibility and reliability of the mapping, a rigorous validation and quality assurance process was implemented. All collected information was manually reviewed to confirm factual accuracy and contextual relevance. This process helped to identify inconsistencies, clarify ambiguous statements, and ensure that the final profiles reflected a balanced and evidence-based assessment of each provider's capabilities. The validation phase also contributed to the overall integrity of the mapping exercise by reinforcing transparency and methodological consistency.

Characteristics description

Each row in the table corresponds to a distinct vendor and captures a set of standardised attributes:

- **Vendor name and location:** identifies the company and its country of origin, providing geographic context relevant to regulatory alignment and market positioning.
- **Creation of synthetic data:** specifies the underlying methodology used to generate synthetic data, distinguishing between generative models (e.g. GANs, VAEs), rule-based synthesis, and hybrid approaches. This classification is essential for understanding the technical foundations and limitations of each solution.
- **Domains of activity:** lists the sectors in which the vendor's synthetic data solutions are applied, such as healthcare, finance, telecommunications, retail, and public administration. This helps assess the breadth of applicability and relevance to EU policy priorities.
- **Type of services:** describes the nature of the offering, including platform capabilities (e.g. SaaS) integration with machine learning pipelines, support for privacy risk analysis, and tools for secure data collaboration. This dimension reflects the operational model and user-facing functionalities.
- **Data source:** indicates whether the vendor's solution requires access to real-world datasets, supports augmentation, or enables independent data synthesis. This is a critical factor in evaluating the legal and ethical implications of data use, particularly in relation to personal and commercially sensitive information.
- **Commercial maturity:** provides the year of establishment or market entry, offering insight into the vendor's expertise, stability, and potential scalability. Where available, this is complemented by indicators of customer adoption and application volume.
- **GDPR compliance:** summarises the vendor's approach to data protection, including the use of differential privacy, anonymisation frameworks, privacy-by-design engineering, and alignment with European and international regulatory standards. This dimension is essential for assessing the legal robustness and trustworthiness of the synthetic data solution.

Limitations

While the aggregated mapping provides a structured and comparative overview of synthetic data providers, several limitations should be acknowledged to contextualise the findings and inform future research and policy development.

Firstly, the mapping relies **exclusively on publicly available information**, including vendor websites, technical documentation, press releases, and open-access reports. As such, it may not capture proprietary methodologies, internal compliance mechanisms, or unpublished use cases that could significantly influence the assessment of a provider's maturity or regulatory alignment. Many vendors operate in competitive or sensitive domains such as healthcare, finance, and defence, where disclosure of technical details or client applications is limited due to confidentiality agreements or commercial sensitivities.

Secondly, **the availability and granularity of data varied considerably across providers**. While some vendors offer detailed descriptions of their synthetic data generation techniques and privacy safeguards, others provide only high-level summaries or marketing-orientated narratives. This unevenness in transparency may affect the comparability of profiles and limit the ability to assess the robustness of GDPR compliance or the scalability of their solutions.

Additionally, the mapping does not account **for dynamic developments in the synthetic data landscape**, including emerging vendors, evolving regulatory interpretations, or technological innovations such as foundation models and multimodal synthesis. Given the pace of change in the field, periodic updates to the mapping would be necessary to maintain its relevance and accuracy.

Despite these constraints, the mapping offers a valuable starting point for understanding the synthetic data provider ecosystem and identifying areas for further investigation, policy support, and market development.

2.3.1.1 Consultation activities

Survey to synthetic data providers

To complement the desk research and vendor mapping presented in Section 2.4.1.1, a targeted stakeholder survey was designed, launched, and completed. The objective of this consultation was to gather primary evidence from synthetic data providers operating within the EEA and UK, focusing on the development, deployment, and governance of their synthetisation solutions. The survey was intended to address key gaps identified during the mapping phase, particularly in relation to commercial maturity, real-world applications, usage metrics, and regulatory compliance practices.

The survey specifically targeted the vendors identified in the aggregated mapping presented above. Each of the 15 mapped companies was contacted directly, with multiple contact points per organisation. In addition, the survey was made accessible to other relevant providers through a snowball sampling approach, encouraging recipients to share the survey within their networks to broaden outreach.

The survey was structured across ten thematic sections, beginning with basic company information and progressing through technical practices, commercial maturity, and compliance approaches. The structure is summarised below:

Table 8. Survey to synthetic data providers (Task 1b) summarised structure

Theme	Content summary
Company profile	Basic information such as company name, sector, geography, and

	year of establishment.
Technical methods	Methods used to generate synthetic data, including use of real-world data and augmentation techniques.
Sectors of application	Identification of sectors where synthetic data methods are applied.
Use case scenarios	Examples of actual or potential implementation scenarios, including source data, involved parties, and problems addressed.
User statistics	Data on the number of customers and applications using the company's solutions.
Commercial maturity	Assessment on the maturity level of the company's synthetic data solutions (e.g. early stage, growth, established, advanced).
GDPR compliance and anonymisation	Techniques used to ensure GDPR compliance and data anonymisation, including risk models and challenges in non-EU jurisdictions.
Future-proofing	Steps taken or planned to ensure the long-term viability and scalability of synthetic data solutions.
Policy recommendations	Suggestions for EU and national regulatory action to support the synthetic data ecosystem.
Comparative legal insights	Reflections on advantages or challenges in data sharing legislation between the EU and selected non-EU countries.

The survey included both closed- and open-ended questions, with routing logic to tailor the experience based on the respondent's operational scope. The survey was designed to elicit detailed responses on topics not easily accessible through desk research, such as anonymisation risk models, future milestones, and policy needs.

The survey was launched on 20 May 2025 via the online platform Forsta and distributed to 64 targeted contacts. Reminder emails were issued on 3 June, 16 June, the latter informing recipients of a two-week extension. The survey closed on 11 July 2025.

Six completed responses have been received from synthetic data vendors. Whilst the response rate remains modest, the data collected provides valuable qualitative and quantitative insights into the operational realities of synthetic data providers. To further enrich the evidence base and address residual gaps, follow-up interviews with selected experts and providers have been conducted, the methodology of which is described in below.

Table 9. Overview of survey responses

Vendor location	Vendor market
France	EU/EEA-wide
Portugal	International
Sweden	Nation-wide
Finland	EU/EEA-wide
France	EU/EEA-wide
Italy	EU/EEA-wide

Interviews with synthetic data providers

Interview outreach began on 10 July 2025, targeting vendors identified during the mapping and survey phases. A follow-up reminder was issued on 7 August 2025 to encourage participation and confirm availability. By the time of submission, three interviews have been successfully conducted with representatives from synthetic data providers.

A baseline topic guide was developed during the inception phase and refined in response to survey findings. This guide was adapted for each interviewee to reflect their specific technologies, business models, and regulatory contexts. The structure ensured consistency across interviews while allowing flexibility to explore stakeholder-specific insights. The topic guide included a combination of open-ended questions, follow-up prompts, and clarification probes to encourage detailed responses and allow interviewers to explore emerging themes in real time. In line with EU requirements, participants received relevant privacy notices and were assured that their responses would be treated confidentially and used solely for the purposes of this study.

Insights from interviews were crucial to the creation of the company factsheets presented below. They have been designed to be simple 2-page documents outlining the key points of query that are outlined in Table 10 below.

Table 10. Overview of the interview programme (Task 1b)

Synthetic data provider	Topics discussed in the interview
AINDO	• Technical methods and innovation • Use cases and market demand • GDPR compliance and legal positioning • Strategic role and market maturity • Policy and ecosystem recommendations
YData	
Octopize	

Limitations of consultation activities

While the survey and interviews provided valuable insights, two overarching limitations should be acknowledged to contextualise the findings.

Low participation

Despite targeted outreach and follow-up, overall participation was modest. This reflects both the relatively small size of the synthetic data vendor community in the EEA and the challenges of engaging commercial actors in research exercises. The limited sample reduces the generalisability of the results and means that some perspectives or practices may not be fully captured.

Confidentiality and sensitivity

Several topics—particularly those involving business models, client data, and compliance practices—were considered commercially sensitive. As a result, respondents often provided high-level answers or declined to share detailed figures. This constraint is common in studies involving emerging technologies and should be taken into account when interpreting the evidence base.

These limitations do not undermine the value of the consultation activities but indicate that certain findings should be treated as indicative rather than exhaustive.

2.4 Findings: Task 1b – Examination of the maturity and offering of data synthetisation solutions

This section presents the findings for Task 1b, focusing on the mapping and analysis of synthetic data vendors in the EEA. It draws on evidence collected through desk research and targeted consultations, presented in a structured and integrated manner. The following subsections are organised as follows:

- **Section 2.4.1** provides a factual account of the data collection process. It begins by outlining the conceptual framework underpinning the mapping exercise, summarises the overall results, and introduces the company factsheets developed for selected vendors.
- **Section 2.4.2** adopts a more analytical perspective, interpreting the evidence to describe the synthetic data landscape across key dimensions, including creation techniques, commercial applications and maturity, and privacy assurance patterns.
- **Section 2.4.3** looks ahead, presenting the outlook for the sector and outlining policy recommendations informed by the preceding analysis.

2.4.1 The European landscape of synthetic data as commercial offer

This subsection presents the factual outputs of the data collection activities carried out under Task 1b. It begins by outlining the conceptual framework used to map synthetic data vendors in the EEA, clarifying the criteria applied for inclusion and classification. It then provides an overview of the mapping results, highlighting the distribution of vendors across countries, their core business characteristics, and the types of synthetic data offered. Finally, it introduces the company factsheets developed for selected vendors, which consolidate key information on technical approaches, application domains, and compliance practices.

2.4.1.1 Conceptual framework

This section presents the conceptual framework that guided the mapping of synthetic data providers across the EEA and selected third countries. The framework was developed through targeted desk research and is grounded in the definitions and variables outlined in the ToR. It serves as the analytical foundation for identifying, categorising, and comparing providers based on their technical approaches, data dependencies, and compliance strategies.

Understanding synthetic data

Synthetic data refers to artificially generated information that replicates the statistical properties and structure of real-world datasets without necessarily containing any actual records. According to IBM, synthetic data is “artificial data designed to mimic real-world data,” generated through statistical methods or AI techniques such as deep learning and generative models³⁴⁹. It can be used to supplement or replace real datasets, particularly in contexts where access to high-quality, representative, or privacy-compliant data is limited.

Synthetic data is increasingly used across a range of sectors, including healthcare, finance, and public administration, to support activities such as model training, system testing, and secure data sharing. Its appeal lies in its potential to provide high-utility datasets that are free from the legal and ethical considerations associated with personal

³⁴⁹ IBM. What Is Synthetic Data? | IBM. <https://www.ibm.com/think/topics/synthetic-data>

or commercially sensitive information. As noted by the Royal Society, synthetic data can enhance privacy, fairness, and data augmentation, while also accelerating innovation³⁵⁰.

From a policy and regulatory perspective, synthetic data is increasingly recognised as a strategic enabler for privacy-preserving innovation, regulatory compliance, and international data governance. According to the EDPS, synthetic data can enhance privacy through data protection by design, allowing individuals' personal data to remain undisclosed while still enabling meaningful analysis and model training³⁵¹. It also holds promise for improving fairness in AI systems by mitigating bias and enabling the creation of more representative datasets.

Types of synthetic data generation

To capture the diversity of approaches and implications, the conceptual framework distinguishes between two primary modes of synthetic data generation. These categories reflect both the technical foundations, and the regulatory considerations associated with different methods of data synthesis.

Creation of synthetic data without the use of specific real-world data assets

This category refers to synthetic data that is generated independently of any identifiable real-world dataset. The data is provided using generalised models, rule-based systems, or simulated environments that do not rely on actual personal or commercial records. According to a systematic review by Goyal and Mahmoud (2024), such data is often generated using generative adversarial networks (GANs), variational autoencoders (VAEs), or large language models (LLMs) trained on abstracted or publicly available data³⁵².

Common techniques include:

- **Simulation-based generation**, where synthetic data is created from predefined rules or statistical assumptions.
- **Generative models**, such as GANs and VAEs, which learn generalised data distributions and produce new instances that reflect plausible patterns without referencing specific source data.

This approach is particularly relevant in:

- **Early-stage development**, where synthetic data enables testing without risking data breaches.
- **Regulatory sandboxes**, where innovations can be trialled under controlled conditions, especially in markets where regulation is prevalent such as the EEA.
- **Education and training**, where realistic but non-sensitive datasets are needed.

The absence of real-world data inputs significantly reduces privacy risks and simplifies compliance. However, the utility of such data depends on the sophistication of the generation model and its ability to reflect realistic distributions. As highlighted in

³⁵⁰ Jordon, J., Houssiau, F., Cherubin, G., Cohen, S. N., Szpruch, L., Bottarelli, M., Maple, C., & Weller, A. (2022). Synthetic Data - what, why and how? Royal Society. https://royalsociety.org/-/media/policy/projects/privacy-enhancing-technologies/Synthetic_Data_Survey-24.pdf

³⁵¹ Synthetic Data. European Data Protection Supervisor. https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en

³⁵² Goyal, M., & Mahmoud, Q. H. (2024). A Systematic Review of Synthetic Data Generation Techniques Using Generative AI. *Electronics*, 13(17), 3509. <https://doi.org/10.3390/electronics13173509>

Nikolenko's (2021) review of synthetic data for deep learning, the realism and domain relevance of fully synthetic datasets can vary widely depending on the model architecture and training strategy³⁵³.

Creation of synthetic data using real-world data assets

This category encompasses synthetic data generation methods that rely on actual datasets as inputs. These datasets may contain personal or commercially sensitive information and are used to train models that produce synthetic versions of the original data. Both the EDPS³⁵⁴ and academic literature³⁵⁵ agree that synthetic data generated from real-world datasets should undergo a privacy assurance assessment to ensure that the output does not constitute personal data under the GDPR.

Three subtypes of synthetic data using real-world data assets are distinguished:

Real-world data assets containing personal data: these include datasets with personally identifiable information, such as health records or financial transactions. Synthetic data generated from such sources must comply with the GDPR, particularly regarding anonymisation, re-identification risk, and data minimisation. A privacy assurance assessment is essential to evaluate whether synthetic outputs could inadvertently reveal information about data subjects.

Real-world data assets containing commercial data: these datasets contain proprietary business information, trade secrets, or internal metrics. While not always subject to GDPR, they raise concerns around confidentiality and intellectual property. In this context, synthetic data must ensure that commercially sensitive information—such as proprietary business processes or confidential performance metrics—is not exposed, while still retaining enough structure and utility to support meaningful analysis or model development.

Augmenting real-world data assets with synthetic data elements: in this hybrid approach, synthetic data is used to enrich or expand existing datasets. This may involve generating additional records to balance class distributions, simulate rare events, or fill gaps in incomplete data. While this enhances analytical robustness, it introduces complexity in terms of data provenance and compliance, especially when personal data is involved.

These methods are more common in applied settings where statistical fidelity and domain specificity are critical. They enable high-utility datasets for machine learning, policy modelling, and operational analytics, but require careful validation and documentation to ensure legal and ethical integrity. As noted in SpringerLink's review of tabular and latent space generation, hybrid approaches are increasingly used to improve model generalisation and reduce bias in training data³⁵⁶.

Overview of GDPR compliance techniques

³⁵³ Nikolenko, S. I. (2021). *Synthetic Data for Deep Learning*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-75178-4>

³⁵⁴ Synthetic Data. European Data Protection Supervisor. https://www.edps.europa.eu/press-publications/publications/techsonar/synthetic-data_en

³⁵⁵ Kurapati, S., & Gilli, L. (2023). Synthetic data: A convergence between Innovation and GDPR. *Journal of Open Access to Law*, 11(2), 12. <https://doi.org/10.63567/gk6xj346>

³⁵⁶ Fonseca, J., & Bacao, F. (2023). Tabular and latent space synthetic data generation: a literature review. *Journal of Big Data*, 10(1). <https://doi.org/10.1186/s40537-023-00792-7>

Synthetic data is often presented as a privacy-preserving alternative to real-world datasets, particularly in contexts where personal data is subject to strict regulatory controls. However, its use does not automatically guarantee compliance with the GDPR. To ensure that synthetic solutions align with GDPR requirements, providers typically implement a combination of technical and organisational safeguards. The most common techniques include:

- **Anonymisation:** under GDPR Recital 26, anonymised data falls outside the scope of the regulation. However, achieving true anonymisation is technically complex and requires careful consideration of all means likely to be used for identification. Academic literature highlights that many so-called anonymisation techniques reduce identifiability but do not yield fully anonymous data³⁵⁷.
- **Pseudonymisation:** defined in Article 4(5) of the GDPR and is recognised as a technical and organisational measure that supports data protection by design. It involves replacing identifiers with artificial tokens, which can be reversed if a key is retained. While pseudonymised data remains within the scope of GDPR, it can reduce privacy risks when properly implemented³⁵⁸.
- **Privacy-by-design engineering:** GDPR Article 25 mandates that data protection be embedded into the design of systems and processes. For synthetic data, this includes secure model training environments, access controls, audit trails, and documentation of privacy risk assessments. The JRC notes that privacy-by-design principles are essential for trustworthy synthetic data applications in policy and research contexts³⁵⁹.
- **Transparency and documentation:** Transparency is a core principle of the GDPR, as outlined in Articles 5(1)(a) and 12, requiring that any information relating to the processing of personal data be provided in a concise, transparent, intelligible, and easily accessible form, using clear and plain language³⁶⁰. In the context of synthetic data—particularly when derived from real-world datasets—this includes publishing privacy notices, technical documentation, and results of privacy assurance assessments.

³⁵⁷ Fischer-Hübner, S., Hansen, M., Hoepman, J. H., & Jensen, M. (2023). Privacy-Enhancing Technologies

and Anonymisation in Light of GDPR and Machine Learning. In F. Bieker, J. Meyer, S. Pape, I. Schiering, &

A. Weich (Eds.), Privacy and Identity Management: 17th IFIP WG 9.2, 9.6/11.7, 11.6/SIG 9.2.2

International Summer School, Privacy and Identity 2022, Proceedings (pp. 11-20). (IFIP Advances in

Information and Communication Technology; Vol. 671 IFIP). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-031-31971-6_

³⁵⁸ Guidelines 01/2025 on Pseudonymisation | European Data Protection Board. EDPB | European Data Protection Board. https://www.edpb.europa.eu/our-work-tools/documents/public-consultations/2025/guidelines-012025-pseudonymisation_en

³⁵⁹ Hradec, J., Craglia, M., Di Leo, M., De Nigris, S., Ostlaender, N. and Nicholson, N., Multipurpose synthetic population for policy applications, EUR 31116 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-53478-5, doi:10.2760/50072, JRC128595.

³⁶⁰ Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). Available at: <https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng>

2.4.1.2 Mapping of current landscape of synthetic data providers in Europe

Table 11 below provides a consolidated, like-for-like snapshot of European providers of synthetic-data solutions and adjacent privacy-enhancing technologies (PETs). It is intended as a neutral factual baseline to support the analysis presented later in the report – namely, the assessment of technical approaches, market maturity and privacy safeguards across the ecosystem.

The categories presented in the table below correspond to the coding framework introduced earlier in this section (see Conceptual Framework). Category A refers to vendors producing fully synthetic data, while Categories B1, B2 and B3 capture different approaches to generating synthetic data from real-world assets, as described above.

Table 11. Types of synthetic data coding for analytical purposes

Category	Description
A	Creation without the use of specific real-world data assets (e.g. physics-based simulation or models trained without identifiable proprietary datasets)
B	Creation using real-world data assets
	B1: personal-data inputs (e.g. health or financial records)
	B2: commercial or proprietary business data
	B3: augmentation of existing datasets (rebalancing, rare-event simulation, or gap-filling)

Codes (A, B1–B3) are applied solely to this variable because it is central to understanding risk profiles and compliance implications, and it provides a consistent reference point for comparing vendors across the mapping. This typology is used solely to aid comparability; it does not pre-judge legal qualification of outputs, which depends on privacy assurance and context-specific risk assessment.

Table 12. Summarised vendor mapping (full mapping can be found in Annex 2)

Name	Location	Creation of synthetic data	Synthetic generation category
MOSTLY AI	Vienna, AT	Generative models for tabular/text; optional Differential Privacy (DP) training; privacy-by-design safeguards (rare-category & extreme-value protection, overfitting prevention)	B (using real-world data): B1/B2 with B3 support
Sarus	Paris, FR	Privacy layer for analytics/ML on non-visible data; DP-protected outputs; DP-trained synthetic data; automatic privacy-first query rewriting	B (using real-world data): B1/B2; B3 supported
Syndata	Stockholm, SE	Generative AI (single & multi-table), presents 'Random'/'Realistic' with evaluation dashboard	B (using real-world data): B1/B2 with B3
Syntheticus	Zurich, CH	Generative AI across tabular/text/image/multimodal; Differential Privacy & Confidential Computing options	B (using real-world data): B1/B2 and B3

YData	Lisbon, PT	Generative models + anonymisation & PII detection; rebalancing/augmentation in Fabric	B (using real-world data): B1 (patient) with B2/B3 as applicable
AETION	Barcelona, ES (EU tech hub) / US HQ	De-identification + synthetic data generation within Aetion Generate; AI-powered privacy risk assessment	B (using real-world data): B1/B2 with B3
AINDO	Trieste, IT	Generative AI for fully synthetic tabular enterprise data; bias removal/compatibilization/completion	B (using real-world data): B1/B2 (anonymized outputs; limited B3)
TRUATA	Dublin, IRE	Anonymisation & privacy-enhanced analytics (clear rooms); not primarily generative SD vendor.	B (using real-world data): B1/B2 (anonymized outputs; limited B3)
OCTOPIZE	Nantes, FR	Avatar method: anonymous synthetic ‘avatars’ aligned to EDPB anonymisation criteria; peer-reviewed,	B (using real-world data): B1 (health) and B2; supports B3
Brighter AI	Berlin, DE	Deep Natural Anonymisation (synthetic overlays) + Precision Blur for faces/plates; full-body blur option	B (using real-world data): B1 (personal in images); augmentation
Sherpa.ai	Erandio, ES	Federated Learning with Secure Aggregation and Differential Privacy options; does not create datasets absent real inputs.	B (using real-world data): training on B1/B2 data in situ; no SD product
Anyverse	Madrid, ES	Fully simulated, physics-accurate synthetic imagery (RGB/LiDAR/Radar) with pixel-level ground truth; no real personal data.	A (without real-world data): simulation-based
Oblivious	Dublin, IE	Privacy infrastructure (Differential Privacy analytics + Confidential Computing);	B (using real-world data): PETs to analyze B1/B2 in situ; no SD product

2.4.1.3 Company factsheets

The following company factsheets provide a concise and structured overview of interviewed synthetic data vendors operating within the EEA. Each factsheet is designed to present key information in a comparable format, supporting the analysis of the maturity, technical approaches, and compliance practices across the synthetic data landscape. The factsheets summarise core attributes such as vendor location, data generation methods, application domains, service offerings, data dependencies, commercial maturity, and GDPR compliance measures. Where available, illustrative use cases are included to highlight real-world applications.

Figure 10. Octopize company factsheet

Overview Legal name: Octopize (SAS). HQ & founding: Nantes, France – founded 2018. Geographic footprint: EU/EEA-wide operations with projects reported in multiple Member States; selected international engagements (e.g., Canada) as described by the company. Public-sector and private-sector customers include hospitals/universities, government bodies, and large enterprises in telecoms and mobility. Mission and value proposition: Deliver anonymous synthetic data for structured/tabular datasets through the Avatar method, enabling sharing and secondary uses that meet GDPR anonymisation criteria, with transparent, explainable steps and documented privacy metrics. Ownership/funding snapshot: Not publicly disclosed in our consultation materials; with EU-funded participation in a consortium on responsible AI. At a glance: • Mode: Real-data-based generation (B) with focus on B1 (personal data) and B2 (commercial/proprietary), plus support for B3 (augmentation). • Primary sectors: Healthcare and public statistics; expansion into telecoms, mobility/transport, defence, and insurance. • Maturity (self-assessed): advanced. • Engagement model: Software + services; flexible deployment (SaaS/on-prem/VPC) aligned to controller governance. • Method highlight: Avatar (patented; peer-reviewed); vendor-published privacy metrics.	 OCTOPIZE MIMETHIK DATA Core offering and technology Primary products & capabilities. Octopize provides software and services to create anonymous synthetic “avatars” from structured/tabular datasets. The Avatar method is positioned as mathematically explainable, using dimensional projections and neighbourhood sampling to produce datasets intended as non-personal under GDPR for sharing and reuse. Features include parameter controls, per-run reports with privacy/utility indicators, and flexible deployment (SaaS or customer-controlled). Generation approach and taxonomy alignment. Octopize operates under Mode B (using real-world data assets). Avatar outputs are designed to qualify as anonymous data when generated within prescribed settings and validated through its metrics suite. The method supports B3 augmentation (e.g., scenario expansion) once privacy requirements are met. Privacy-by-design features. The pipeline applies dimensional projections to capture correlations, then K-neighbourhood sampling to generate new points while re-centring outliers that may create re-identification risk. Octopize references DP-MST (a differential-privacy-inspired component) and a published privacy-metrics catalogue. The company stresses explainability—documented, stepwise transformations rather than opaque model fitting. Deployment models and integration. Solutions are delivered on-premises, in private cloud/VPC, or as SaaS, with emphasis on customer-controlled execution for sensitive datasets. Interfaces include UI and APIs; each run produces a reproducible artefact pack (parameters, metrics, plots) for governance files (DPIA, audits).
Domains of activity and illustrative use cases Primary sectors served. Octopize has a strong healthcare and public-statistics profile (hospitals, research institutions, public bodies). It reports expansion into telecoms (customer analytics and data sharing), mobility/transport and defence (threat/scenario modelling), and insurance/financial services where privacy-by-design sharing is required. Contracts have been delivered in Luxembourg and Belgium; additional projects are cited in Canada. Healthcare and public research • Secondary research / open data: creation of anonymous synthetic datasets from clinical or administrative records to enable publication and cross-partner collaboration, while keeping source tables protected. • Clinical AI experimentation: privacy-preserving training/validation datasets for exploratory modelling or feasibility studies, with Avatar outputs taking the place of directly identifiable records. Telecoms, mobility/defence, insurance • Telco analytics and data sharing: anonymous synthetic datasets to support multi-party analytics (e.g., customer journey insights) without releasing row-level originals. • Mobility & defence: scenario generation where rare or sensitive events are difficult to share; policy emphasis on explainability and documented anonymisation.	



Privacy assurance and compliance

Evaluation practices (as reported)

- **Published metrics and methodology.** Octopize provides a privacy-metrics catalogue tied to the Avatar method, covering nearest-neighbour distances, outlier handling, and tests consistent with GDPR anonymisation criteria; documentation is publicly accessible to improve transparency and reproducibility.
- **Attack-aware posture.** The vendor references DP-MST and supports integration of external privacy-metrics libraries so controllers can run additional probes (e.g., membership-inference-style checks) where appropriate. Per-run reports bundle utility and privacy indicators to evidence that outputs meet the configured privacy bar.
- **Process controls.** Outputs can be gated by pass/fail thresholds; rare-category suppression and extreme-value re-centring are used to mitigate re-identification vectors. The firm encourages controller-side execution so governance (access, logging, DPIAs) remains under the customer's policies.

Legal positioning & supervisory signals

- **Anonymisation claim.** Octopize asserts that Avatar outputs are anonymous (GDPR sense) when generated and validated per method documentation; this position is supported by peer-reviewed material and a history of French DPA-related expertise/evaluation cited by the company.
- **Regulatory expectations.** In the survey and interview, the firm asked EU institutions to enforce GDPR consistently, clarify anonymisation criteria, and recognise transparent metrics used by controllers, noting that such recognition would reduce duplicated assessment effort across projects.

Commercial maturity and market signals

Self-assessed maturity. Advanced.

Adoption signals. The company reports delivery to public administrations, health institutions, and large enterprises across several Member States; lifecycle length remains long where buyer privacy culture is still maturing. International traction includes Canada and selected non-EU projects, while EU activity in telecoms and mobility is noted as growing.

Engagement model. Although software can be provided off-the-shelf, most enterprise deployments pair the product with configuration, evaluation runs, and documentation packaging to meet governance thresholds. Controllers often request evidence packs per generation—parameters, visuals, privacy metrics—so they can append them to DPIAs and audit trails.

Differentiators (as positioned):

- Explainable anonymisation pipeline (Avatar) for structured/tabular data with published metrics.
- Anonymous outputs designed for sharing and publication where justified.
- Flexible deployment (SaaS/on-prem/VPC) with emphasis on controller-side execution.
- Metrics openness (ability to plug in external privacy-metrics libraries).

Strategic outlook and challenges

Emerging opportunities. The firm sees continued growth in healthcare (secondary research, statistics, clinical-AI exploration) and public-sector statistics, with increasing telco and mobility/defence interest where cross-partner analytics and scenario generation are needed. It anticipates further demand for anonymous data publications and controlled data-sharing in EU projects, provided that anonymisation criteria and evidence expectations are consistently understood by buyers and authorities.

Operational challenges.

- Awareness and culture gaps—some sectors still treat GDPR as a minimal checkbox rather than a privacy-engineering discipline, elongating sales cycles.
- Certification and cost burdens—formal schemes help in procurement but can be resource-intensive for SMEs; recognition of metrics-based evidence can mitigate this.
- Market misconceptions—confusion between anonymous synthetic data and generic “fake data” can derail scoping unless explainability and method documentation are provided early.

Methodological stance. Octopize's interview position stresses explainability and transparency in anonymisation, advocating documentation and metric-driven validation; the firm supports broader use of standard privacy-metrics across tools to reduce fragmentation and accelerate confident adoption.

Figure 11. Data company factsheet

Overview Legal name: YData Labs, Inc. (operating with EU presence via subsidiary) HQ & founding: originated in Portugal; founded 2019; subsequent US HQ move with engineering primarily in Europe and go-to-market in the US. Geographic footprint: International operations; active across EU markets with stronger near-term banking, insurance, pharma and healthcare demand reported in the US. Mission and value proposition: Equip data teams with high-quality, privacy-preserving data by combining data quality assessment, anonymisation/masking, and synthetic data generation within a single workflow that runs entirely in the customer's environment. YData is a core member of the world's largest consortium dedicated to Responsible AI, which brings together startups, research centres, and industry leaders with an €80 million investment to deliver 21 AI projects by 2030 (Center for Responsible AI). Ownership/funding snapshot: Public funding : https://www.crunchbase.com/organization/ydata Gonçalo and Fabiana founded YData and received investment from Flying Fish Partners, Techstars, Real Ventures, Faber and EDP, among other angel investors. At a glance: • Mode: Primarily real-data-based synthesis (B) with B1/B2 inputs; B3 augmentation is a core use. Data generation from scratch is also available (A). • Primary sectors: Financial services (banks, insurance, fintech); healthcare/pharma; telecoms. • Maturity (self-assessed): established. • Engagement model: Platform + services; designed for client-side execution (no vendor custody of raw data). • Public sector: Selected US/EU public-sector pilots/engagements.	 Core offering and technology Primary products & capabilities. YData's platform combines three pillars: (i) data profiling and quality assessment to address upstream issues, (ii) privacy tooling—anonimisation/masking and PII detection for quasi-identifiers, and (iii) synthetic data generation for statistically faithful datasets that enable safe analytics and model development. The stack is positioned as a single workflow for data quality and synthesis, running entirely in the customer's environment. Generation approach and taxonomy alignment. YData operates under Mode B (synthesis using real-world data assets). Generative models learn distributions from the controller's datasets, with quasi-identifiers removed or masked before training. Augmentation (B3) is a core feature, especially for imbalanced classes such as fraud and anomalies. All processing occurs inside the client's infrastructure—on-premises or in their cloud—so the vendor does not take custody of raw data. Privacy-by-design features. The platform combines anonymisation/masking for direct identifiers with synthetic generation for analytical attributes; Differential Privacy (DP) is available as an option. The company argues that for ML-oriented datasets—once quasi-identifiers are removed—synthetic generation breaks one-to-one traceability, with DP and other controls applied as needed. Deployment models and integration. All processing is designed to occur within the client environment, with YData collecting only minimal telemetry for product improvement. Integration patterns include APIs/SDKs into data pipelines, with evaluation artefacts available for internal governance. Domains of activity and illustrative use cases Primary sectors served. YData reports sustained demand from financial services (banks, insurance, fintech) and growing interest from healthcare/pharma and telecoms. Public-sector engagements exist but are not the main focus.
	Finance/insurance examples • Fraud detection: augment minority classes with synthetic records to improve model learning for rare events; claimed impact on detection performance and business KPIs. The interviewee referred to a public case study (available externally) documenting uplift. • Credit risk & marketing analytics: generation and quality steps to create shareable, high-utility datasets for model development when direct use of customer-level data is constrained by policy or regulation. Healthcare examples • R&D and analytics enablement: synthetic datasets derived and evaluated within the controller's boundary to accelerate experimentation and collaboration; US uptake described as stronger, with select EU examples Telecoms & public sector examples • Telco analytics: privacy-preserving datasets for customer analytics and experimentation. • Public sector: specific references include a US defence entity and a Swiss city exploring synthetic data for municipal innovation.
	
	



Privacy assurance and compliance

Evaluation practices (as reported)

- **Anonymisation + synthesis:** direct identifiers (e.g., names, IDs) are masked or replaced; analytical attributes are then handled through synthetic generation to break one-to-one relationships
- **Optional DP and data-privacy evaluation:** DP can be used depending on use case and risk appetite; the survey response also referenced a general "data privacy evaluation" practice. The company mentioned an evaluation white paper and external initiatives oriented to standardising evaluation.
- **Client-side controls:** because execution occurs within the customer environment, security and logging align with the controller's governance; artefacts can be appended to DPIAs and internal audits.

Legal positioning (as stated by the company)

- The interviewee argued that well-executed synthetic generation for ML datasets (after removing quasi-identifiers) can produce outputs that no longer fall under GDPR, as long as good practice is followed and re-linkage risk remains remote; they also acknowledged that legal qualification is context-dependent and poor implementations could invalidate privacy assurances.
- In the survey, YData's response indicated that if the data is synthetic it does not need anonymisation, reflecting the firm's stance on the role of synthetic generation in achieving privacy for ML-ready datasets; this was contrasted by other respondents in the survey who favour explicit anonymisation metrics.

Supervisory engagement & initiatives:

- The firm reported participation in non-profit and consortium initiatives (e.g., AI 2030; EU-funded Centre for Responsible AI), focusing on responsible AI and data privacy standardisation efforts, and contributing to guidance on concepts, methodology, and best practices.

For more information: <https://ydata.ai/resources/syntheticdata-quality-metrics>

Commercial maturity and market signals

Self-assessed maturity. Established.

Adoption signals. The interviewee described an international customer base with finance as a long-standing anchor vertical, healthcare/pharma gaining traction (particularly in the US), and telecoms as an additional growth area. Public-sector cases were noted but characterised as opportunistic rather than core go-to-market.

Engagement model. Although the survey ticked "both options are available" (off-the-shelf and comprehensive engagement), the company described a delivery reality in which platform + services are paired to implement client-side execution, configure privacy/quality parameters, and generate evaluation artefacts for the customer's governance processes.

Differentiators (as positioned):

- Client-side executive by design (no vendor custody of raw data).
- Combined data quality + privacy tooling + synthesis in one flow.
- Focus on augmentation for rare events (e.g., fraud) where business impact is direct.
- Participation in standardisation/consortium work on responsible AI and privacy

Strategic outlook and challenges

Emerging opportunities. The company anticipates continued demand in finance (fraud, credit, analytics) and growing uptake in healthcare/pharma, with telecoms interest persisting. It sees an inflection in openness to experimentation post-GenAI, provided buyers grasp the difference between LLM-generated "fake data" and statistically grounded synthetic data from enterprise datasets.

Operational challenges.

- Awareness gaps and misconceptions—post-GenAI, some buyers equate synthetic data with generic LLM outputs, requiring re-education and elongating sales cycles.
- Governance prerequisites—projects must align with internal policies (DPIAs, access approvals, lawful bases), making client-side deployment essential but time-consuming.
- Regional asymmetries—healthcare momentum is currently stronger in the US; EU procurement often demands richer evidence packs prior to first deployment.

Ecosystem participation. YData reports involvement in standard-setting efforts and non-profit alliances to converge definitions, methods and evaluation—aimed at reducing fragmentation and accelerating adoption across regions.

Figure 12. AINDO company factsheet

Overview Legal name: AINDO S.p.A HQ & founding: Trieste, Italy – founded 2018 Geographic footprint: Primary customer base is in Italy, alongside EU-wide engagements e.g. France, UK. Mission and value proposition: AINDO provides an enterprise-grade solution for generating fully synthetic tabular data, with a strong focus on health data. The technology is GDPR-compliant by design, eliminating doubts around privacy risks and enabling organisations to share and use data with confidence. AINDO's synthetic data preserves the statistical fidelity of the original while ensuring robust privacy protection, with built-in privacy controls and rigorous post-generation verification. By doing so, AINDO empowers clinical research and AI development, enabling compliant data sharing and advanced analytics without compromising data integrity or individual privacy. Ownership/funding snapshot: not disclosed At a glance: • Mode: real-data-based synthesis (B) with B1/B2 inputs and B3 augmentation. • Primary sectors: healthcare/life sciences (70%); others (30%). • Maturity (self-assessed): established. • Engagement model: typically delivered as part of a comprehensive agreement (not off-the-shelf). • Certification: Europrivacy certification – scoped initially to healthcare deployments. AINDO is the only synthetic data company to have achieved this certification so far (as of Sept 2025).	 Core offering and technology Primary products & capabilities. AINDO positions a Synthetic Data Platform focused on tabular enterprise data. The platform supports: (i) generation of fully synthetic datasets that preserve statistical relations, (ii) augmentation to rebalance classes or simulate rare events, and (iii) data “compatibilization” and completion steps that aid downstream analytics. These functions are offered with user-controlled privacy parameters and automatic reporting on data quality and privacy. Privacy-by-design features. AINDO reports user-controlled Differential Privacy (DP) and generalisation options at generation time, alongside nearest-neighbour-based diagnostics and attack-based tests. The product surfaces automatic quality reports and a privacy score per generation run, helping data owners evidence both utility and privacy. Deployment models and integration. Deployments are customer-controlled (on-premises or private cloud), aligning with buyers’ preferences in regulated sectors. Interfaces are exposed via UI and APIs; generation jobs produce reproducible artefacts (parameters, metrics, test results) designed to be shared with assurance functions (DPO, security, legal) as part of internal governance. Domains of activity and illustrative use cases Primary sectors served. The firm’s activity concentrates in healthcare/life sciences (clinical research, RWE/RWD use, data sharing) and finance/insurance (internal data access, model development and validation). In interview, AINDO described a customer mix of ~20 organisations, with ~70% in healthcare and ~30% in other sectors.
 Healthcare examples • Hospitals, research centres, pharma: creation of fully synthetic clinical datasets to accelerate AI-based research, enable statistical analyses and support trial design, while keeping identifiable source data within the controller’s boundary. Post-generation gates are applied before any external sharing. • RWD Liberation to support RWE generation of higher quality	 Other examples (insurances, banks, Telecommunications, PA) • Banking/insurance analytics, telecommunications, and public administration: synthetic copies and augmentation to speed proof-of-concepts and improve model development workflows when direct use of customer-level data is constrained; annual subscriptions or multi-phase projects are typical.  Other observations • AINDO noted growing interest beyond the core sectors, including telecommunication procurements referencing synthetic data, particularly in the UK public tender context, which the firm views as a signal of rising buyer literacy.



Privacy assurance and compliance

Evaluation practices (as reported)

- **Metrics and diagnostics.** AINDO describes a privacy score computed using nearest-neighbour logic, checks for distance to closest record, and rare-category/extreme-value safeguards that reduce memorisation and overfitting. Quality reports include gap analyses of statistical information.
- **Attack-based testing.** The firm indicates deliberate attack simulations (e.g., membership-inference-style probes) and post-generation control that blocks outputs failing pre-set privacy thresholds. This is positioned as a standard gate before any sharing outside the controller's environment.
- **Process and artefacts.** Each run produces reproducible evidence—parameters, metrics, and test results—that can be appended to DPIAs or shared in audits, aligning with buyer expectations identified elsewhere in our study.

Certification and supervisory engagement

- **Europrivacy certification.** AINDO holds Europrivacy certification for synthetic-data processing in healthcare (the only synthetic company in Europe to have achieved it so far)—cited by the firm as an external assurance signal that complements technical evaluation. Our desk sources logged the certification in July 2024; the firm frames this as part of its healthcare go-to-market.
- **Position vis-à-vis GDPR.** The company maintains that anonymisation of outputs is evidenced via pre-/post-generation tests and third-party assessment, not assumed. It emphasises the need for clear, shared criteria to determine when outputs qualify as anonymous data under the GDPR.

Legal framing (as stated by the company)

- AINDO argues that clearer definitions of anonymised data and explicit guidance on what controllers may do with synthetic outputs would reduce friction and support scale, provided decisions are evidence-based (metrics + attacks + governance).

Commercial maturity and market signals

Self-assessed maturity. Established.

Customer base and traction (as reported). Approx. 20 customers, ~70/30 split healthcare/others. Geographically, the Italian customer base represents the majority of AINDO's clients. Additionally, cluster in France and the UK, with additional EU clients. (Counts are indicative, based on interview statements.)

Engagement model. The firm generally provides synthetic data as part of a comprehensive engagement rather than off-the-shelf, combining software with privacy assessment, parameterisation, evaluation runs, and report packaging to meet enterprise assurance thresholds.

Differentiators (as positioned):

- Post-generation privacy gates that block unsafe outputs.
- User-controlled DP/generalisation at generation time.
- Europrivacy badge for healthcare deployments.
- Emphasis on evidence packets that can be appended to DPIAs and audit files.

Strategic outlook and challenges

Emerging opportunities. The firm anticipates continued demand in healthcare (RWD/RWE; research data sharing; clinical AI), steady usage in finance (fraud, risk, internal data access), and broader public procurement references to synthetic data across other sectors. The team expects evaluation discipline to keep rising as buyers standardise evidence expectations.

AINDO is positioning itself most strongly around the regulatory momentum in healthcare, with the recent Italian Bill 1146 serving as a cornerstone development and a starting point for broader EU-level adoption.

Operational challenges.

- Awareness gaps and misconceptions can prolong discovery phases; education remains part of each engagement.
- Certification costs and varied buyer evidence bars can slow first deployments, especially for SMEs; proportionate recognition of existing metrics/attack evidence would help.
- Legal clarity on anonymisation and lawful bases for synthesis/sharing remains uneven across Member States, increasing transaction costs for cross-partner projects.

Methodological breadth. AINDO encourages policy and research to support multiple synthetic data techniques, "including more simulation where possible," so controllers can choose methods that best balance utility and privacy.

Additional information provided by the AINDO team:

On the 17th of September 2025, the joint Italy approved a series of amendments to Bill 1146 on artificial intelligence, proposed by the Government. Among the most relevant changes is the new text of Article 8, dedicated to healthcare research, which introduces three unprecedented legal provisions: (i) the processing of personal data for scientific research and experimentation is declared to be of public interest, provided it is carried out by public or non-profit private entities; (ii) secondary use of personal data without direct identifiers is authorized without the need for additional consent, subject to a general notice published on the controller's website; and (iii) anonymization, pseudonymization, and synthesization of personal data are expressly permitted, with the National Agency for Regional Health Services tasked—following consultation with the Data Protection Authority and taking into account international standards and state of the art—to issue and update guidelines on these practices.

This development represents the culmination of several years of work carried out by Aindo in demonstrating the value of synthetic data for healthcare research. At the same time, it should be regarded as a first step in shaping a broader regulatory framework in this field. Italy is now positioned at the forefront of these initiatives, and it is expected that similar measures will be considered at EU level by other Member States.

2.4.2 Analysis of the current landscape of offering synthetic data solutions

This section presents a structured analysis of the European synthetic data vendor landscape, drawing on mapping results, survey responses, and targeted interviews. The findings are organised to provide a clear overview of (1) the types of synthetic data production in use, (2) the commercial maturity and sectoral application of vendors, and (3) the range of privacy assurance modes adopted across the market. Each subsection uses both quantitative and qualitative evidence, highlighting key patterns, sectoral trends, and the practical realities shaping the adoption and deployment of synthetic data solutions in Europe.

Types of synthetic data production

This subsection applies the coding introduced in the Conceptual Framework and reflected in Table 10: types of synthetic data coding for analytical purposes. Category A represents fully synthetic data, and category B represents data derived from real-world assets, with subcategories B1 personal-data sources, B2 commercial/proprietary sources, and B3 hybrid augmentation.

The vendor mapping shows a clear pattern **12 out of 13 vendors** generate synthetic data using real-world datasets (**Category B**). These solutions typically learn from structured enterprise data or sector-specific datasets and then produce synthetic versions for analysis, testing, or sharing. Most vendors also incorporate **B3 augmentation** features, such as class balancing or rare-event simulation, which were once niche but are now standard.

Only **two vendors in the sample – Anyverse and YData – fall under Category A**. Anyverse creates fully simulated, physics-accurate imagery for computer-vision applications such as advanced driver assistance systems (ADAS) and robotics, whilst YData delivers synthetic data generation, profiling, and orchestration tools that enhance data quality and accelerate AI development across industries like finance, healthcare, and

retail. These approaches avoid GDPR concerns linked to personal data but is limited to specific use cases and does not address the needs of organisations working with tabular or transactional data. Within Category B, **B1 (personal data)** and **B2 (commercial/proprietary data)** are almost universal. Vendors frequently work with sensitive datasets from finance, healthcare, and telecoms, where privacy and confidentiality requirements are strict. This reliance on real data explains why privacy assurance and deployment choices are central to vendor strategies.

Deployment models reflect both compliance obligations and client expectations. On-premises, virtual private cloud, and clean-room environments dominate, even when vendors offer SaaS options. Large organisations—especially in banking, healthcare, and the public sector—prefer to keep processing inside their own perimeter to reduce legal and reputational risk.

Interviews confirm this trend:

- **Aindo** completes generation and **post-generation verification** before any data leaves the controller's environment.
- **YData** designed its platform to run entirely within the client's infrastructure, sharing only minimal telemetry for product improvement.
- **Octopize** applies its Avatar methodology to produce outputs considered anonymous under a documented process, enabling controlled sharing when justified.

Survey responses indicate that many vendors provide evidence packs detailing what was executed, the parameters used, and the resulting privacy and utility metrics. These practices now act as signals of maturity, reassuring buyers that privacy and governance requirements are met.

The dominance of Category B solutions shapes the current market. Vendors primarily handle tabular enterprise data (finance, insurance, health) and, in some cases, image or video streams, where anonymisation or synthetic overlays are applied.

- B1 (personal data) and B2 (commercial/proprietary data) are almost universal: 12 out of 13 vendors work with at least one of these data types.
- B3 augmentation is common, showing that enrichment techniques—once considered advanced—are now part of standard offerings.

Interviews reinforce these findings: **YData** and **AINDO** emphasise local execution on client data; **Octopize** focuses on privacy-assured anonymisation; and **Anyverse** (not interviewed, identified through the mapping) stands out as the only vendor relying solely on simulation.

Commercial maturity and sectorial application

The mapping shows that synthetic data vendors in Europe are most active in **finance, banking, insurance, and healthcare/life sciences**. About 10 out of 13 vendors target financial services, and 9 out of 13 are active in healthcare. These sectors are **natural anchors for synthetic data because they handle high-sensitivity, high-value datasets and face strict regulatory requirements**. In finance, synthetic data is used for fraud detection, risk modelling, and analytics where confidentiality is paramount. In healthcare, it enables research on real-world data, privacy-preserving data sharing, and clinical trial support where direct use of patient data is restricted. Consultation evidence supports this: YData reports strong demand from banks and insurers, with growing interest from healthcare and pharma, while AINDO notes a roughly 60/40 split between healthcare and finance across its customer base. Survey responses also highlight healthcare research and fraud detection as key use cases.

Beyond these anchors, the market is diversifying. **Telecoms and marketing analytics** are targeted by 3 out of 13 vendors, while **public-sector/statistics** and **automotive/ADAS** (including in-cabin monitoring) are each served by 3 to 7 vendors. Telecoms use cases often involve cross-affiliate data sharing and customer analytics under heightened privacy expectations. Public-sector applications include privacy-enhanced statistics and clean-room collaborations, while automotive/ADAS splits between visual anonymisation and high-fidelity simulation for perception tasks. Interviews confirm this spread: Octopize foresees increased number of use cases defence and mobility, and computer-vision vendors illustrate both anonymisation and simulation approaches. This breadth suggests a market that is diversifying, even as privacy assurance and deployment constraints remain central to purchase decisions.

Most vendors in the mapping were founded between 2017 and 2021, reflecting the recent commercialisation of deep generative modelling and the operationalisation of GDPR in enterprise data. The field is a mix of established platforms and growth-stage companies. Some vendors report enterprise references and certifications (e.g., AINDO's Europrivacy for healthcare), while others remain in early commercialisation. Many solutions combine software with consulting to address governance and integration challenges, reflecting the complexity of embedding synthetic data into existing workflows.

Survey responses confirm this pattern. Vendors self-assessed their maturity as ranging from **"growth" to "advanced"**, with references to regulatory sandboxes as stepping stones. For example, Octopize selected "Advanced", AINDO and YData "Established", and others "Growth". Milestones cited include regulatory sandboxing with data protection authorities, publication of privacy-metrics methodologies, and expansion of evaluation coverage (e.g., membership-inference, adversarial accuracy).

Although several vendors market licence-based platforms, the prevailing delivery model is blended: **software is paired with dataset vetting, parameter configuration, evaluation runs, and reproducible reporting**. Survey respondents indicated that synthetic data is often provided "only as part of a more comprehensive agreement" or "both options" (including off the shelf), but open answers describe workflows that bundle professional services with the product. Interviews clarify why: YData emphasises strict customer-side execution to avoid vendor data custody; AINDO runs post-generation privacy controls and blocks outputs that fail checks; Octopize anchors engagements in an anonymisation-first methodology with published metrics and documentation. For buyers, this hybrid model compresses assurance tasks and provides artefacts for security, legal, and DPO functions. For vendors, it shortens the time from pilot to production by addressing governance gaps that would otherwise stall procurement.

Despite platformisation, **sales cycles in regulated sectors remain long and are punctuated by governance checkpoints**. Interviews identify three recurring frictions: (1) awareness gaps—buyers often conflate robust tabular synthesis with "generic fake data", requiring re-education; (2) internal data-governance prerequisites—projects stall until datasets are inventoried and lawful bases clarified; (3) procurement assurance—security reviews, DPIAs, and contract schedules can extend timelines regardless of technical readiness. YData notes that misconceptions have increased in the last two years, elongating discovery. Octopize reports strong traction but lengthy lifecycles in mobility/defence, where partners mature unevenly. Survey entries reinforce these points and add requests for clearer lawful-basis guidance for synthesis and sharing.

Privacy assurance modes

For synthetic data vendors in Europe, privacy is claimed to be built into the core of their products. However, the ways vendors protect privacy can be quite different. Some use differential privacy (DP) in their models, others focus on anonymisation, and some combine these with secure computing environments or "clean rooms". There's also a great deal of variety in how vendors test and prove their privacy claims: some provide

detailed privacy scores and run attack tests, while others rely on external certifications like Europrivacy. This mix of approaches shows that there's no single standard yet, and that balancing privacy with usefulness is complex.

Across the market, four main types of privacy safeguards are common, and many vendors use more than one:

- **Differential privacy (DP).** DP is used either during model training (setting a privacy “budget”) or as a filter for outputs. Vendors often let clients adjust the level of privacy, trading off some data utility for stronger protection. DP is especially common in tabular data platforms and privacy middleware. **Used by at least 5 out of the 13 vendors.**
- **Anonymisation frameworks.** These include removing or masking identifiers in data, creating “avatars” for structured data, or using synthetic overlays for images. When done carefully, these methods can meet GDPR requirements. For example, Octopize’s Avatar method is designed to meet strict EU privacy standards, with supporting documentation and expert review. **Used by all identified vendors.**
- **Privacy-by-architecture.** This means designing systems so that data stays where it is, and only the results leave. Examples include confidential computing, clean rooms, and federated learning. Features like single sign-on, access controls, and audit logs are often included. Vendors like YData and AINDO run their platforms entirely within the client’s infrastructure, so no raw data ever leaves the customer’s control. This is now the norm for sensitive sectors like finance and healthcare. **Used by at least 10 out of 13 vendors.**
- **Process controls and documentation.** Vendors are increasingly providing approval workflows, detailed reports for each data generation run, and “evidence packs” that show what was done, with what settings, and what privacy and utility were achieved. These artefacts help buyers’ security and legal teams check compliance. **Used by at least 9 out of 13 vendors.**

A key trend is the move towards regular, quantitative privacy testing alongside checks for data usefulness:

- **VEIL.AI** uses DP together with classic privacy checks (like k-anonymity and l-diversity), and tests for risks such as membership inference and overfitting.
- **AINDO** checks how close synthetic records are to real ones and blocks outputs if privacy risks are found.
- **Octopize** uses a published privacy-metrics framework to check that its Avatar outputs are truly anonymous.

Many vendors (at least 7 out of 13) now offer dashboards that show privacy and quality metrics for each dataset, making it easier for clients to audit and trust the results.

There’s no single answer to when synthetic data is truly anonymous under GDPR. **Octopize** claims its Avatar outputs are anonymous, based on expert review and published criteria. **AINDO** relies on Europrivacy certification and a mix of tests and third-party checks. **YData** focuses on removing direct links to real people, but admits that context and good practice matter.

Survey responses show this diversity: for example, YData stated that “if the data is synthetic, it does not need anonymisation”, while others—such as VEIL.AI — called for clear anonymisation metrics, including inference-attack testing, or expert determination before considering data anonymous. This illustrates an area where practice is ahead of official guidance.

Some vendors go further by getting formal certifications or working with regulators, but this is still not the norm:

- AINDO has Europrivacy certification.
- Some vendors (2/13) have had their methods reviewed by national data protection authorities.
- Clean room deployments and confidential computing are sometimes used as extra assurance (3/13 vendors).

However, certifications like ISO or Europrivacy can be expensive, especially for smaller companies, so many rely on transparent metrics and process controls unless a client or regulator requires formal certification.

Finally, most vendors now offer solutions that keep all processing within the customer's own systems—on-premises, in private clouds, or in clean rooms. Raw data never leaves the client's control; only privacy-protected results do. This is not just about ticking compliance boxes—it's about building trust, especially in regulated sectors. Many vendors also bundle their software with services to help clients set up, configure, and document privacy controls.

2.4.3 Current and future challenges and policy implications

2.4.3.1 International dimensions and comparative insights

The Comparisons with the US market were mainly drawn from consultation interviews with synthetic data vendors.

The **US market for synthetic data is moving faster than the EU**, mainly because **privacy rules are lighter and adoption is easier**. Vendors interviewed consistently highlighted this difference. In the **EU**, strict **GDPR requirements** shape the entire approach to synthetic data. Companies like **AINDO** and **Octopize** emphasised that compliance involves **detailed documentation, privacy metrics, and sometimes costly certifications such as Europrivacy**. While these measures are seen as **best practice** and a potential **competitive advantage**, they also **slow down adoption** and **increase costs**, particularly for **smaller firms**. Several respondents noted that **legal uncertainty** around what qualifies as **anonymised data** adds further delays, as buyers and legal teams often require extensive evidence before approving projects.

In contrast, the **US market offers a more flexible regulatory environment**, which **accelerates commercial uptake**. YData explained that this was a key reason for **moving its headquarters to the US** while keeping technical teams in Europe. The company described the US as a **larger, more dynamic market** with **fewer procedural hurdles** and a **stronger appetite for experimentation**, especially in healthcare and finance. This openness allows vendors to launch pilots and scale solutions more quickly than in Europe. Interviewees also pointed out that US incentives, such as **R&D tax credits**, provide a straightforward way to support innovation, whereas EU funding is often seen as **bureaucratic and slow**.

The competitive impact of these differences is clear. While European vendors argue that GDPR-aligned methods should position them as leaders in privacy-preserving technologies, the reality is that **fragmented rules** and **long procurement cycles** make it **harder to compete internationally**. Some firms are shifting resources to the US to capture faster revenue opportunities. Both markets share strong demand in sectors like healthcare, finance, and telecoms, **but the drivers differ: EU buyers prioritise compliance and explainability**, while **US buyers focus on speed and integration with AI workflows**. Across interviews and survey responses, vendors called for clearer global standards and “portable” privacy evidence to reduce duplication and help European companies compete on equal terms.

2.4.3.2 Challenges and outlook

The European synthetic-data market faces a set of recurring obstacles that slow adoption and complicate scaling. These challenges are primarily governance and regulatory rather than technical.

A first issue is the **fragmentation of assurance practices**. Vendors adopt different approaches to demonstrating compliance: AINDO combines privacy-preserving techniques with post-generation checks and privacy scores; Octopize relies on its CNIL-reviewed Avatar method; YData publishes evaluation frameworks and white papers. While these practices show maturity, the absence of a common evidentiary baseline means buyers face a comparison problem, extending review cycles and complicating procurement.

A second challenge is **legal uncertainty**. Survey responses repeatedly call for clearer definitions of anonymised data and lawful bases for synthesis and sharing. Vendors stress that qualification of synthetic outputs as “anonymous” remains context-dependent and risk-based, leaving legal teams cautious and slowing cross-partner projects.

The **cost and complexity of certification** is another recurring theme. Europrivacy and similar schemes are valued as trust signals, particularly in healthcare, but they are expensive and resource-intensive. AINDO highlighted the burden of multi-scheme compliance when combined with long enterprise sales cycles, while Octopize noted that certification costs can deter SMEs from entering the market.

Finally, **market dynamics add pressure**. YData explicitly stated that it moved its headquarters to the US to access a larger, more dynamic market with fewer procedural hurdles. Survey comments echo this, noting that non-EU countries are adopting synthetic data faster, aided by lighter privacy rules and incentives such as R&D tax credits.

Outlook: what the next 12 – 24 months may bring

The evidence base suggests a market in transition. While structural challenges remain, several trends point towards greater maturity, standardisation, and competitive repositioning.

Increasing frequency of evidence packs in vendor deliverables

Interviews confirm that vendors already provide structured privacy and utility reports as part of their delivery model:

- **AINDO** offers automatic quality reports with privacy scores and attack-based checks.
- **Octopize** shares CNIL-reviewed metrics and detailed documentation.
- **YData** publishes evaluation frameworks and white papers.

This practice may harden into a market norm, with evaluation artefacts treated as first-class deliverables rather than optional extras. Over the next two years, there may be:

- **Automation of evidence generation**, reducing manual effort and human error.
- **Standardised templates** for portability across internal reviews.
- **Evaluation-as-a-service models**, offering dashboards and APIs for continuous monitoring.

This shift could reflect a move from ad hoc evidence to repeatable, auditable processes that buyers can trust. Vendors that make assurance easier may gain a competitive edge.

Push for interpretative clarity

Legal ambiguity remains the single biggest governance bottleneck. Survey respondents explicitly requested:

- **Clearer anonymisation criteria**
- **Guidance on lawful bases for synthesis and sharing**

While legislative change seems unlikely in the short term, **soft-law instruments**—such as EDPB opinions, sectoral codes of conduct, or national DPA guidance—may emerge within 12–18 months. These could provide interpretative clarity that lowers legal risk and shortens review cycles. Vendors may also advocate for **recognition of existing privacy metrics** as part of compliance evidence, which could accelerate convergence.

Certification evolves

Europrivacy is likely to remain relevant, particularly in healthcare and finance, but its current cost profile may be unsustainable for SMEs. The next phase could involve:

- **Modular or proportionate certification**, allowing vendors to certify specific components (such as privacy metrics or deployment architecture) rather than entire platforms.
- **Sandbox-based validation**, as seen in Swedish IMY pilots referenced in survey responses, offering structured but lighter-weight routes to trust.
- **Reuse of artefacts** generated in these settings to avoid duplication across multiple reviews.

Such changes may be critical to balancing trust with affordability.

Controller-side execution consolidates

The compute-to-data model—running synthesis inside the customer’s perimeter—has already become common for sensitive workloads. AINDO and YData both design products to run entirely within client environments, avoiding data transfer and reducing compliance risk. This pattern may harden into a baseline expectation, with differentiation shifting to:

- **Ease of integration** with enterprise security stacks (identity, logging, attestation)
- **Pre-certified deployment blueprints**, which could become a selling point for regulated sectors.

Sectoral growth patterns hold

Healthcare and finance may continue to dominate adoption, driven by **strong compliance incentives and clear business cases** such as fraud detection and clinical research. Telecoms and mobility could grow more slowly, constrained by fragmented governance and lower immediate ROI. Public-sector uptake may hinge on:

- Clarity around lawful bases.
- Budget for certification.

Simulation-based approaches could **expand in computer vision and robotics**, where they bypass personal-data concerns entirely.

Hybrid and augmentation use cases expand

Synthetic data may increasingly **complement** rather than replace real data. YData and AINDO both described workflows where synthetic data augments real datasets—for example:

- Rebalancing rare classes in fraud detection.
- Enriching clinical datasets for research.

There may be **growth in augmentation scenarios**, paired with stronger documentation of provenance. Vendors could position these workflows as risk-managed innovation, aligning with governance norms while improving model performance.

Competitive pressure from the US intensifies

YData's move to the US and survey comments underline the risk: **lighter US privacy rules and R&D incentives** may enable faster scaling. Unless EU guidance and assurance pathways become more predictable, European vendors could lose ground in high-growth sectors like healthcare AI. Likely responses may include:

- **Greater emphasis on privacy as a differentiator**, with EU vendors marketing GDPR alignment as a trust advantage.
- Calls for **EU-backed incentives**, such as simplified R&D credits or funding tied to evidence-based compliance.

Convergence on core metrics

Survey responses list common checks—membership inference, nearest-neighbour distance, adversarial accuracy—alongside classical disclosure-risk criteria such as k-anonymity, t-closeness, and l-diversity. Over time, these could form a de facto baseline, supported by:

- **Open-source libraries** maintained by industry consortia.
- **Horizon Europe projects** acting as conveners for reference implementations.

This convergence may reduce duplication and make assurance more portable across institutions.

Broader governance signals

Bias and **explainability** were flagged in interviews as emerging concerns. While not yet formalised, these issues **may enter procurement criteria** within the next two years, especially in public-sector and regulated domains. Vendors that can demonstrate auditable fairness checks alongside privacy metrics could enjoy a competitive edge.

Likely market scenarios

Conservative: Adoption remains concentrated in finance and healthcare, with bespoke evidence and slow cross-sector diffusion. Certification stays costly; legal ambiguity persists.

Mainstream (most likely): Standardised evidence packs become common; modular certification and controller-side blueprints reduce friction; uptake broadens to telecoms and mobility.

Ambitious: EU-recognised evidentiary baseline emerges; portable sandboxes and proportionate certification accelerate scaling; synthetic data becomes a default privacy-enhancing tool across sectors.

2.4.3.3 Policy recommendations

This section summarises the actions that companies and vendors explicitly asked public authorities to consider during our consultations (Task 1b survey and interviews). The recommendations reflect practical frictions encountered in procurement, assurance, legal

qualification, and market development. They are grouped into two main strands, as repeatedly emphasised by company stakeholders.

Clarify GDPR scope and adopt a more permissive approach

Companies consistently called for clearer, more practical, and more favourable rules for synthetic data under GDPR. There is a strong desire for interpretative clarity so that controllers can make defensible, repeatable decisions supported by metrics and attack tests, rather than relying on informal judgement. Specifically, companies requested:

- **Clear criteria for anonymity:** Define when synthetic data qualifies as “anonymous” and what evidence is sufficient to support that conclusion. Provide a framework and explicit guidance on the lawful basis for synthesis and sharing, especially in cross-partner projects.
- **Standardised evaluation practices:** Recognise and encourage the use of transparent privacy-evaluation metrics and attack tests (e.g., membership inference, nearest-neighbour distance, adversarial accuracy, k-anonymity, t-closeness, l-diversity). The aim is to normalise a minimum evidentiary baseline, so evaluation becomes portable across projects and authorities, without mandating a single tool.
- **Consistent enforcement and awareness:** Enforce GDPR consistently across sectors and raise awareness of robust privacy engineering. Companies highlighted the need for myth-busting and authoritative explanations to distinguish mature solutions from superficial claims, and to ensure privacy is treated as a design discipline, not a checkbox.
- **Recognition of anonymised/synthetic data for AI/ML training:** Explicitly recognise the use of anonymised synthetic data for AI/ML training when evidence-based thresholds are met, provided outputs pass documented privacy tests and are accompanied by reproducible evaluation artefacts.
- **Expand and coordinate regulatory sandboxes:** Continue and expand sandbox environments for validating evaluation methods and governance packs. Broader availability and light coordination could reduce duplicative effort across Member States.
- **Proportionate and accessible certification:** Ensure certification schemes (e.g., Europrivacy) are proportionate and mindful of SME resources, building on—not duplicating—the evaluation evidence vendors already produce.

Support EU companies with grants and by shielding them from unfair competition

Companies repeatedly emphasised the need for practical support to compete internationally, especially against US firms. They called for a more level playing field and incentives that are outcome-focused and accessible. Their requests include:

- **Practical incentives for privacy engineering and evaluation:** Prioritise outcome-linked incentives (such as funding for evaluation efforts or core R&D on data quality and synthesis), rather than complex, long-horizon grants. R&D credit-style incentives, as used in the US, are seen as more effective for supporting innovation and aligning with how startups build products.
- **Level the playing field for EU-based suppliers:** Provide more help for software vendors competing against US companies, addressing heavier assurance expectations and higher certification costs faced by EU SMEs. Companies want Europe’s privacy-first approaches to be an advantage, not a handicap, and called for policy and programme choices that reduce structural disadvantages—such as access to supervised pilots, visibility in public sector projects, and recognition of evaluation artefacts.

- **Continue supporting methodological breadth:** Support a range of synthetic data generation techniques, including simulation, alongside tabular generation and augmentation. This is seen as a way to foster innovation and allow controllers to choose the best technique for their privacy and utility needs.

3 Task 2: Using automated data reporting mechanisms for demonstrating regulatory compliance

Task 2 investigates how **automated data reporting mechanisms** can be leveraged to demonstrate regulatory compliance within the European Union. To ensure a robust and representative analysis, **the study focuses on five sectors**, selected according to the criteria established in the methodology. By examining these sectors in detail, the study aims to identify current compliance burden and assess benefits and challenges related to the different techniques used for reporting.

3.1 Overall methodological approach

3.1.1 Selection of five sectors

Selection criteria

The identification of the five sectors for in-depth analysis was based on a structured and transparent methodology aimed at ensuring that the study would focus on areas where automated data reporting has both the greatest relevance and potential impact. The objective was not only to capture sectors already experimenting with or deploying such technologies, but also to identify those where regulatory obligations create strong incentives for automation, even if adoption remains limited at present. This dual focus allowed the selection to combine **technological opportunity** with **regulatory significance**, ensuring that the chosen sectors are illustrative of broader trends while also providing practical insights into how automation could alleviate compliance burdens. Two main factors were applied to determine which sectors would be included in the analysis:

- **Evidence of current or potential use of automated reporting technologies.** Preference was given to sectors where digital or automated tools are already in use, or where strong potential exists for their application in the near future. This includes industries undergoing active digital transformation, those with established reporting systems that can be enhanced through automation, and sectors where early pilots of advanced solutions such as artificial intelligence (AI), robotic process automation (RPA), or blockchain are emerging. The presence of such initiatives indicates both technological readiness and a capacity to integrate new solutions into existing compliance workflows.
- **Regulatory burden stemming from EU legislation.** Sectors subject to substantial EU-level reporting obligations were prioritised. These are areas where compliance activities are complex, resource-intensive, and recurrent, often spanning multiple legislative acts and requiring frequent submissions of detailed information to public authorities. In such contexts, advanced and automated techniques have a clear potential to streamline processes, reduce administrative costs, enhance the quality and consistency of reporting, and enable more timely monitoring of compliance.

The combination of these two criteria ensured that the selected sectors are highly relevant both from a technological standpoint and in terms of their regulatory context.

Selection process and outcome

A long list of sectors was compiled through desk research, focusing on industries characterized by high regulatory intensity and some degree of technological readiness for automation. Two sectors (Agriculture and Automotive) were analysed first to validate the methodology and data collection instruments. These initial case studies helped refine the analytical framework, survey design, and interview guidelines. Three additional sectors were then chosen to complete the analysis. The final set of five provides a balanced

coverage across different parts of the EU economy, including primary production, manufacturing, and services. Further details are provided in the summary table below

Table 13. Overview of selected sectors for Task 2

Sector	Rationale for selection
Agriculture	The sector is subject to extensive reporting under CAP and related environmental requirements, generating structured data flows through established systems (IACS, AMS). It includes a wide range of actors with varying levels of digital maturity, making it representative for assessing scalability and inclusiveness of automated solutions. Its strong link to sustainability policies adds relevance for broader EU objectives.
Automotive	Automotive manufacturers must comply with extensive EU regulations and directives that require recurrent reporting using structured templates. The presence of defined indicators and schemas provides a solid foundation for testing automation, while the sector's complex, multi-actor ecosystem offers an opportunity to examine interoperability challenges. Given its high level of investment in new technologies, the sector also offers a suitable environment for evaluating next-generation compliance solutions.
Healthcare	Healthcare faces complex obligations across multiple EU platforms (e.g., MDR/IVDR, CTIS, pharmacovigilance), involving sensitive data and multi-actor workflows. These characteristics make it a relevant context for exploring secure and reliable automation approaches. Findings can inform practices in other highly regulated domains where data integrity and transparency are critical.
Chemical	The sector is governed by comprehensive EU legislation (i.e., REACH, CLP) and relies on centralised IT systems for dossier generation and submission. This creates a structured environment for piloting digitalisation initiatives. Compliance activities are resource-intensive, making efficiency gains measurable in terms of reduced administrative burden. Moreover, the sector's role within complex supply chains provides an opportunity to assess the broader applicability of proposed solutions.
Energy	Energy combines environmental, efficiency, and market transparency obligations, resulting in frequent and data-intensive reporting to EU bodies. Existing digital channels and schemas create a foundation for testing automation in high-volume, time-sensitive contexts. Insights from this sector can support broader objectives related to decarbonisation and system reliability.

Caveats and limitations

The analysis of sectoral uptake of automated reporting technologies is subject to several important caveats and limitations that should be considered when interpreting the findings.

- **Low uptake of automated solutions in certain sectors**

In some sectors, the deployment of automated reporting technologies remains limited. As a result, the analysis shifts from examining existing practices to exploring potential applications and future opportunities. This necessarily places greater emphasis on stakeholder consultations—through surveys and interviews with businesses, industry bodies, and technology providers—as the primary source of evidence. Desk research alone provides only partial insights in such cases. Consequently, the analysis focuses more on barriers to adoption, such as high costs, lack of sector-specific solutions, or low digitalisation levels, rather than on detailed accounts of actual implementation. Findings in these sectors should therefore be interpreted as indicative or exploratory rather than fully evidence based.

- **Variability in data availability and evidence depth**

The availability and quality of information vary significantly across sectors. In domains where automated reporting is better documented, case studies can be more comprehensive and grounded in concrete examples of technologies, benefits, and challenges. By contrast, in sectors with less evidence, the analysis is necessarily more speculative, often drawing on expert judgment and hypothetical scenarios.

- **Differences in stakeholder engagement**

The depth of analysis is also influenced by the degree of stakeholder participation. Variations in the willingness of organisations to engage, their familiarity with automated compliance tools, and the representativeness of the consulted groups may affect the robustness and granularity of findings. Limited or uneven participation can constrain the ability to provide systematic comparisons across sectors.

- **Resulting heterogeneity of findings**

Taken together, these factors mean that the final set of sectoral studies will likely display a heterogeneous character. This outcome reflects the differing levels of technological maturity and documentation across sectors and should be clearly acknowledged when interpreting the study results.

3.1.2 Methodological toolbox for sectorial analysis

3.1.2.1 Desk research

The desk research formed the backbone of the sectoral analyses, providing the legal, technical, and contextual basis for understanding compliance and reporting practices. It combined the review of EU legislation, regulatory guidance, evaluation studies, academic literature, and industry reports with supporting documentation from compliance software providers and sector-specific platforms

Agriculture

The desk research for the agriculture sectorial analysis drew on a broad range of EU-level documentation, technical reports, scientific literature, and resources from private tool providers. Key topics explored included regulatory compliance obligations under the CAP, administrative and reporting burdens for farmers, and the adoption of digital and automated tools in agricultural compliance.

Core materials included regulatory and analytical outputs from the European Commission—particularly DG AGRI—such as the *Analysis of administrative burden arising from the CAP* (2024),³⁶¹ guidance on *Conditionality under the CAP* (2024),³⁶² and the *Annual Burden Survey* (2021).³⁶³ These were complemented by recent targeted outputs, such the *Simplification survey on the specific issues faced by farmers across the EU* (2024)³⁶⁴, which sheds light on specific regulatory burdens faced by farmers.

The review also covered recent academic literature on digitalisation in agriculture, particularly studies examining the adoption of precision farming techniques and digital recordkeeping tools. This literature review was further complemented by the analysis of websites and documentation from private farm management software providers.

³⁶¹ European Commission, *Analysis of administrative burden arising from the CAP*, 2024. https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cmef/regulation-and-simplification/analysis-administrative-burden-arising-common-agricultural-policy_en

³⁶² European Commission, *Conditionality, Common Agricultural Policy*, 2024. https://agriculture.ec.europa.eu/common-agricultural-policy/income-support/conditionality_en

³⁶³ European Commission, 2021. *Annual Burden Survey* 2021. https://commission.europa.eu/system/files/2022-07/annual-burden-survey_2021_en.pdf

³⁶⁴ European Commission (2024). *Simplification survey sheds light on specific issues faced by farmers across the EU*. https://agriculture.ec.europa.eu/media/news/simplification-survey-sheds-light-specific-issues-faced-farmers-across-eu-2024-04-12_en

Automotive

The desk research for the automotive sector drew on a wide array of sources, including official reports from European institutions, impact assessments conducted within the EU legislative process, academic literature, technical documents and guidelines issued by professional organisations, as well as information provided by automotive manufacturers and developers of commercial reporting software.

The legal analysis relied on summaries of EU legislative acts provided by the Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (DG GROW).³⁶⁵ In addition, the *2023 Automotive Regulatory Guide* prepared by the European Automobile Manufacturers' Association (ACEA) was consulted to gain a better understanding of the scope of existing directives and regulations.³⁶⁶ These documents informed the selection of the legislative acts examined in the case study. The research focused on EU regulations and directives that impose substantial compliance and reporting obligations in the automotive industry, including requirements for data collection, processing, and submission of reports in standardized formats. Priority was given to legislation with broad applicability over narrowly focused regulations. The analysis also assessed the potential for these activities to be automated, ensuring alignment with the study's objectives.

The review was complemented by additional searches of scientific and grey literature to identify current practices in regulatory compliance and reporting. Professional service providers' websites were also consulted to gain insights into commercially available solutions and their comparison with existing methods.

Healthcare

The desk research for the healthcare case study draws on a broad mix of EU-level legal and technical documentation, scientific literature, and materials from compliance software providers. Core sources include implementation guidance and operational documents published by the European Commission and the European Medicines Agency (EMA), with a particular focus on the regulatory use of EU-level platforms such as EUDAMED,³⁶⁷ Clinical Trials Information System (CTIS),³⁶⁸ and EudraVigilance.³⁶⁹

In addition, academic studies analysing the impact of regulatory changes on healthcare institutions, laboratories, and manufacturers were reviewed, especially in relation to clinical evaluation, post-market surveillance, and data reporting obligations. A key reference in this regard was the 2024 MedTech Europe survey on the administrative

³⁶⁵ Summaries of the EU legislation concerning the automotive sector are available at the following address: <https://ec.europa.eu/docsroom/documents/45484>. For a list of regulations and directives related to the manufacturing of motor vehicles, their trailers, systems, and components see https://single-market-economy.ec.europa.eu/sectors/automotive-industry/legislation/motor-vehicles-trailers_en.

³⁶⁶ The guidelines can be accessed at <https://www.acea.auto/publication/automotive-regulatory-guide-2023>.

³⁶⁷ European Database on Medical Devices - <https://ec.europa.eu/tools/eudamed/#/screen/home>.

³⁶⁸ CTIS - <https://euclinicaltrials.eu/search-for-clinical-trials/trial-map/?lang=en>.

³⁶⁹ EMA. (2024). *Annual Report 2024*. https://www.ema.europa.eu/en/documents/annual-report/2024-annual-report-european-medicines-agency_en.pdf.

European Commission (2024). *Factsheet for healthcare professionals and health institutions*: https://health.ec.europa.eu/document/download/afeafd1b-5f7e-4029-8201-4c9b1cc12e0c_en?filename=healthcareprofessionals_factsheet_en.pdf

burden of MDR and IVDR,³⁷⁰ which provides detailed evidence on rising compliance costs, documentation workloads, and staffing pressures faced by industry actors.

The research further examined grey literature and product documentation from providers of healthcare compliance software, as well as companies applying advanced technologies to electronic health records (EHRs). This review identified emerging trends, such as the use of AI in adverse event tracking or blockchain for audit trails, which informed the formulation of the survey questionnaire.

Chemical

The desk research for the chemicals sector builds on the approach developed for the automotive sector and draws on a wide range of sources, including both official publications and private-sector information – such as data shared by companies or providers of commercial solutions.

The legal analysis builds on the summary of EU legislation relevant to the chemicals sector, compiled by DG GROW and available online.³⁷¹ In addition, the Study team consulted the European Chemicals Agency's (ECHA) Legislation Finder (EUCLEF) to identify further regulations and directives that may be applicable to the industry.³⁷²

The review focused on legislative acts with significant compliance and reporting implications, including the REACH Regulation, the Classification, Labelling and Packaging (CLP) Regulation, the Biocidal Products Regulation, the Fertilising Products Regulation, and the Corporate Sustainability Reporting Directive. Additional acts such as the POPs Regulation, the Prior Informed Consent Regulation, the Waste Framework Directive, the Detergents Regulation, and legislation concerning explosives, pyrotechnics, and drug precursors were also considered relevant.

Energy

The desk research for the energy sector drew on a comprehensive review of EU legislation, regulatory guidance, evaluation studies, and supporting documentation from compliance platforms.

In addition to the legal texts, the research drew on the European Commission's ETS Reporting Tool and related templates and guidance, the ACER REMIT Portal and Transaction Reporting User Manual (TRUM), as well as supporting documentation from ENTSO-E's Transparency Platform. These sources provided insights into how compliance and reporting requirements are operationalised across regimes.

To capture user-side challenges, evidence was collected from evaluations and consultations, including the *Joint Research Centre's review of the first Article 8 compliance cycle under the Energy Efficiency Directive*,³⁷³ the *2025 study Harmonising Energy Audit Reporting: Addressing Data Loss and Policy Challenges in the EU Member*

³⁷⁰ MedTech Europe (2024), *The Administrative Burden of MDR and IVDR: Evidence from Industry Survey*. https://www.medtecheurope.org/wp-content/uploads/2025/03/250318_mte-report-on-admin-burden-ivdr_mdr_final.pdf.

³⁷¹ Relevant legislative acts regarding the chemicals sector are available at the following address: https://single-market-economy.ec.europa.eu/sectors/chemicals/chemicals-legislation_en.

³⁷² EUCLEF can be accessed at the following address: <https://echa.europa.eu/legislation-finder>.

³⁷³ Serrenho, T., *Analysis on the practices to collect, store and assess information arising from Energy Audits in the EU-28 - EED Article 8 provisions on Energy Audits' information management, monitoring and verification*, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-08766-3, doi:10.2760/82756, JRC109382.

States,³⁷⁴ and the 2024 ACER public consultation on REMIT reporting obligations.³⁷⁵ Additional perspectives were derived from industry position papers, particularly from the electricity sector, and from the external review of the ENTSO-E Transparency Platform, which assessed data completeness, timeliness, and usability.

3.1.2.2 Survey

The surveys aimed to collect evidence on the techniques currently used for regulatory compliance and reporting across the selected sectors. In addition, it sought to gather stakeholder perspectives on the potential benefits and challenges of adopting advanced and automated technologies.

Survey questionnaires and dissemination

The desk research played a key role in shaping the design of the survey questionnaire. It first identified the most commonly used techniques and tools for data reporting, enabling the Study team to include concrete and relevant examples in the questionnaire. This approach helped ensure that the survey was both relatable and accessible to respondents. In addition, the desk research shed light on existing challenges and limitations associated with these techniques. These insights were then incorporated into the questionnaire to gather respondents' feedback on the identified issues.

The questionnaires were developed to explore the current use, limitations, and future outlook of digital and automated tools used for regulatory compliance across the five sectors under study. The survey is structured into five sections:

1. **Profiling:** Background information on respondents, including the type of organisation they represent, the activities or subsectors they cover, and their level of involvement in reporting and compliance tasks.
2. **Overview of current use of reporting and compliance techniques:** Mapping of compliance tasks under the selected regulations and identification of the tools or techniques used to fulfil them, classified as paper-based/manual, digital spreadsheets, advanced digital solutions, sector-specific management systems, or fully automated tools.
3. **Challenges and limitations:** Information on barriers linked to specific tools or techniques, such as cost, usability, or interoperability issues.
4. **Expected costs and benefits:** Exploration of the organisational and economic implications of moving towards more digital or automated systems.
5. **Outlook and recommendations:** Stakeholder views on future trends, funding needs, and policy changes required to improve reporting and compliance processes.

To ensure clarity and ease of response across stakeholder types, the questionnaire includes plain-language explanations and concrete tool examples in each section.

³⁷⁴ Pellegrini-Masini, G., Leal-Arcas, R., & Colmenar-Santos, A. (2025). *Harmonising energy audit reporting: Addressing data loss and policy challenges in the EU Member States*. Energy Policy, 191, 113112. <https://doi.org/10.1016/j.enpol.2024.113112>.

³⁷⁵ Agency for the Cooperation of Energy Regulators, 2024, *Evaluation of responses to the public consultation of ACER on the revision of the Annex of the REMIT Implementing Regulation*. https://www.acer.europa.eu/sites/default/files/documents/Official_documents/Public_consultations/PC_2024_R_01/PC_2024_R_01_Evaluation_of_Responses.pdf.

The agriculture, automotive, and healthcare surveys were administered through the online platform Forsta, while stakeholders in the chemicals and energy sectors were invited to provide input by completing a questionnaire attached to the consultation request.

All surveys were conducted between May and August 2025. Invitations were sent to targeted stakeholder lists in each sector, with recipients encouraged to circulate the link within their networks to broaden outreach. Across all surveys, reminders were issued midway through the response period and again shortly before closure to maximise participation.

Table 14. Overview of survey responses

Sector	Survey period	Total numbers of contacts	Total number of completed responses
Agriculture	21 May – 31 July, 2025	235	4
Automotive	12 May – 31 July, 2025	317	8
Healthcare	7 July – 8 August, 2025	293	6

Despite substantial outreach—including targeted invitations, multiple reminders, and encouragement to circulate the survey within stakeholder networks—the **number of completed responses was limited across all sectors**. This restricts the representativeness of the findings, so results should be interpreted as indicative rather than comprehensive. Several factors likely contributed to the low response rate: the **technical and sensitive nature of regulatory compliance**, the perception of automation as a **highly advanced topic** (especially in certain sectors), and possible survey fatigue among targeted groups. To address these limitations and enrich the evidence base, additional targeted interviews were planned and delivered, enabling the Study team to gather more in-depth qualitative insights from key stakeholders.

Survey sample

A total of 19 completed responses were received across the three sectors, representing a mix of associations, industry representatives, service providers, and knowledge actors. While the response rate was relatively low—reflecting the stakeholder engagement challenges outlined in the Task 2 caveats—the survey was designed with extensive detail to ensure comprehensive coverage of relevant topics and issues. To address remaining information gaps, additional interviews were conducted.

In agriculture, responses mainly came from EU-level associations covering crop production, mixed farming, and agricultural support services. The automotive sector sample combined vehicle manufacturers, component suppliers, and actors from the broader e-mobility ecosystem, including both EU-level and global organisations. In healthcare, most responses came from Notified Bodies and authorised representatives, complemented by one national-level healthcare association, one industry representative, and one academic stakeholder. Overall, the sample provided a broad geographic spread, with most respondents operating at EU level, alongside national, regional, and global perspectives.

Table 15. Overview of survey samples

Sector	Type of stakeholders	N.	Geographical scope	
Agriculture	Crop production (e.g., cereals, vegetables, fruits, oilseeds, pulses)	2	EU Level	3
	Mixed farming (combination of crop and livestock farming)	1		

	Agricultural support services (e.g., soil preparation, harvesting, pest control, advisory services)	1	National	1
Automotive	Vehicle manufacturers	2	EU-level	3
	Component suppliers	4	Regional	1
	Vehicle manufacturers & suppliers	1	Other (i.e., global)	4
	E-mobility eco-system	1		
Healthcare	Healthcare association	1	Other	5
	Notified Bodies	4		
	Industry and manufacturers' representatives	1		
	Academia	1	National	1

3.1.2.3 Follow up targeted interviews

The follow-up interviews were designed to complement the data collected through the survey and provided an opportunity to capture more detailed qualitative insights. They were particularly valuable in highlighting nuances around the challenges faced by different stakeholder groups and in allowing participants to articulate their main concerns and ideas for possible policy options. The discussions were tailored to the profile of each interviewee, ensuring that the questions focused on the most relevant compliance and reporting issues in their sector.

Table 16. Overview of interview respondents

Sector	Type of stakeholders	N.	Topics discussed in the interview
Agriculture	Farmers association	1	Proportionality of administrative burden for small farms; limits of automated controls; digital divide between large and small farmers; preference for human-centred inspections; call for free public digital tools.
	Farmer association	1	Uneven digital adoption; tools bought for operations first (compliance secondary); cost and interoperability issues; offline functionality needs; concerns about vendor lock-in; benefits for results-based schemes.
Automotive	Automobile manufacturers' association	3	Reporting activities required under EU legislation; challenges faced by manufacturers in conducting compliance and reporting activities; methods and tools used in reporting; potential of automated tools to streamline the process
Healthcare	Authorised representatives for medical devices	2	High MDR/IVDR workload; EUDAMED delays and lack of alerts; fragmented vigilance formats; non-standard NB submissions; need for AR-led data entry; early-stage, costly, non-interoperable digital tools.
Chemical	Industry association for chemical substance manufacturers	1	Reporting obligations under REACH and CLP; challenges faced by manufactures and the type of techniques used to carry out compliance and reporting activities; potential of automated tools to streamline the process
	Representatives of an industry which is a downstream user of chemicals	3	Reporting obligations under the EU legislation; challenges faced by manufacturers and the type of techniques used to carry out compliance and reporting activities; potential of automated tools to streamline the process

	European Chemicals Agency (ECHA)	6	Reporting requirements under the ECHA-relevant legislation; tools developed by ECHA to streamline reporting process; challenges experienced by users; potential to automate reporting process; future policy changes
Energy	Regulatory compliance software provider	3	Functionality and use cases of an AI-powered regulatory compliance platform; benefits and drivers of adoption among companies of automated tools; challenges related to interoperability, integration and scalability; reflections on gradual uptake of automation and the future of digital compliance tools in the energy sector.

Disclaimer: limited stakeholder inputs for Energy sectorial analysis

Despite outreach to over 100 stakeholders, only one interview was successfully conducted for the Energy sector analysis. Most contacted stakeholders either declined or did not respond, often citing the subject matter as too technical. This outcome reflects the broader limitations outlined in Section 3.1.1 of the report.

Consequently, the analysis for the Energy sector relies primarily on desk research, complemented by insights from the single interview. As a result, the level of detail and granularity may be lower compared to other sectors, as noted in the general caveats and limitations for Task 2.

Furthermore, the policy options presented for the Energy sector should be regarded as the authors' interpretation based on identified gaps in existing frameworks. These options are not supported by direct stakeholder validation and should therefore be considered indicative rather than consensus based.

3.2 Conceptual framework

3.2.1 Overview of reporting activities and processes included in the analysis.

Reporting and compliance activities in this context refer to the structured processes by which organizations collect, process, validate, and submit information to demonstrate adherence to regulatory requirements.

These activities are rarely isolated. They are typically **embedded within broader business processes** such as procurement, production, service delivery, risk management, and sustainability monitoring. As a result, compliance reporting often involves a complex interplay between dedicated regulatory technology solutions (e.g., reporting platforms, compliance software) and multifunctional sector and business specific systems. The analysis therefore aims to trace the direct and indirect linkages between technology adoption and regulatory reporting obligations, while recognising that compliance functions are not always served by dedicated tools.

The specific reporting activities tend to be sector specific as they are inherently related to the type of compliance requirements stemming from the legislative framework. However, in general, the relevant activities included in the analysis are the following:

- **Data collection and recordkeeping:** Gathering and maintaining records required for compliance, such as technical documentation, operational logs, traceability data, and audit trails.
- **Data validation and quality assurance:** Checking data accuracy, completeness, and consistency against regulatory thresholds and standards.

- **Preparation and submission of reports:** Compiling required information into prescribed formats (e.g., XML, CSV, PDF), and submitting to authorities via digital portals, email, or integrated platforms.
- **Incident and event reporting:** Logging and notifying authorities of adverse events, non-compliance, or safety issues, often through dedicated vigilance or surveillance systems.
- **Monitoring and audit readiness:** Ensuring that records and systems are prepared for inspection, audit, or verification by competent authorities.
- **Continuous compliance management:** Ongoing activities such as updating records, responding to regulatory changes, and implementing corrective actions.

As shown in Section 3.2.3. below, each sectorial analysis contains a detailed description of both the requirements and the modalities of reporting activities.

3.2.2 Overview of type of technological solutions included in the analysis

Reporting activities are supported by a spectrum of technological solutions, which can be classified according to their degree of automation and integration within organizational processes. These technologies have been grouped into three main categories: **standard techniques**, **advanced techniques**, and **automated techniques**.

Standard techniques

Standard techniques encompass tools and methods that **rely primarily on manual processes and human intervention**. These solutions are typically used for basic data collection, recordkeeping, and reporting, and are characterized by limited automation and integration. Standard techniques vary significantly across sectors, but some examples can be provided below:

- **Paper-based records:** Handwritten forms, printed documentation, physical logbooks.
- **Spreadsheets:** Microsoft Excel, Google Sheets, LibreOffice Calc used for manual data entry, tracking, and basic calculations.
- **Word processing documents:** Microsoft Word, PDF files for reports, technical documentation, or regulatory submissions.
- **Manual submission:** Sending reports and supporting documents via email or physical mail to authorities.
- **Basic file storage:** Local folders or shared drives for archiving compliance documentation.

Advanced techniques

Advanced techniques refer to the use of **specialized software applications and integrated digital systems** that facilitate the management of compliance-related data and reporting processes. These solutions are designed to move beyond basic manual methods by introducing **structured digital workflows**, centralized data management, and partial automation of routine tasks. However, while they streamline and standardize many aspects of compliance, they still require some degree of human oversight for configuration, validation, and interpretation. Advanced techniques typically enable organizations to collect, organize, and store compliance data in a systematic way, often integrating information from multiple business functions or operational areas.

A key feature of advanced techniques is that **they are usually not exclusively oriented toward compliance and reporting**. Instead, these systems often underpin several aspects of broader business processes—such as procurement, production, quality management, finance, or human resources—which are related or integrated with reporting functions. As a result, compliance activities are frequently embedded within multifunctional platforms that support both operational and regulatory needs.

Some advanced techniques are **horizontal technologies and software** which are used across different sectors such as:

- **Compliance software suites:** Platforms specifically designed to manage regulatory data, generate compliance reports, and maintain documentation (e.g., Rimsys, RAMS, Formwork by OpenRegulatory).
- **Enterprise Resource Planning (ERP) systems:** Integrated business management software (such as SAP S/4HANA, Oracle Fusion Cloud) that includes compliance modules for finance, HR, production, and reporting.

In other cases, advanced technologies and tools are sector specific and designed to integrate with particular business processes, reflecting the unique regulatory requirements and operational workflows of each industry. These sector-specific systems are developed to address the compliance and reporting needs that arise from the legislative framework governing the sector, and they often form part of broader digital transformation strategies within organizations.

Automated techniques

Automated techniques represent the **highest level of technological advancement in reporting and compliance activities**. In principle, these solutions are designed to minimize manual intervention by leveraging emerging technologies that can execute, monitor, and validate compliance processes with limited human oversight. Automated techniques are typically implemented to address the increasing complexity, volume, and frequency of regulatory reporting obligations, and are especially relevant for organizations seeking to enhance efficiency, accuracy, and scalability in their operations. Like advanced techniques, these tools **are often not exclusively designed for reporting and compliance**. Indeed, they are frequently integrated with broader business systems and processes, supporting multiple operational and regulatory functions. Some examples of the underlying technologies used for automated reporting is provided below.

- **Robotic Process Automation (RPA):** Software robots automate routine, rule-based tasks such as extracting data from source systems, generating compliance reports, and submitting documents to regulatory portals.
- **Artificial Intelligence (AI) and Machine Learning (ML):** Algorithms analyse large datasets, detect anomalies, structure unstructured information, and support predictive compliance monitoring (e.g., automated review of clinical notes, risk detection).
- **Blockchain and Distributed Ledger Technologies:** Decentralized platforms create tamper-proof, transparent, and auditable records of compliance-related data, enabling secure traceability and verification across multiple parties.

3.2.3 Structure of the sectorial analysis

Although the selected sectors display inherent differences, the analysis has been carried out following a common structure which enhances consistency and potential comparability of the results. While the level of detail highly depends on sector specific features (such as data availability, use of reporting tools, stakeholder engagement), the table below presents the overall content of the analytical sections.

Table 17. Structure of the sectorial analysis

Section	Sub-sections	Purpose and content
Overview of the sector	Definition	To establish the boundaries and formal definition of the sector under analysis, ensuring clarity about what is included and excluded. <i>Official classification of the sector (e.g., NACE codes, regulatory definitions).</i> <i>Description of the main activities, products, or services that characterize the sector.</i> <i>Explanation of sector boundaries, including any relevant exclusions (e.g., what is not considered part of the sector).</i>
	Size, value, employment	To provide context for the sector's economic and social significance. <i>Key statistics on sector output, gross value added, and market share.</i> <i>Employment figures, workforce characteristics, and demographic trends.</i> <i>Key trends and outlooks</i>
Regulatory landscape, reporting and compliance activities	Legal and regulatory framework	To identify and describe the main legislative acts, regulations, and standards that govern compliance and reporting in the sector. <i>List and summary of the five selected EU legislations.</i> <i>Full legal analysis (in Annex).</i> <i>Reference to additional relevant regulatory sources</i>
	Reporting and compliance activities	To describe the specific compliance and reporting obligations imposed by the regulatory framework, and how these are operationalised in practice Overview of compliance activities for each selected EU legalisation Description of processes and procedures
	Overview of existing challenges	To summarise the main challenges and pain points encountered by stakeholders in meeting compliance and reporting obligations under the current regulatory framework. Content: Identification of key sources of administrative burden, such as time spent on paperwork, recordkeeping, and responding to inspections. Analysis of the perceived issues and challenges in compliance mechanisms, including concerns about over-reliance on digital or automated controls and the risk of errors or misinterpretation.
Overview of reporting and compliance techniques	Description of techniques and tools	To categorise and describe the main types of technological solutions used for reporting and compliance, focusing on their level of automation and integration. Content: Neutral description of standard, advanced, and automated techniques. Explanation of the technological principles underlying each category. Concrete examples of tools or systems, illustrating their typical use and integration with business processes.
	Current use of reporting	To analyse how the different techniques and tools are actually used in practice within the sector.

	techniques	<i>Summary of adoption rates and usage patterns for each technique, based on survey, interview, or desk research evidence.</i> <i>Comparative tables or qualitative summaries showing prevalence across regulatory frameworks.</i> <i>Identification of predominant, supporting, or marginal techniques.</i>
Challenges, barriers, and qualitative cost-benefit analysis	Overview of challenges per reporting technique	To provide a detailed analysis of the main obstacles and barriers associated with each reporting and compliance technique. <i>Technique-by-technique overview of challenges, based on desk research and consultations</i> <i>Comparative summary of barriers for different categories of tools.</i>
	Barriers to adoption	To analyse the main obstacles that hinder the uptake and effective use of compliance and reporting tools and techniques within the sector. <i>Identification of barriers, such as labour intensity, financial costs, and technical requirements.</i> <i>Stakeholder perceptions on usability and scalability</i>
	Qualitative cost-benefit analysis	To evaluate the costs and benefits associated with different reporting and compliance techniques, supporting evidence-based recommendations. <i>Qualitative or quantitative assessment of costs (such as labour, licensing, integration, training).</i> <i>Analysis of benefits (such as efficiency, accuracy).</i> <i>Comparative tables or figures summarising findings for each category of technique.</i>
Policy recommendations		To synthesise findings and propose actionable recommendations for improving compliance and reporting practices in the sector. <i>Concise summary of the main findings and insights from the analysis.</i> <i>Recommendations for potential measures to address identified challenges,</i>

3.3 Agriculture

3.3.1 Overview of the sector

3.3.1.1 Definition

The agricultural sector in the European Union (EU) is defined according to the Statistical Classification of Economic Activities in the European Community (NACE Rev. 2). Specifically, it corresponds to Section A (primary sector: agriculture, forestry and fishing), Division 01, entitled “Crop and animal production, hunting and related service activities.”³⁷⁶ This definition encompasses all **farming activities focused on crop cultivation and animal husbandry, along with associated service activities directly related to agricultural production** (for example, agricultural contracting or farm management support). It even includes activities like hunting and trapping as part of the sector. Notably, this classification excludes any further processing of agricultural products beyond the farm gate – such as food manufacturing – which is handled in separate industrial sectors.³⁷⁷ The agriculture sector as defined above is therefore distinct from forestry (Division 02) and fisheries (Division 03), covering only those activities involved in producing primary agricultural goods.³⁷⁸

³⁷⁶ European Commission: Eurostat, *NACE Rev. 2 – Statistical classification of economic activities in the European Community*, Publications Office, 2008.

³⁷⁷ Ibid.

³⁷⁸ Ibid.

Within this broad sector, there are three key subsectors of economic activity:

- **Crop cultivation:** This includes the growing of field crops and horticultural products – for example, cereals (such as wheat and maize), oilseeds, fruits, vegetables, and other crops grown in open fields or greenhouses. These crop products form a large component of EU agriculture by volume and value.
- **Livestock farming:** This covers the breeding and rearing of farm animals for meat, dairy, eggs, and other animal products. Major categories include cattle (for beef and milk production), pigs, poultry (chickens and other birds for meat and eggs), sheep and goats, among others. Livestock farms are widespread across the EU and contribute significantly to agricultural output, particularly through products like milk and pork.
- **Agricultural support services:** This subsector encompasses ancillary activities that support farming itself. It includes services such as crop harvesting and planting contract work, farm maintenance and management services, veterinary and animal husbandry services, and other related support. Hunting and trapping activities are also classified here as part of agriculture's related services in the NACE framework. These services facilitate core crop and livestock production, though they represent a smaller share of the sector's value.

3.3.1.2 Size, value, employment

Agriculture remains a significant part of the EU economy, though its relative weight has gradually declined over the decades. In 2023, the **total value of agricultural output in the EU was approximately EUR 537.1 billion** (in basic prices).³⁷⁹ This figure represents the gross output of all farms in the EU, encompassing the value of crop products, animal products, and agricultural services (as well as certain inseparable secondary activities on farms). After accounting for the costs of inputs used in production, the EU's agriculture sector generated a gross value added (GVA) of about EUR 223.9 billion in 2023.³⁸⁰ This means agriculture contributed roughly 1.3% of the EU's Gross Domestic Product (GDP) in 2023.³⁸¹

In terms of composition, output is fairly balanced between crop and livestock production. Just over half of the total output value (around 51%) came from crop products in 2023 – equivalent to roughly EUR 274 billion, with particularly high outputs from categories like cereals and horticultural products.³⁸² Almost 40% of output value was generated by animals and animal products (about EUR 214 billion, driven largely by the dairy and meat segments)³⁸³. The remaining ~9% of output value derived from agricultural services and secondary (non-agricultural) activities on farms.³⁸⁴ This breakdown highlights that crop cultivation and livestock rearing together accounted for the vast majority of agricultural production by value in 2023.

The EU's agriculture sector also employs a **substantial workforce, albeit one that has been shrinking over time**. To account for the prevalence of part-time and seasonal work in farming, Eurostat measures agricultural labour input in Annual Work Units (AWUs) – a

³⁷⁹ Eurostat, *Performance of the agricultural sector*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Performance_of_the_agricultural_sector

³⁸⁰ Ibid.

³⁸¹ Ibid.

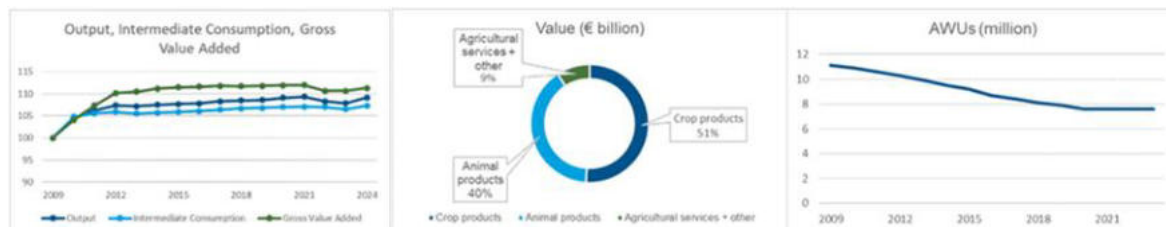
³⁸² Ibid.

³⁸³ Ibid.

³⁸⁴ Ibid.

metric that equates the total work performed to a number of full-time equivalent jobs.³⁸⁵ By this measure, the agricultural labour input in 2023 was about 7.6 million AWUs, meaning the work done was equivalent to roughly 7.6 million full-time workers across the EU.³⁸⁶ This figure reflects the many small or family-operated farms and part-time farmers in Europe, converted into full-time job equivalents for statistical purposes.

Figure 13. The European Agriculture Sector – Key Facts³⁸⁷



Importantly, **agricultural employment has been on a long-term downward trend**. The total volume of agricultural labour in 2023 was about 1.8% lower than in 2022,³⁸⁸ continuing a steady decline observed over the past decade (on average –2.6% per year between 2008 and 2023).³⁸⁹

3.3.2 Regulatory landscape, reporting and compliance activities

3.3.2.1 Legal and regulatory framework

This overview gathers the main reporting and compliance duties that apply to agricultural actors operating in the European Union. It outlines, in broad terms, the obligations that farmers, cooperatives, and agribusinesses must fulfil under the CAP Strategic Plans Regulation (EU 2021/2115), the CAP Financing, Management and Monitoring Regulation (EU 2021/2116), the EU Deforestation Regulation (EUDR), the Plant Protection Products Regulation (EC No 1107/2009), and the Corporate Sustainability Reporting Directive (CSRD). The full legal review is available in Annex 3.1.

Deforestation Regulation (Regulation (EU) 2023/1115)

The Deforestation Regulation establishes **due diligence obligations for operators and traders placing certain commodities**—such as cattle, cocoa, soy, and palm oil—on the EU market. All such products must be deforestation-free and legally produced.

Operators are required to **collect and verify geolocation data** of production plots, assess risk levels, and adopt mitigation measures where necessary. They must also **submit an annual due diligence statement** (Article 12) summarising their compliance efforts and results. A simplified **due diligence procedure** is foreseen for products originating from countries classified as low risk (Article 13). All relevant **records must be kept for five years** and made available for inspection.

³⁸⁵ Ibid.

³⁸⁶ Ibid.

³⁸⁷ The data on agricultural output, intermediate consumption, and gross value added are sourced from Eurostat's Economic Accounts for Agriculture (online code: *aact_eaa05*). The breakdown of output value by sector is based on Eurostat's agricultural production statistics (online code: *aact_eco01*). Employment figures measured in Annual Work Units (AWUs) come from Eurostat's Agricultural Labour Input dataset (online code: *aact_ali02*).

³⁸⁸ Ibid.

³⁸⁹ Ibid.

While the direct legal obligations apply to operators and traders, upstream suppliers—including farmers—may be required to provide traceability information and other supporting data to meet the regulation's compliance standards.

CAP Strategic Plans Regulation (Regulation (EU) 2021/2115)

The CAP Strategic Plans Regulation provides the legal framework for Member States to develop National Strategic Plans that translate EU-level policy goals into national conditionality requirements and support schemes. While the regulation does not impose direct reporting duties on individual farmers, it sets eligibility criteria for CAP funding that national authorities must verify. These include compliance with **Good Agricultural and Environmental Conditions (GAEC)** and **Statutory Management Requirements (SMRs)**, as listed in Annex III.

Farmers must meet these **conditionality requirements** to access CAP payments. This involves adhering to environmental, climate, and agricultural standards, such as maintaining permanent grassland, implementing crop diversification, or respecting water protection zones. Although the obligations are harmonised at EU level, Member States retain flexibility in how they operationalise reporting formats and verification systems.

The regulation also allows Member States to establish specific **support schemes**, such as complementary income support for young farmers or participation in eco-schemes. These schemes may require additional evidence of qualifying practices during the application process.

CAP Financing, Management and Monitoring Regulation (Regulation (EU) 2021/2116)

The CAP Financing, Management and Monitoring Regulation lays out the administrative architecture for implementing and controlling CAP support. It mandates the use of **geospatial aid application systems for area-based payments and sets requirements for effective verification, audit, and monitoring**.

Member States must establish integrated management and control systems, such as the **Integrated Administration and Control System (IACS)**, which includes tools like the Land Parcel Identification System (LPIS) and the Geo-Spatial Aid Application (GSAA). These systems are used to ensure that declared land parcels, activities, and interventions comply with CAP conditions. The regulation also defines rules for on-the-spot checks, remote sensing, and recordkeeping obligations to support CAP expenditure control.

Although the regulation sets out broad technical and procedural requirements, its implementation varies across Member States, depending on the configuration of national systems and platforms.

Plant Protection Product Regulation (Regulation (EC) No 1107/2009)

The Plant Protection Product Regulation governs the authorization, use, and monitoring of pesticides (plant protection products) within the EU. Its primary aim is to ensure that only safe and effective pesticides are used, minimizing risks to human health, animal welfare, and the environment. Farmers are directly affected by this regulation as it determines which pesticides they can legally use on their crops. To comply, farmers must adhere to strict **guidelines on the application of authorized products**, including dosage limits, application methods, and pre-harvest intervals. They are also required to maintain records of pesticide use, which may be subject to inspection by national authorities.

The **authorization procedure** for plant protection products involves stringent reporting requirements to ensure the safety and efficacy of pesticides before they can be marketed or used in the EU. Applicants seeking authorization must submit a comprehensive **dossier to the relevant national or EU authority**, containing detailed scientific data on the product's chemical composition, toxicological profile, environmental impact, and efficacy. This includes studies on residues, effects on non-target organisms, and potential risks to human health and the environment. The dossier must also demonstrate

compliance with Integrated Pest Management principles and provide evidence that the product does not have harmful effects on groundwater, biodiversity, or ecosystems. During the evaluation process, applicants may be required to provide additional information or clarify data in response to requests from regulatory authorities. Once authorized, ongoing reporting obligations may include monitoring and reporting adverse effects, as well as periodic re-evaluation of the product's safety.

Corporate Sustainability Reporting Directive (Directive (EU) 2022/2464)

The Corporate Sustainability Reporting Directive (CSRD) primarily targets large companies, listed SMEs, and certain non-EU companies with significant operations in the European Union. Specifically, it applies to entities that meet specific size thresholds (e.g., balance sheet total exceeding EUR 20 million, net turnover exceeding EUR 40 million, or more than 250 employees) or are listed on regulated markets. Most farmers, particularly individual agricultural producers and small farming operations, fall outside the scope of the CSRD because they typically do not meet these size criteria or operate as listed entities. However, **larger agricultural businesses, such as agribusinesses or cooperatives** that meet the thresholds or are listed on regulated markets, will be required to comply. Additionally, farmers supplying to large companies covered by the CSRD may face indirect reporting requirements as part of their supply chain commitments.

If farmers fall within the scope of the CSRD, they will be required to comply with detailed **sustainability reporting obligations** as outlined in Articles 19a and 29a of the directive. Article 19a mandates that in-scope entities disclose information on environmental, social, and governance (ESG) matters, including their impact on sustainability and how sustainability issues affect their business. Article 29a further requires the use of standardized reporting frameworks, specifically the European Sustainability Reporting Standards (ESRS), ensuring consistency and comparability of data. The ESRS can be found in the Annexes to Commission Delegated Regulation (EU) 2023/2772. They require reporting on environmental topics like climate change, pollution, water and marine resources, biodiversity and ecosystems, as well as social issues like labour, affected communities and consumers. The ESRS also mandate reporting on business conduct.

3.3.2.2 Reporting and compliance activities

CAP Framework (Regulation (EU) 2021/2115 and Regulation (EU) 2021/2116)

Under the CAP framework, farmers and agricultural businesses are required to comply with a set of obligations related to **conditionality, income support, and voluntary environmental commitments**. These are implemented and monitored through the Integrated Administration and Control System (IACS), a central mechanism linking farmer declarations, geospatial tools, and verification systems.

Each Member State must establish an **IACS to manage CAP payments and ensure compliance**. Core components include the Land Parcel Identification System (LPIS) for referencing agricultural land, the Geo-Spatial Aid Application (GSAA) for area-based claims, the Area Monitoring System (AMS) using satellite data, and registers for beneficiaries, animals, and entitlements. Paying Agencies, which are national or regional departments designated by Member States and accredited by the European Commission, operate the system, disburse funds, and oversee administrative and control procedure.³⁹⁰ They are responsible for managing expenditure from the CAP's two funds (EAGF and EAFRD), ensuring that payments are legal, regular, and properly accounted for under the

³⁹⁰ https://agriculture.ec.europa.eu/common-agricultural-policy/financing-cap/cap-paying-agencies_en#:~:text=,agencies%E2%80%99%20annual%20accounts%20and%20is.

principle of shared management. All activities are documented through annual accounts and performance reports.

The IACS operates on an annual cycle. Farmers **submit a single aid application**—commonly via a national online portal—typically pre-filled with parcel data and entitlements. These applications undergo automated cross-checks (e.g. field size validation, overlap detection, animal count verification) to ensure consistency across IACS databases. **Compliance is assessed via a combination of administrative checks, physical inspections, and increasingly, satellite-based monitoring through the AMS.**³⁹¹ Since 2018, EU Member States have been required to implement the GSAA for 100% of area-based direct payment claims, replacing paper forms with interactive digital systems. These geographic information system (GIS)-based platforms allow farmers to declare parcels, select support schemes, and verify compliance with eligibility criteria using digital maps and pre-filled data. The systems are designed to flag inconsistencies in real time—such as overlapping boundaries or missing information— with the aim of reducing errors and improve LPIS accuracy.³⁹²

Each Member State implements the IACS according to its own administrative arrangements, resulting in varying degrees of digitalisation and support for farmers. In France, the TelePAC platform includes an interactive parcel map and the TelePAC Géophotos mobile app, which allows farmers to upload geo-tagged photos to support compliance. Authorities use satellite data (Copernicus) and AI-based classification to detect irregularities, requesting additional information only when anomalies are flagged.³⁹³ In Poland, the eWniosekPlus system has been the sole application channel since 2018. It offers pre-filled data, real-time validation, and extensive user support, helping to reduce errors and administrative burden despite initial challenges during its rollout.³⁹⁴ In Germany, each federal state maintains its own GSAA portal, all interoperable with regional LPIS databases; these systems often integrate with farm management software, allowing farmers to exchange parcel data across platforms.³⁹⁵ Some Member States have gone further in integrating CAP applications with wider e-government services. For example, the Estonia's e-PRIA system³⁹⁶ is highly digital-by-default, linking CAP applications with other e-government services and farm databases. In summary, each Member State (or region) has its own branded online application portal tailored to its farmers and administrative systems.

GSAA platforms also support compliance beyond the application phase. **Farmers must implement and document the practices required under Good Agricultural and Environmental Conditions (GAECs) and Statutory Management Requirements (SMRs)** throughout the year. These cover a wide range of areas, from soil protection and crop rotation to buffer zones, animal welfare, and food safety. Compliance with

³⁹¹ https://agriculture.ec.europa.eu/common-agricultural-policy/financing-cap/assurance-and-audit/managing-payments_en.

³⁹² European Court of Auditors (2016), *Special Report No 25/2016: The Land Parcel Identification System – a useful tool to determine the eligibility of agricultural land, but its management could be further improved*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52016SA0025>.

³⁹³ <https://www.tools4cap.eu/tool/area-monitoring-system-3str-telepac-geophotos/>.

³⁹⁴ European Commission, *Analysis of administrative burden arising from the CAP*, 2024: https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cmef/regulation-and-simplification/analysis-administrative-burden-arising-common-agricultural-policy_en.

³⁹⁵ <https://www.nextfarming.de/hilfe/mehrfachantrag-und-schlagimport/mehrfachantrag-und-schlagexport/>.

³⁹⁶ <https://ec.europa.eu/digital-building-blocks/sites/pages/viewpage.action?pageId=533365952>.

complementary income support schemes or voluntary agri-environmental measures may require additional documentation or multi-annual commitments. To demonstrate compliance, farmers must maintain detailed and up-to-date records—including fertiliser and pesticide use, sowing and harvesting dates, grazing periods, and environmental practices. Records may be kept in printed logbooks, spreadsheets, or digital farm management systems and must be made available during audits. LPIS remains central to monitoring declared land areas and eligibility, and farmers are expected to validate parcel data annually and report any changes.³⁹⁷

To support remote verification, some authorities have introduced **LIDAR for field characterisation**, especially in complex terrains. Since 2023, the Area Monitoring System (AMS) is mandatory across the EU, using **Sentinel satellite imagery from the Copernicus programme** to monitor practices such as mowing or fallow periods. Authorities can use this data to trigger alerts or request additional evidence, such as geo-tagged photos via GSAA, when potential non-compliance is detected. This marks a shift from episodic inspections to continuous monitoring, improving both oversight and efficiency.³⁹⁸

Deforestation Regulation (Regulation (EU) 2023/1115)

To comply with the EU Deforestation Regulation, operators and traders placing or exporting relevant agricultural products—such as cattle, soy, palm oil, cocoa, coffee, rubber, or wood—must carry out a series of ongoing compliance and reporting activities to ensure that products are deforestation-free and legally produced.

Day-to-day compliance begins with **collecting detailed information about the product and its origin**, including the geolocation coordinates of the land where it was grown or harvested, along with documentation proving legal land use, compliance with national environmental laws, and other production-related authorisations. This information must then be assessed **against EUDR risk indicators**. Operators must evaluate the likelihood of non-compliance, taking into account supply chain complexity, supplier track records, and country-specific risks. In many cases, this involves maintaining batch-level traceability records, consulting satellite imagery, and using supplier questionnaires or third-party audits—although these tools do not substitute the operator's legal responsibility.

Where risks are identified, operators must adopt **mitigation measures** before placing the product on the market. These may include seeking additional documentation, conducting on-site checks, or excluding certain supply chains. All decisions must be documented within a structured due diligence system.³⁹⁹

Plant Protection Product Regulation (Regulation (EC) No 1107/2009)

Under the Plant Protection Product Regulation, professional users of plant protection products (PPPs) must carry out a range of **documentation and operational tasks to ensure that PPPs are used safely and exclusively under authorised conditions**, in compliance with EU and national legislation.

³⁹⁷ European Commission, *Analysis of administrative burden arising from the CAP*, 2024. https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cmef/regulation-and-simplification/analysis-administrative-burden-arising-common-agricultural-policy_en.

³⁹⁸ Ibid.

³⁹⁹ European Union (2023), Regulation (EU) 2023/1115 of the European Parliament and of the Council of 31 May 2023 on the making available on the Union market and the export from the Union of certain commodities and products associated with deforestation and forest degradation, OJ L 150, 9.6.2023. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32023R1115>.

A key obligation is the maintenance of detailed **application records**. For each use of a PPP, the user must log the product name, crop and plot treated, dose and volume, date and time of application, method and equipment used, as well as buffer zones, weather conditions, and any relevant re-entry intervals. These records must be completed promptly and kept for at least three years (or longer, if required nationally), and made available during inspections or audits.

Compliance also includes proper **storage and handling of PPPs**. These must be kept in locked, ventilated facilities, clearly labelled to prevent environmental or health risks. Containers must be regularly checked for damage or expiry, and leftover products or rinsing liquids disposed of in line with environmental regulations.

Users are further required to **calibrate and maintain spraying equipment**, as set out in Directive 2009/128/EC. Calibration ensures the correct dosage is applied, preventing both underuse (ineffective control) and overuse (excess residues). This involves maintaining service logs, inspecting tanks, and replacing worn nozzles.

Before using a new product or changing crops, farmers must check that the PPP is authorised for that specific crop and growth stage and complies with **pre-harvest intervals and maximum residue limits (MRLs)**. This often requires consulting product labels and online authorisation registries maintained by national food safety authorities.

At the commercial level, companies placing PPPs on the market must follow stringent **authorisation procedures**, including the submission of scientific dossiers for approval or renewal. These dossiers cover toxicological, ecotoxicological, environmental fate, and efficacy data, and are assessed by a Rapporteur Member State and the European Food Safety Authority (EFSA). Post-authorisation obligations include reporting adverse effects or product inefficacy and taking corrective measures.

In operational terms, the regulation demands continuous supply chain monitoring, regular communication with suppliers, and thorough documentation of compliance decisions.⁴⁰⁰

Corporate Sustainability Reporting Directive (Directive (EU) 2022/2464)

Under the Corporate Sustainability Reporting Directive (CSRD), large agricultural businesses are required to include detailed **sustainability disclosures in their annual management reports**, covering environmental, social, and governance (ESG) factors in line with the European Sustainability Reporting Standards (ESRS). In practice, this entails continuous data collection, validation, and reporting throughout the year—well beyond a single annual exercise. The CSRD is being phased in and is not yet fully applicable; small farmers remain outside its direct scope.

On a day-to-day basis, agricultural enterprises must **monitor key environmental indicators** such as greenhouse gas emissions, fertiliser and pesticide use, water consumption, and biodiversity-related measures. These data are typically captured using digital farm management systems or sustainability software, which help centralise and standardise recordkeeping.

To meet the CSRD's value chain transparency requirements, agribusinesses must also **collect relevant ESG information** from upstream suppliers, including individual farmers. This often involves emissions estimates, land use data, or evidence of sustainable practices. Ensuring data accuracy and traceability is critical, as CSRD reporting is subject

⁴⁰⁰ European Union (2009), Regulation (EC) No 1107/2009 of the European Parliament and of the Council of 21 October 2009 concerning the placing of plant protection products on the market and repealing Council Directives 79/117/EEC and 91/414/EEC, OJ L 309, 24.11.2009. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32009R1107>.

to external assurance. Companies must therefore establish internal controls—such as automated checks, digital logs, and audit trails—to validate and document their inputs.

Once verified, the data feed into CSRD-compliant **reports** that combine quantitative indicators (e.g. CO₂-equivalent emissions, hectares under biodiversity actions) with qualitative information (e.g. sustainability risks, strategies, and governance structures). Although submitted annually, these reports require year-round monitoring and regular updates, especially in response to regulatory or operational changes. Many businesses rely on integrated ESG dashboards to track progress and prepare disclosures efficiently.⁴⁰¹

3.3.2.3 Overview of existing challenges

Farmers across Europe have consistently highlighted the **administrative burden** involved in demonstrating compliance with EU legislation. This includes preparing and submitting CAP applications, maintaining detailed records, responding to eligibility checks, and ensuring readiness for inspections. Such tasks often fall outside the core agricultural activities and require significant time, planning, and familiarity with increasingly technical procedures. Findings from the European Commission's 2024 EU-wide survey highlight the extent of this burden. Around 33% of farmers reported spending more than six full working days annually on administrative tasks related to CAP applications, with an additional 24% spending five to six days, and 38% dedicating one to four days.⁴⁰² Farmers' organisations report that compliance-related tasks often spill over into evenings, requiring farmers to complete digital paperwork, update parcel maps, and upload documents after full days in the field. Despite this digital transition, CAP compliance continues to involve a wide range of recurring tasks—from land declaration and data validation to practice implementation and recordkeeping. Tools such as GSAA, LPIS, and AMS have reduced administrative burden and improved control mechanisms, but farmers must still engage regularly with regulatory systems and maintain operational familiarity with both compliance rules and digital platforms.

A further challenge relates to the **growing complexity and fragmentation of compliance processes**. Farmers are frequently required to report the same information through multiple platforms or instruments, such as conditionality declarations, eco-scheme commitments, and agri-environmental schemes. In several cases, farmers' organisations noted that national IT systems do not communicate effectively with one another, forcing farmers to manually re-enter data already provided elsewhere. This duplication not only increases administrative workload but also raises the risk of inconsistencies or errors, which can trigger inspections or payment reductions.⁴⁰³ Farmers highlighted that this situation is compounded by dynamic eligibility criteria and late policy adjustments, such as changes to GAEC standards or eco-scheme rules, which require farmers to revise plans mid-season and adapt their records accordingly. These shifting requirements contribute to regulatory uncertainty and undermine long-term planning.⁴⁰⁴ One survey stakeholder

⁴⁰¹ European Union (2022), Directive (EU) 2022/2464 of the European Parliament and of the Council of 14 December 2022 amending Regulation (EU) No 537/2014, Directive 2004/109/EC, Directive 2006/43/EC and Directive 2013/34/EU, as regards corporate sustainability reporting, OJ L 322, 16.12.2022. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32022L2464>.

⁴⁰² European Commission (2024). *Simplification survey sheds light on specific issues faced by farmers across the EU*. https://agriculture.ec.europa.eu/media/news/simplification-survey-sheds-light-specific-issues-faced-farmers-across-eu-2024-04-12_en.

⁴⁰³ EU CAP Network (2025). *Study on Simplification of the Common Agricultural Policy: Analysis of administrative burden and recommendations*.

⁴⁰⁴ Based on interviews with representatives of farmers' associations conducted for this study.

specifically highlighted this issue by calling for “machine readability of data” and “full interoperability” across systems, underscoring the need for agreed EU-wide data standards.⁴⁰⁵

Finally, a recurring concern relates to the perceived **lack of flexibility and responsiveness** in current compliance mechanisms. Interviewees expressed unease about the increasing reliance on remote controls, geo-tagged photos, and algorithm-based checks, which often leave little room for contextual explanation. Farmers worry that the absence of human assessment can lead to penalties based on inaccurate or incomplete digital evaluations. This is especially problematic for farms operating in atypical or complex environments, where automated tools may misinterpret landscape features, cropping patterns, or natural elements not easily captured by satellite imagery or automated models.⁴⁰⁶

3.3.3 Overview of reporting and compliance techniques

3.3.3.1 Description of techniques and tools

Standard tools for data compliance in the agriculture sector

Standard techniques for compliance in agriculture include simple tools used to record farm activities and meet regulatory requirements. These include **handwritten notebooks, paper forms, and digital spreadsheets**. Handwritten records are often used to note things like fertiliser use, pesticide treatments, or planting dates, but they can be hard to organize or store over time.

Digital spreadsheets offer a more structured approach. Tools like Microsoft Excel enable farmers to record operational data (fertilizer doses, pesticide applications, seeding dates, livestock movements) in predefined fields. Compared to handwritten records, spreadsheets support better consistency, easier archiving, and simplified sharing. They also offer built-in functions (drop-down menus, conditional formatting, basic formulas) that reduce the likelihood of mistakes and improve compliance readiness. Nonetheless, they still require careful manual data entry and a certain level of digital literacy to avoid inaccuracies.

To support farmers in using these tools more effectively, cooperatives, advisory services, and public networks have developed **pre-filled or customizable spreadsheet templates** aligned with national CAP compliance requirements. These templates cover areas such as crop rotation records, integrated pest management logs, and nutrient application tracking. Although not integrated with external databases or automated systems, these spreadsheet-based techniques serve as a practical and scalable compliance solution – particularly for farms in transition towards more advanced digital systems. It should be noted, however, that when using manual or spreadsheet records, farmers (or their advisors) often need to re-enter or upload this information into national CAP compliance portals during annual subsidy applications or inspections.⁴⁰⁷

Advanced tools for data compliance in the agriculture sector

Advanced tools support the collection of data, the organisation of records, and the generation of reports—tasks that were traditionally managed manually—thereby helping to reduce administrative burdens and improve accuracy. Key technologies in this space

⁴⁰⁵ Based on stakeholder survey responses.

⁴⁰⁶ Based on interviews with representatives of farmers' associations conducted for this study.

⁴⁰⁷ Basir, M. A., Buckmaster, D. R., Raturi, A., & Zhang, Q. (2024). *From pen and paper to digital precision: a comprehensive review of on-farm recordkeeping*. *Precision Agriculture*. <https://doi.org/10.1007/s11119-024-10172-7>.

include digital logbooks, and Farm Management Information Systems (FMIS). Each offers distinct capabilities to assist farmers in meeting compliance more efficiently.

Digital Logbooks

Digital logbooks are the **electronic equivalent of the traditional farm diary or record book**. Farmers have long been required to keep logs of activities (spraying, irrigation, animal treatments, etc.) to comply with environmental and safety standards and moving this practice to computers or mobile devices greatly improves efficiency. Digital logbooks can be standalone applications (e.g. a smartphone app solely for recording farm tasks) or part of larger farm management platforms. The transition from paper-based to digital record-keeping addresses the limitations of manual logs, which are labour-intensive, error-prone, and difficult to archive. Digital systems streamline these processes, providing structured and accessible tools that support better farm management and regulatory adherence.⁴⁰⁸

Advanced digital logbooks have evolved into comprehensive tools that help farmers efficiently manage compliance-related data. These tools, such as **Agri-Trak**⁴⁰⁹ or **Fieldmargin**⁴¹⁰ enable farmers to efficiently **record and organise compliance-relevant data** directly in the field, supporting both daily management and regulatory obligations. These tools are typically built around **mobile-first interfaces**, allowing users to log tasks such as **spraying, fertilising, seeding, or harvesting** as they are being carried out. Importantly, most systems function offline and synchronise automatically when a connection is re-established, ensuring that data can be captured even in remote areas with limited connectivity. This real-time functionality significantly reduces the likelihood of missed or inaccurate entries that can occur when relying on memory or end-of-day reporting.

Each activity entry is usually linked to a specific field or parcel and may include a range of metadata, such as timestamps, weather conditions at the time of the activity, names of operators or equipment used, and even photographs or notes. This level of detail supports traceability and provides strong evidence of compliance during inspections or audits. For example, a pesticide application can be recorded with the precise product, dosage, date, time, and location, while also documenting operator certification or re-entry intervals—all in a structured and easily retrievable format.

The ability to **geolocate field activities and maintain chronological, searchable logs** improves the reliability of records, making them both more useful for farm management and more robust in the context of CAP reporting, food safety certification, or environmental auditing. Furthermore, digital logbooks often incorporate standardised templates or pre-defined activity types, which help ensure that required compliance information is captured in a consistent manner, reducing omissions and easing the reporting burden. This structured, verifiable approach transforms compliance documentation from a burdensome task into a routine part of day-to-day farm operations.

In Fieldmargin, for example, farmers can log key activities such as pesticide applications, fertiliser use, mowing, or planting directly onto mapped fields, with each entry linked to a specific date, location, and operator. They can also attach photos, notes, and weather data to each task, creating a detailed and traceable record of operations that supports compliance with environmental and CAP-related obligations. They also have access to **reporting tools** that summarise field jobs, input usage, and crop rotations—allowing them

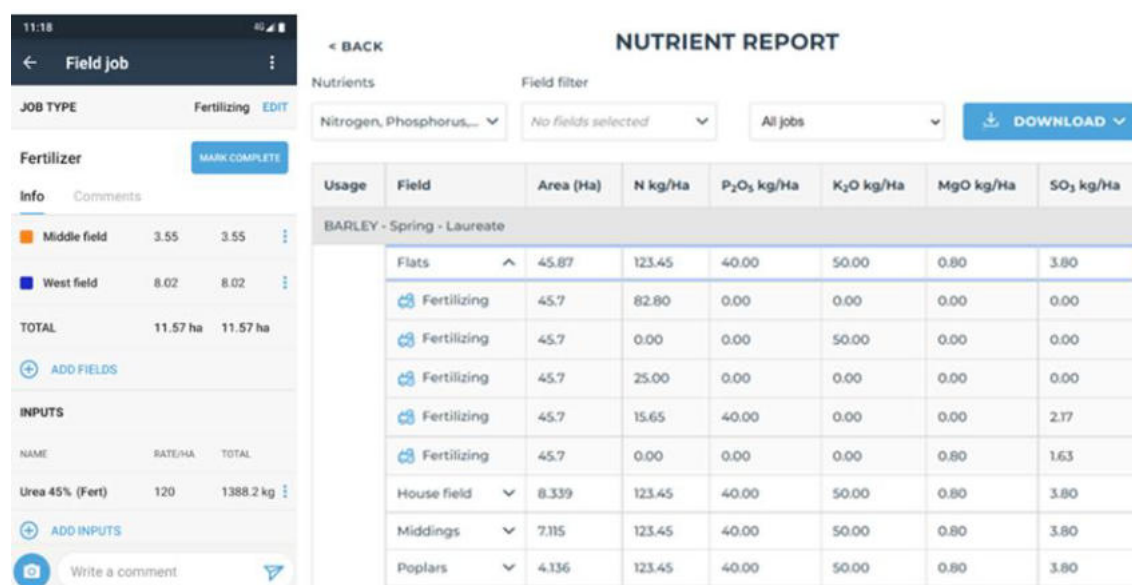
⁴⁰⁸ Ibid.

⁴⁰⁹ For further information: <https://www.agri-trak.com/>.

⁴¹⁰ For further information: www.fieldmargin.com.

to export structured records (e.g. as PDFs or spreadsheets) for inspections, organic certification, or subsidy applications. These features help demonstrate adherence to GAEC rules, IPM practices, and eco-scheme commitments.

Figure 14. Input Logging and Nutrient Reporting in Fieldmargin⁴¹¹



Farm Management Information Systems (FMIS)

Farm Management Information Systems (FMIS) are **comprehensive digital platforms designed to support farm operations and regulatory compliance**. Unlike basic digital logbooks, which focus on recording individual field tasks, FMIS integrate a wide range of functions—including crop planning, input management, machinery monitoring, inventory control, and compliance reporting—into a centralised system.⁴¹²

These systems **collect and synchronise data from multiple sources**, such as IoT sensors, GPS-equipped equipment, weather services, and satellite imagery. This integration allows farmers to plan and monitor agricultural activities while maintaining traceable records for inspections or audits. A typical FMIS interface displays mapped fields, activity logs (e.g. spraying, fertilising, harvesting), input stock levels, and performance indicators. Some platforms also include modules for livestock tracking, traceability logs, or audit trails relevant to CAP conditionality, GAEC standards, and plant protection product (PPP) regulations.

FMIS increasingly support automated data capture through technologies such as ISOBUS (ISO 11783), enabling seamless communication between machinery and software. Operational data—such as seeding rates or pesticide applications—can be directly transmitted and matched to specific parcels. This automation improves the accuracy and timeliness of compliance records.

Adoption of FMIS is growing. Global estimates suggest that around 20% of farmers currently use farm management software—making it the most widely adopted category of

⁴¹¹ Fieldmargin interface showing two core compliance functionalities. On the left, an input application record captures key data. On the right, a nutrient summary report allows monitoring of nitrogen, phosphorus, and potassium levels across fields.

⁴¹² 4Growth Project (2023). *Farm Management Information Systems Explained: A Closer Look. (FMIS overview and adoption trends)*. <https://4growth-project.eu/agriculture/farm-management-information-systems-explained-a-closer-look/#:~:text=Adoption%20Trends%20and%20Barriers>

agricultural technology.⁴¹³ In Europe, uptake remains limited among smaller holdings due to cost, complexity, and lack of technical support. One recent survey found that 46% of European farmers had adopted at least one digital tool, implying significant room for further FMIS adoption, especially among family-run or diversified farms.⁴¹⁴

Commercial platforms such as xFarm⁴¹⁵, AGRIVI⁴¹⁶, Agroptima⁴¹⁷, or 365FarmNet⁴¹⁸ explicitly promote their ability to simplify compliance tracking, facilitate audit readiness, and align farm data with regulatory requirements across the EU farming sector.⁴¹⁹ For example, xFarm is a FMIS that illustrates how compliance tracking is embedded into broader farm management. Through its web dashboard, users can view field maps, crop schedules, completed tasks, and receive alerts for anomalies—such as overdue spraying or low input stock. The platform allows real-time logging of field operations with specific details on input use (e.g. pesticide, fertiliser), operator name, application rate, and weather conditions. All records are automatically linked to the appropriate parcel and can be exported in standardised formats for inspections or subsidy applications. Farmers can also view analytics dashboards comparing input use over time or between fields, helping them stay within legal limits and improve resource efficiency.⁴²⁰ Beyond CAP-related reporting, xFarm also supports compliance with the CSRD by offering modules that calculate carbon footprint, water use, and other environmental indicators required under the European Sustainability Reporting Standards (ESRS). This enables agri-food companies to monitor ESG performance and prepare traceable, audit-ready sustainability reports.⁴²¹

⁴¹³ 4Growth Project (2023). *Farm Management Information Systems Explained: A Closer Look. (FMIS overview and adoption trends)*. <https://4growth-project.eu/agriculture/farm-management-information-systems-explained-a-closer-look/#:~:text=Adoption%20Trends%20and%20Barriers>.

⁴¹⁴ McKinsey & Co. (2023). *Agtech: Breaking down the farmer adoption dilemma. (Global agtech adoption statistics)*. <https://www.mckinsey.com/industries/agriculture/our-insights/agtech-breaking-down-the-farmer-adoption-dilemma>.

⁴¹⁵ For further information: <https://xfarm.ag/en>.

⁴¹⁶ For further information: <https://www.agrivi.com/en>.

⁴¹⁷ For further information: <https://www.agroptima.com/en/>.

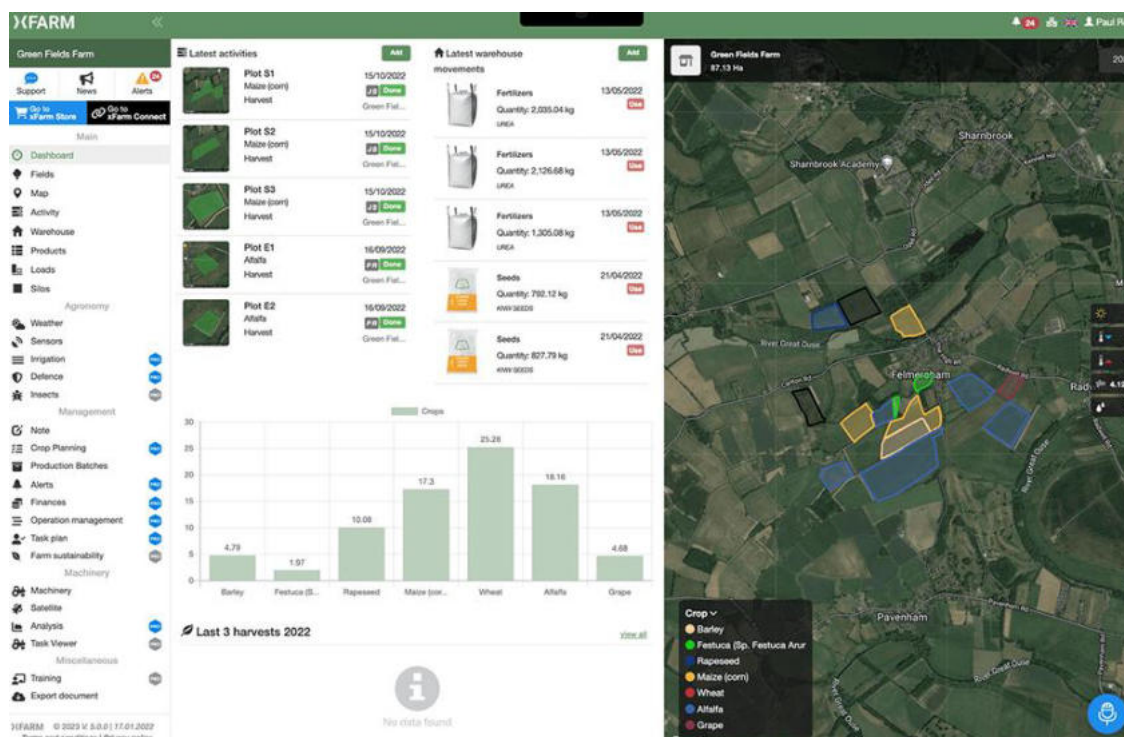
⁴¹⁸ For further information: <https://www.365farmnet.com/en/>.

⁴¹⁹ AGRIVI (2024). *Smart Farm Management Software*. <https://www.agrivi.com>; Agroptima (2024). *Field record keeping made simple*. <https://www.agroptima.com/en>; 365FarmNet (2024). *Your digital farm management solution*. <https://www.365farmnet.com/en>.

⁴²⁰ xFarm Technologies (2023). *Farm Management Software – Product Description*.

⁴²¹ For further information: <https://xfarm.ag/en/blog-posts/xfarm-come-strumento-di-reportistica-per-la-direttiva-sulla-rendicontazione-della-sostenibilita-aziendale-csrd>.

Figure 15. Overview of xFarm's Free Compliance and Management Functions⁴²²



Automated tools for data compliance in the agriculture sector

Automated tools represent the most advanced category of compliance technologies in agriculture. These tools aim to **reduce or eliminate the need for manual data entry by capturing and verifying information automatically**—often in real time—through sensors, machine integration, or distributed digital systems. Their use can enhance accuracy, reduce administrative burden, and facilitate audit-ready reporting.

While still emerging, one of the most promising technologies in this space is blockchain. The European Green Deal and Farm-to-Fork Strategy have heightened interest in blockchain as a means to improve sustainability reporting and build consumer trust.⁴²³ Blockchain systems offer a decentralised and tamper-evident way to log and verify compliance-related data, particularly in areas such as traceability, certification, and sustainable supply chain reporting. In 2024, the global market for blockchain in agriculture was valued at USD 352 million, with forecasts projecting substantial growth to over USD 5.6 billion by 2032.⁴²⁴ This trend highlights growing recognition and investments in blockchain technology across the agricultural sector. Nevertheless, adoption remains at an early stage, with most applications still in pilot or exploratory phases.

Blockchain's main advantage lies in its **capacity to create a shared, immutable record of compliance-related data**, which can be accessed by multiple actors—farmers, inspectors, certification bodies, cooperatives—without risk of manipulation or unauthorised changes. Each transaction or data entry (e.g. pesticide application, harvest date, input

⁴²² The xFarm dashboard provides access to 16 core functions—including field mapping, activity logging, input tracking, and weather monitoring.

⁴²³ https://blockchain-observatory.ec.europa.eu/events-pool/use-blockchain-technologies-agri-food-sector-2020-11-23_en?

⁴²⁴ Credence Research (2024). *Blockchain in Agriculture Market*. <https://www.credenceresearch.com/report/blockchain-in-agriculture-market>.

purchase) is recorded with a timestamp and cryptographic signature, forming a traceable, tamper-evident chain of events.⁴²⁵

One major area of application is compliance with the CAP. In principle, **blockchain systems could be used to record key field operations—such as fertilisation, mowing, or pesticide application—as they occur, using IoT sensors, GPS-equipped machinery, or FMIS platforms.** These entries could be logged on a permissioned blockchain ledger, accessible to farmers, public authorities, and certification bodies. By providing a tamper-evident audit trail of activities, blockchain could reduce the need for repeated declarations, lower verification costs, and ensure consistency between farm records and subsidy applications. Pilot projects have begun to explore this potential. In Spain, for example, a CAP-related initiative tested the integration of blockchain with Radio-frequency identification tags and IoT sensor data to automatically register irrigation cycles and machinery use, with data fed into a ledger accessible to subsidy-paying agencies. The system aimed to enable passive, real-time verification of compliance with CAP conditionality and eco-scheme requirements.⁴²⁶

Beyond the CAP context, **blockchain is also being explored and deployed to support compliance in other agri-food regulatory areas,** particularly in relation to traceability, certification, and sustainability reporting obligations. These applications are becoming increasingly relevant due to evolving legal requirements such as the EU Deforestation Regulation and the CSRD, which demand verifiable information on input origin, production practices, and environmental impact. A notable example is the EU-supported **AgriChains programme**, which works in 24 countries to help governments and agri-businesses build traceability and forest monitoring systems aligned with EUDR and due diligence legislation.⁴²⁷ Several EU-funded initiatives have also explored blockchain's compliance potential. One such example is the **SmartAgriHubs project**, which supported a series of blockchain experiments across Europe. Within this initiative, a Slovenian-led pilot applied the OriginTrail Decentralised Knowledge Graph (DKG) to strengthen trust and traceability in organic beef and dairy supply chains. By linking farm management data, certification records, and transport logs into a tamper-proof blockchain structure, the system enabled secure, verifiable compliance documentation across all stages of the chain.⁴²⁸ The **TRACE4EU project** is developing cross-border blockchain systems to verify agri-food certificates using decentralised digital identities.⁴²⁹ In parallel, the TrustEat project developed blockchain-compatible nano-sensors and capacity-building resources to support food safety monitoring.⁴³⁰

In the private sector, companies are operationalising blockchain **to support traceability, certification, and audit-readiness.** Platforms like **IBM Food Trust**, adopted by Carrefour and other European retailers, allow producers to record origin and quality data on a

⁴²⁵ Food and Agriculture Organization of the United Nations and International Telecommunication Union, 2019. *e-Agriculture in Action: Blockchain for Agriculture – Opportunities and Challenges*. <https://www.fao.org/3/ca2906en/CA2906EN.pdf>.

⁴²⁶ Martínez-Castañeda, M. & Feijoo, C. (2022). *The usage of blockchain in agri-food value chains: State of the art in Spain and an analysis from public perspective*. <https://www.sciencedirect.com/science/article/pii/S030859612300085X>.

⁴²⁷ For further information: <https://zerodeforestationhub.eu/projects/agrichains/>

⁴²⁸ For further information: <https://www.smartagrihubs.eu/latest-news/BLOG-IFOAM-Application-of-OriginTrail-to-the-organic-food>

⁴²⁹ For further information: <https://trace4eu.eu/>

⁴³⁰ For further information: <https://cordis.europa.eu/article/id/449259-building-a-safer-food-chain-starts-with-blockchain>

shared ledger, boosting transparency and consumer trust. Carrefour, for instance, lets consumers scan QR codes on products like poultry or dairy to view compliance and origin details directly on their phones.⁴³¹ Similarly, Certified Origins Italia uses **Oracle Blockchain** to ensure the traceability of extra virgin olive oil exports,⁴³² while **SAP's Farm-to-Consumer** blockchain modules are used by agri-food companies in Italy and Spain to connect farm ERP systems with retailers, reducing the cost and effort of compliance audits.⁴³³ Platforms like **TE-FOOD**⁴³⁴ and **Open Food Chain**⁴³⁵ support food safety compliance, while tools like **AgriDigital**⁴³⁶ and blockchain-linked crop insurance pilots offer automated solutions for contract enforcement and weather-indexed payouts.

In sum, blockchain is already being deployed in a variety of commercial compliance contexts—from input traceability to certification and insurance—while EU-funded public pilots continue to explore its longer-term potential for digitalised, cross-border regulatory systems. These developments suggest that blockchain could soon serve as a foundational infrastructure for trusted and automated compliance reporting in the EU agri-food sector.

3.3.3.2 Current use of reporting techniques

Compliance and reporting in agriculture rely on a diverse set of techniques, ranging from paper-based tools to fully automated systems. A qualitative analysis of survey responses, complemented by desk research and interviews, suggests that standard techniques continue to dominate, while advanced tools play a supporting role and automated tools remain marginal (Table 18).

Table 18. Use of reporting techniques for carrying out compliance and reporting activities

Techniques	Standard tools - Paper	Standard tools	Advanced digital logbooks	Farm Management Information Systems	Fully automated tools
Overall use	Predominant	Predominant	Supporting	Supporting	Marginal
Deforestation Regulation	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
CAP Strategic Plans Regulation	Predominant	Predominant	Supporting	Supporting	Marginal
CAP Financing, Management and Monitoring Regulation	Supporting	Supporting	Supporting	Supporting	Marginal
Plant Protection Products Regulation	Inconclusive	Inconclusive	Inconclusive	Inconclusive	Inconclusive
Corporate Sustainability Reporting Directive	Likely supporting	Likely predominant	Likely supporting	Likely supporting	Likely marginal

Source: The data used to compile this table were collected through a survey of industry associations.

⁴³¹ See: European Commission (2023). *EIP-AGRI seminar on blockchain technologies in agri-food*. Available at: <https://ec.europa.eu/eip/agriculture/en/event/use-blockchain-technologies-agri-food-sector.html>

⁴³² For further information: <https://www.oliveoiltimes.com/sponsored/certified-origins/110304>

⁴³³ For further information: <https://www.oiv.int/public/medias/8593/digital-trends-applied-to-the-vine-and-wine-sector.pdf>

⁴³⁴ For further information: <https://te-food.com/>

⁴³⁵ For further information: <https://www.openfoodchain.com/>

⁴³⁶ For further information: <https://www.agridigital.io/>

The comparative overview of current compliance and reporting techniques across the five selected EU regulations highlights the persistence of standard methods alongside the gradual uptake of more advanced systems. Across regulations, spreadsheet solutions (mainly Excel) represent the predominant digital tool, while logbooks and FMIS are used in a supporting role and fully automated systems remain marginal.

Paper-based tools remain significant overall, particularly under the CAP Strategic Plans Regulation, where farmers still rely on handwritten records, printed declarations, and sketch maps to demonstrate compliance with conditionality and eco-scheme requirements.

Spreadsheets, especially Excel, emerge as another significant technique, confirmed by survey responses identifying them as predominant for recording and reporting under the CAP Strategic Plans Regulation. Stakeholders also indicated their use under the Plant Protection Products Regulation for tracking pesticide authorisations and applications, while in the CSRD context spreadsheets were reported as a regular instrument to compile ESG data and prepare reports.

Survey responses also alluded to the use of custom-built in-house tools, though these were described as basic solutions limited to manual data entry and document storage. One stakeholder noted that such tools were mainly used for submitting data to public authorities, but they lacked advanced functions for integration, monitoring, or automated reporting. This reinforces the overall picture that, while farms may experiment with tailored solutions, these tend to replicate the limitations of spreadsheets rather than represent a step toward advanced digitalisation.

Advanced tools, including digital logbooks and Farm Management Information Systems (FMIS), **increasingly function as supporting techniques** across the regulatory framework. While survey evidence suggests their use is not yet widespread among farmers, it is likely to be more frequent among larger holdings, which are better positioned to benefit from their specific compliance functionalities under different regulations.

Fully automated tools, such as blockchain-based systems, **were consistently reported as marginal across all regulations**. This aligns with the broader picture of these technologies being at an early stage of deployment, with survey evidence confirming their limited role in current farmer practices.

One key conclusion that emerges on the use of tools for compliance is that standard digital tools—most notably spreadsheets—continue to dominate compliance-related tasks at the farm level. Their usefulness lies in the fact that they can be easily tailored to meet the specific needs of different production systems and regulatory requirements. For example, within CAP, cooperatives and advisory bodies often provide pre-filled spreadsheet templates that help farmers document crop rotations, pesticide use, and fertiliser applications in line with CAP conditionality requirements. In parallel, some farmers develop their own spreadsheet tools using programs like Excel or LibreOffice, tailoring them to their specific compliance and planning needs. These self-developed templates are sometimes shared informally within farming communities, reinforcing their value as a flexible and low-cost alternative.⁴³⁷ Survey data confirms this pattern: digital spreadsheets were reported as the only tool regularly used for CAP compliance, especially in area measurement and audit preparation.⁴³⁸ By contrast, more advanced or automated tools remain largely underutilised. This underuse is consistent across regulations and suggests that these tools are still perceived as either too complex, insufficiently integrated, or not yet delivering clear added value. Fully automated tools

⁴³⁷ Based on interviews with representatives of farmers' associations conducted for this study.

⁴³⁸ Based on stakeholder survey responses.

such as blockchain platforms or AI-powered assistants were almost entirely absent from reported practices. Even custom-built in-house software, where used, was limited to basic functions like submitting data to authorities via manual entry.⁴³⁹ More broadly, several interviewees noted that farmers are more inclined to adopt digital technologies when they directly support core agronomic activities—such as planting, irrigation, equipment automation, or remote sensing—rather than when they are solely designed for compliance. Tools that align with productivity goals are seen as a better use of limited time and resources, whereas compliance-specific technologies are often perceived as secondary, burdensome, or poorly integrated into daily operations.⁴⁴⁰ As one farmer put it, “we didn’t get into farming to do data entry”.⁴⁴¹ This helps explain why grassroots digital solutions—such as self-built templates or peer-shared tools—sometimes gain more traction than formal platforms.

3.3.4 Challenges, barriers, and qualitative cost-benefit analysis

3.3.4.1 Overview of challenges per reporting technique

Standard tools

Paper-based techniques remain **a common but problematic** approach to compliance and reporting in agriculture. Survey stakeholders unanimously stressed that these methods are highly time-consuming, requiring repeated manual work to complete CAP application forms, maintain pesticide treatment logs, or keep livestock registers. The reliance on handwriting not only prolongs data entry but also increases the risk of errors and inconsistencies. Farmers also face persistent difficulties with the storage, retrieval, and archiving of physical records. Paper documentation is prone to damage, loss, or incompleteness, and lacks the resilience of digital records for long-term traceability. Stakeholders further noted that the absence of standardised templates across CAP schemes and Member States adds to the burden, as farmers must adapt their record-keeping to fragmented or shifting formats. Duplication of effort was also identified as a key drawback, with many farmers required to later re-enter paper-based data into digital portals such as GSAA or national registries. These overlapping demands underline why paper remains a weak point in the compliance chain, contributing to inefficiency and frustration across farm types.⁴⁴²

Digital spreadsheets were viewed as a step forward from paper, but survey evidence suggests that **they carry their own set of limitations**. Stakeholders highlighted the incompatibility of spreadsheets with government portals and official compliance systems, which prevents seamless transfer of information and obliges farmers to manually re-enter the same data. Similarly, spreadsheets cannot automatically import information from sensors, machines, or FMIS, reducing their effectiveness in modern, digitalised farm environments. Although they offer flexibility, spreadsheets still depend heavily on manual data entry, which remains time-consuming and prone to error. Farmers also identified the risk of data loss or corruption as a significant concern, particularly where backup and archiving practices are weak. In practice, this means that spreadsheets often serve as an intermediary tool rather than a final reporting solution—useful for structuring and tracking

⁴³⁹ Ibid.

⁴⁴⁰ Ibid.

⁴⁴¹ Based on interviews with representatives of farmers' associations conducted for this study.

⁴⁴² Based on stakeholder survey responses.

information internally but ultimately limited by their lack of interoperability with regulatory systems.⁴⁴³

Advanced tools

Farm Management Information Systems (FMIS) are seen **as promising tools for streamlining compliance, yet their uptake remains constrained** by several practical challenges. Survey responses highlighted high software licence and subscription costs, which limit accessibility for smaller farms, alongside persistent gaps in digital literacy among users. Compatibility with national CAP reporting systems is another recurring obstacle: farmers noted that FMIS data cannot be exported seamlessly into GSAA or IACS, forcing manual duplication of entries. Concerns around data privacy and ownership, particularly when platforms rely on cloud-based storage, further discourage widespread adoption. In addition, integration with farm machinery is still uneven, requiring additional effort to reconcile data from tractors, sensors, or other equipment.⁴⁴⁴

Advanced digital logbooks face **many of the same constraints**. Farmers reported difficulties migrating historical paper records into digital systems, raising questions about the completeness of compliance documentation. Uncertainty also persists over whether authorities formally recognise digital logs as valid evidence during audits. As with FMIS, uptake is limited by poor internet connectivity, patchy training and support, and compatibility issues with government portals and farm equipment. These limitations are particularly acute for smaller or older farms, where digital skills are less developed and the benefits of investing in digital compliance systems are less immediately tangible.⁴⁴⁵

Automated tools

The main obstacle to the adoption of fully automated tools is **the absence of clear guidance and concrete reference cases** for their use in CAP compliance or sustainability reporting. Without practical examples, farmers and cooperatives struggle to see the benefits of blockchain-based traceability platforms or to integrate them into existing reporting routines. Although technical costs were reported as negligible in this sample, stakeholders highlighted greater challenges around usability and trust, which keep such tools largely experimental and at an early stage of adoption in agriculture.

Concerns over vendor lock-in were also raised, with two stakeholders noting the risks of dependency on proprietary platforms. This challenge is compounded by uncertainties around scalability, particularly in contexts requiring the management of large volumes of farm-level data across multiple actors in the supply chain. Interoperability is another limiting factor: blockchain systems are not yet aligned with national CAP portals or farm management software, which reduces their usefulness as a compliance tool.

Trust in the technology itself remains mixed. Some stakeholders expressed scepticism about the reliability and transparency of blockchain, with additional concerns about data security and confidentiality in distributed ledgers. Although blockchain is often promoted as a solution for auditability and traceability, these survey results suggest that the absence of regulatory recognition, practical integration pathways, and stakeholder confidence currently keeps such tools on the margins of agricultural compliance.⁴⁴⁶

⁴⁴³ Ibid.

⁴⁴⁴ Ibid.

⁴⁴⁵ Ibid.

⁴⁴⁶ Ibid.

3.3.4.2 Barriers to adoption

The analysis of survey responses highlights five main categories of barriers that affect the use of compliance and reporting tools in agriculture: resource intensity, losses and risks, human capital and digital skills, scalability and tailoring, and financial costs. These barriers manifest differently across techniques, shaping both the ease of adoption and their long-term sustainability (Table 19).

Table 19. Barriers to adoption of reporting techniques for compliance and reporting activities

Barrier	Standard tools - Paper	Standard tools - Spreadsheets	Advanced digital logbooks	Farm Management Information Systems	Fully automated tools
Resource intensive	High	High	Medium	Medium	Medium
Losses and risks	High	High	Medium	Medium	Medium
Human capital and digital skills	Low	Low	Medium	Medium	High
Scalability and tailoring	Low	High	Medium	Medium	High
Financial costs	High	Low	Medium	High	High

Source: The data used to compile this table were collected through a survey of industry associations.

Paper-based records and spreadsheets remain the most burdensome solutions. They are highly resource-intensive, requiring substantial time to complete forms, re-enter data, and maintain files. Their reliance on manual entry creates a high risk of errors, inconsistencies, and data loss, while their lack of scalability makes them ill-suited for modern compliance processes that increasingly rely on digital submission. Despite their low financial costs, the structural inefficiency and duplication of effort they generate make them unsustainable for farmers expected to meet complex, recurring CAP and sustainability obligations.⁴⁴⁷ However, their continued widespread use reflects certain strengths, as they remain adaptable to different farm contexts, can be customised for specific reporting requirements, and are easy to scale without requiring specialised digital infrastructure.

Advanced tools such as digital logbooks and Farm Management Information Systems (FMIS) **represent a clear improvement in terms of efficiency and structuring of information, but survey results show that they introduce new barriers of their own.** On the positive side, they reduce some of the risks of physical data loss or corruption that are common with paper and spreadsheets. However, the survey highlighted a different set of concerns: farmers reported uncertainty over whether digital logbooks are officially recognised for compliance purposes, as well as doubts about the security and ownership of data stored in cloud-based FMIS. This points to a trade-off between reduced risk of accidental data loss and heightened anxiety about privacy, confidentiality, and control over sensitive farm-level data. Several stakeholders linked this to a wider mistrust of external platforms, noting that limited training and poor connectivity can further erode confidence in relying solely on digital systems for compliance.⁴⁴⁸ One interviewee recounted being unable to submit a geo-tagged image to confirm crop presence due to the lack of signal on-site, resulting in a missed verification window. Other described being penalised after satellite tools misclassified land cover or crop type, with little opportunity to

⁴⁴⁷ Based on stakeholder survey responses.

⁴⁴⁸ Ibid.

contest automated errors or access user support.⁴⁴⁹ These connectivity challenges are compounded by digital skill gaps, particularly among older farmers. The average EU farmer is 57 years old, and many interviewees emphasised that this group often lacks the confidence or familiarity needed to use online platforms or smartphone apps effectively. Overall, this digital divide not only hinders access to advanced compliance tools but also contributes to growing inequalities within the sector. Survey responses also confirm that connectivity and training gaps are major barriers to using remote sensing and geo-tagging tools.⁴⁵⁰

Both FMIS and digital logbooks also face challenges in terms of digital skills and usability. Migrating historical paper records into digital systems was described as difficult, particularly for older farmers or smaller farms lacking IT support. Compatibility with CAP portals such as GSAA and IACS remains patchy, often forcing farmers to duplicate entries, while interoperability with machinery or sensors is still limited. These barriers mean that, even when the tools are adopted, their compliance value is undermined by the need for manual cross-checks and parallel recordkeeping. Financial costs further exacerbate these difficulties, with FMIS in particular requiring high licence and subscription fees that are manageable for large farms but prohibitive for smaller operations.⁴⁵¹

Fully automated tools, such as blockchain-based traceability platforms, remain marginal in practice and face the steepest adoption barriers. Survey stakeholders **pointed less to technical or infrastructure costs—rated as negligible at this stage—and more to issues of scalability, interoperability, and trust.** These platforms also demand high levels of digital expertise to operate and integrate, while trust in automated outputs remains fragile, with stakeholders expressing concerns over data security, transparency, and long-term reliability.⁴⁵² In addition, there is also a broader hesitation about shifting too much control to automated systems—especially when farmers do not understand how decisions are made or cannot easily correct mistakes. For some, the risk of “invisible errors” in automated systems is worse than the burden of manual compliance. Survey stakeholders expressed only partial trust in automated outputs, stating that significant manual checks remain necessary.⁴⁵³

In addition, there is a persistent **divide in capacity between small and large farms.** Larger farms or agri-food businesses are generally better positioned to integrate FMIS, digital logbooks, or automated monitoring tools into their compliance routines. This is often due to the availability of dedicated personnel who can handle digital record-keeping, interface with regulatory portals, and ensure that reporting obligations are met on time. In contrast, smaller farms typically rely on the labour and oversight of one or two individuals, often family members, who must juggle production tasks with increasingly complex administrative demands. This structural imbalance is compounded by differences in financial capacity. Larger operations are more likely to afford commercial software licences, hardware upgrades, and advisory services, while smaller farms must weigh such investments against tighter margins and often cannot justify the cost of a system primarily used for compliance. As a result, large farms benefit from streamlined digital processes and early access to innovation, while smallholders face increasing expectations to digitalise without equivalent support or infrastructure. In some cases, interviewees

⁴⁴⁹ Based on interviews with representatives of farmers' associations conducted for this study.

⁴⁵⁰ Based on stakeholder survey responses.

⁴⁵¹ Ibid.

⁴⁵² Ibid.

⁴⁵³ Ibid.

reported that smaller farms had chosen to opt out of CAP subsidies altogether, citing the administrative burden and digital complexity as too high a barrier. Moreover, larger farms tend to be more homogeneous in terms of production systems, making it easier to adopt off-the-shelf compliance software designed for monoculture or single-product operations. In contrast, smaller farms often find that existing tools are poorly suited to their needs, leading to partial use, workarounds, or complete disengagement from digital solutions. This contributes to a growing digital gap in compliance practices across the sector, with smallholders increasingly at risk of regulatory non-compliance or exclusion from support schemes that presume a baseline of digital capability. This challenge was echoed in several survey responses, where stakeholders noted that advanced tools are difficult to use in general, but especially for smaller farms.⁴⁵⁴

3.3.4.3 Qualitative cost-benefit analysis

Overview of costs

Establishing and operating compliance and reporting systems in agriculture involves a wide spectrum of costs. These include labour, licensing, storage, training, interoperability, and, in the case of advanced and automated systems, investments in hardware, maintenance, and research. The survey responses show that the cost profile varies significantly by technique: paper and spreadsheets are dominated by labour and error risks, while digital and automated solutions shift the burden towards licensing, training, and integration.

The **financial cost of compliance** remains one of the most significant barriers to the wider uptake of advanced digital tools, particularly for small farms. Expenditure varies widely across Member States, with average annual costs per farm ranging from EUR 12.50 in Malta to over EUR 10,000 in Germany. The EU median is EUR 236 per farm, representing around 0.4% of total farm costs, though this rises to over 0.7% in countries such as Lithuania.⁴⁵⁵ Labour costs are a major component. In the absence of dedicated administrative staff, compliance tasks are often managed by family members or outsourced to consultants—particularly when requirements are complex or digital tools difficult to navigate. Findings from the European Commission's 2024 survey confirm this trend: 78% of farmers rely on external assistance to complete CAP applications, drawing on cooperatives, advisors, or private service providers.⁴⁵⁶ External support services are frequently used to manage the overall complexity. The annual cost of such support can exceed EUR 6,500 per farm, particularly in regions with a high reliance on consultants due to digital or regulatory challenges.⁴⁵⁷ These cumulative demands not only increase transaction costs for individual farmers but may also discourage participation in voluntary agri-environmental schemes or lead to inadvertent non-compliance due to misunderstanding or administrative error.⁴⁵⁸ These additional costs are especially burdensome for smaller farms, where margins are tight and capacity is limited.

Paper-based techniques are by far the most labour-intensive. Farmers reported that maintaining handwritten logs, parcel maps, and compliance forms requires substantial time, often outside normal working hours. Labour costs and risks of error were

⁴⁵⁴ Ibid.

⁴⁵⁵ European Commission, *Analysis of administrative burden arising from the CAP*, 2024. https://agriculture.ec.europa.eu/common-agricultural-policy/cap-overview/cmef/regulation-and-simplification/analysis-administrative-burden-arising-common-agricultural-policy_en

⁴⁵⁶ Ibid.

⁴⁵⁷ Ibid.

⁴⁵⁸ Ibid.

unanimously rated as significant, alongside the costs of physical storage and the potential consequences of mistakes or incomplete records. Other cost components—such as licensing or training—were negligible, underscoring that the main burden of paper systems lies in human time and inefficiency rather than financial outlay.⁴⁵⁹ For paper records and spreadsheets, the financial outlay remains low, which explains their continued prevalence among small farms.

Spreadsheets reduce some of the burden of physical storage but remain strongly dependent on manual data entry. Labour and error risks continue to dominate the cost profile, with stakeholders rating them as significant across the board. Integration costs were also highlighted, since spreadsheets are incompatible with CAP portals like GSAA or IACS, forcing manual re-entry. While hardware, licensing, and training costs were mentioned, they were assessed as manageable or minor in comparison. Microsoft Office ranges from around EUR 65 to EUR 260 per year, depending on whether the farm uses a personal, family, or business plan with multiple users.⁴⁶⁰ Free alternatives like LibreOffice or Google Sheets are also widely used. Overall, spreadsheets are cheap in monetary terms but costly in labour and error correction.⁴⁶¹

Advanced digital logbooks introduce a more balanced cost structure. Unlike paper or spreadsheets, they require significant upfront hardware or software set-up and generate licensing and interoperability costs. Entry-level digital logbooks like Agri-Trak and Fieldmargin provide accessible options. For example, Fieldmargin costs range between approximately GBP 99.96 (EUR 115) per year for the Essentials plan and GBP 449.88 (EUR 520) per year for the Pro plan, billed monthly. However, for compliance and reporting purposes, the Plus package at GBP 20.82 per month—or about GBP 249.84 annually (EUR 290)—is the minimum required.⁴⁶² Stakeholders also highlighted human capital costs, as training is needed to migrate from paper and to operate the tools effectively. At the same time, they noted that the risk of pure data loss is lower than for standard tools, but issues of privacy and data security emerged as new concerns. Integration with government portals and machinery was flagged as significant, pointing to the need for better compatibility.⁴⁶³

Farm Management Information Systems (FMIS) represent the most resource-demanding advanced tool. Farmers identified licensing fees and human capital costs as significant barriers, alongside the need for hardware, maintenance, and continuous support. Full-fledged Farm Management Systems like xFarm, AGRIVI, or Agroptima offer a broader suite of features – beyond just logging, they include planning, analytics, inventory, financial tracking, and integrations (e.g. IoT sensors, machinery data). These are more expensive but provide the highest compliance and efficiency benefits. For example, xFarm has tiered packages from free basic versions up to advanced plans EUR 195/year (Essential), EUR 570/year (Plus), EUR 1,640/year (Decision) and EUR 2,400/year (Professional). For compliance and reporting purposes, the Plus package (EUR 570/year) is the minimum recommended plan, as it includes Pro-level documentation, environment and climate tracking, and economic management—features necessary for CAP record-keeping, input traceability, pesticide logs, and financial reporting. Larger or more complex farms may opt for the Decision or Professional plans,

⁴⁵⁹ Based on stakeholder survey responses.

⁴⁶⁰ For further information: <https://www.microsoft.com/en-us/microsoft-365/buy/compare-all-microsoft-365-products>

⁴⁶¹ Based on stakeholder survey responses.

⁴⁶² For further information: <https://www.fieldmargin.com/pricing>

⁴⁶³ Based on stakeholder survey responses.

which include additional tools such as crop protection compliance, irrigation monitoring, and advanced reporting. Alternatively, xFarm also allows users to custom-build their setup by purchasing individual modules (ranging from EUR 30 to EUR 350 each per year), offering flexibility to complement or selectively activate features according to specific compliance needs.⁴⁶⁴ Research and development costs were also noted, reflecting the ongoing adaptation of FMIS to CAP reporting requirements. While labour costs were not considered as high as for paper or spreadsheets, they remain relevant given the manual checks required due to weak interoperability with GSAA, LPIS, or machinery. FMIS thus redistribute compliance costs towards licensing, integration, and training, making them more accessible to large farms than to smallholders.⁴⁶⁵

Fully automated tools such as blockchain-based traceability systems currently account for the smallest share of cost exposure in practice, given their limited adoption. However, the one response received highlighted that they are associated with significant licensing, hardware, and research costs, as well as high training needs. Labour costs were considered negligible, reflecting the automation of routine reporting tasks, but interoperability and scalability issues still introduce hidden costs of integration.⁴⁶⁶

Taken together, the findings suggest a clear shift in the cost structure as farms move from standard to advanced and automated solutions. Paper and spreadsheets are dominated by labour intensity and error risks, while digital and automated tools reduce these inefficiencies but replace them with high licensing, interoperability, and training costs. In practice, this means that small farms tend to stick to low-cost but labour-heavy methods, while larger farms are better positioned to invest in FMIS or automation—potentially widening the digital and financial gap across the sector.

Expected benefits

The introduction of advanced and automated compliance tools in agriculture promises a broad range of benefits, cutting across **efficiency, accuracy, and audit readiness**. The survey responses and interview inputs highlight that the strongest gains are associated with advanced digital logbooks and farm management systems, while fully automated solutions remain more prospective, with their potential yet to be realised at scale.

Advanced digital logbooks are most **valued for streamlining day-to-day compliance activities**. Stakeholders emphasised that digital logbooks reduce the costs of audits by making records easily accessible and traceable, while also automating data collection for reporting on pesticide use, fertiliser application, or water consumption. By eliminating the expenses associated with manual data entry and paper logs, they help reduce both labour intensity and error risks. Looking to the future, stakeholders highlighted the potential of digital logbooks to integrate directly with CAP portals, which could allow subsidy applications to be pre-filled or even automated. Features such as smart tagging of data to specific parcels and automatic detection of missing or inconsistent entries were also seen as promising developments to further reduce the risk of subsidy rejections.

Farm Management Information Systems (FMIS) offer a more comprehensive set of benefits. Stakeholders highlighted their **capacity to consolidate all compliance-related information into a single platform**, which directly reduces administrative overhead. The ability to centralise inputs from IoT sensors, weather stations, or soil tests strengthens the reliability of compliance data, while integrated alert functions can flag potential non-compliance risks (e.g., excessive fertiliser use, missed crop rotation requirements) before

⁴⁶⁴ For further information: <https://xfarm.ag/en/versions-and-prices>.

⁴⁶⁵ Based on stakeholder survey responses.

⁴⁶⁶ Ibid.

they occur. These predictive capabilities mark a **shift from reactive to preventive compliance management**. Looking ahead, stakeholders see particularly strong potential in the automated verification of eco-scheme eligibility, real-time cross-checking with LPIS databases, and automated validation of fertiliser and pesticide limits—all functions that would reduce errors and prevent costly subsidy rejections.

Fully automated tools, such as blockchain-based traceability platforms or AI-supported reporting assistants, remain at a much earlier stage of adoption but are associated with **high expectations**. Survey results show that stakeholders value their potential to eliminate manual compliance work altogether, by creating tamper-proof digital audit trails and automatically generating regulatory submissions. Blockchain-enabled smart contracts are viewed as especially promising for automating compliance validation. Future scenarios also envision AI-driven monitoring of regulatory updates to ensure that farms remain aligned with evolving EU obligations. However, these benefits are currently aspirational: implementation challenges, high costs, and regulatory uncertainty limit the practical deployment of fully automated tools.

Taken together, the survey and interviews suggest a clear pattern of progressive benefits. Paper and spreadsheets provide virtually no efficiency or accuracy gains, while advanced digital tools already deliver tangible improvements by reducing duplication, standardising records, and preventing errors in subsidy applications. FMIS take this a step further by consolidating data sources and proactively flagging compliance risks, while the full potential of blockchain and AI tools lies in future integration with EU-level systems and in their ability to shift compliance from a labour-intensive burden to an automated, preventive process. In practical terms, the most immediate benefits are realised in repetitive and data-heavy tasks, such as preparing eco-scheme reports, managing pesticide use records, or documenting CAP subsidy eligibility, while long-term benefits are expected in areas such as real-time monitoring, automated validation, and blockchain-based traceability. However, compliance is rarely the primary driver of adoption. Instead, most farms invest in digitalisation to improve productivity, optimise resource use, or simplify logistics, with compliance functionalities typically regarded as a useful by-product rather than the main incentive.⁴⁶⁷

Cost benefit assessment

The cost–benefit landscape suggests that **advanced digital logbooks currently represent the best balance between costs and benefits**. They provide significant efficiency and accuracy gains at a moderate and manageable cost profile, making them accessible to a broad range of farms, including smaller holdings.

Table 20. Qualitative assessment of the costs and benefits associated with advanced reporting techniques

Type of reporting technique	Costs				Relative Benefits			
	Labour cost	Financial cost	Technical cost	Human capital	Efficiency	Accuracy	Audit readiness & traceability	Real-time monitoring
Advanced digital logbooks	Reasonable	Reasonable	Reasonable	Reasonable	Significant	Significant	Reasonable	Reasonable
Farm Management Information Systems (FMIS)	Reasonable	Significant	Significant	Significant	Significant	Significant	Significant	Significant
Fully automated tools	Negligible	Significant	Significant	Significant	Significant	Significant	Significant	Significant

Source: The data used to create this table were collected through a survey of industry associations.

⁴⁶⁷ Based on interviews with representatives of farmers' associations conducted for this study.

By contrast, **FMIS deliver the most comprehensive compliance support**, integrating multiple data streams and offering strong benefits across all compliance dimensions. However, their significant financial and technical costs mean that they are only realistic for larger or more technologically advanced farms.

Fully automated tools remain largely aspirational. While they promise the most transformative benefits in terms of traceability, real-time monitoring, and audit readiness, their high financial, technical, and human capital costs—combined with the lack of proven large-scale deployment—mean they cannot yet be considered the best option in practice.

Therefore, digital logbooks emerge as the most effective and scalable solution under current conditions, striking a workable balance between affordability and tangible benefits. FMIS represent the next step for larger farms able to invest in more sophisticated solutions, while fully automated tools may become the preferred choice in the longer term, once costs fall and integration with EU platforms improves.

Overall, the findings suggest a trade-off between accessibility and sophistication. Simpler tools such as paper and spreadsheets are low-cost but inefficient, while advanced and automated systems promise greater accuracy and streamlining but require higher investment, digital skills, and trust. Without targeted support in areas such as interoperability, training, and financial assistance, there is a risk that the uneven adoption of advanced tools will deepen the digital divide between large and small farms, leaving the latter at risk of exclusion from increasingly digitalised compliance regimes.⁴⁶⁸

3.3.5 Policy recommendations

Survey responses and desk research pointed to a consistent concern: current digital compliance systems tend to favour larger farms, potentially deepening inequalities across the sector. Several key policy recommendations emerged to address these imbalances.

Public funding for digital compliance tools

Ensure that all farms, especially small and medium-sized holding, can meet CAP digital compliance requirements without financial or technical exclusion.

Operational actions:

- **National and EU authorities should jointly fund the development and maintenance of at least one free, user-friendly CAP compliance platform per Member State.** This platform must support all core CAP reporting functions (e.g., area declarations, livestock tracking, eco-scheme documentation) and be accessible via web and mobile devices.
- **Launch targeted grant or voucher schemes for small farms** to subsidize the purchase or subscription of certified digital compliance tools, with eligibility criteria based on farm size and digital readiness.
- **Mandate that all publicly funded tools are interoperable with leading commercial farm management systems (FMIS),** using open APIs and standardized data formats to ensure seamless data exchange.
- **Tie a portion of CAP technical assistance budgets to the ongoing development, user support, and regular updating of open-source or low-cost compliance platforms.** Require annual reporting on uptake and user satisfaction.

⁴⁶⁸ Based on stakeholder survey responses.

- **Monitor uptake and effectiveness** through annual surveys and CAP audit data, adjusting funding and support mechanisms as needed to ensure no farm is excluded due to digital barriers.

Interoperability and EU-wide data standards

Reduce administrative burden and data duplication by establishing a harmonized, machine-readable data ecosystem for agricultural compliance.

Operational actions:

- **Establish an EU-level working group** (including Member States, farmer organizations, and technology providers) to define, publish, and maintain harmonized data standards for all CAP compliance processes.
- **Require that all new digital compliance tools—public and private—adhere to these standards** as a precondition for public funding, certification, or CAP eligibility.
- **Publish data standards and interoperability requirements at the start of each CAP legislative cycle**, ensuring that tool developers and farmers have clarity before implementation deadlines.
- **Accelerate the creation of agricultural data spaces under the EU Data Strategy**, connecting public and private systems and enabling secure, efficient, and error-reducing data flows.
- **Mandate regular interoperability testing and certification** for all compliance tools, with results published in a public registry to promote transparency and continuous improvement.

Investments in rural digital infrastructure

Eliminate connectivity barriers that prevent rural farms from participating in digital compliance and automation.

Operational actions:

- **Prioritize rural broadband and mobile coverage in national recovery and cohesion funding plans**, with explicit targets for minimum connectivity standards in all agricultural regions (e.g., minimum 30 Mbps download speed, 4G/5G coverage).
- **Link eligibility for digital compliance incentives and CAP support to demonstrated improvements in rural connectivity**, ensuring that investments directly benefit the most underserved areas.
- **Fund local pilot projects** that demonstrate the impact of improved connectivity on compliance efficiency, with results disseminated across Member States to encourage best practice adoption.
- **Require annual reporting by Member States on rural connectivity progress**, with corrective action plans for regions that fall behind.
- **Coordinate with telecom providers and local authorities** to ensure that infrastructure investments are aligned with the needs of the agricultural sector, including support for remote sensing, mobile applications, and real-time data submission.

3.4 Automotive

3.4.1 Overview of the sector

This section offers an overview of the European automotive industry. It begins by outlining the sector's core activities, followed by an analysis of key economic indicators.

3.4.1.1 Definition

The automotive sector encompasses a **wide range of core products** – including motor vehicles and components – and features a **diverse landscape of economic operators**, ranging from Original Equipment Manufacturers (OEMs) to suppliers.

Motor vehicles – the industry's primary output – are classified as either light-duty or heavy-duty based on their weight and intended use, **a distinction that carries significant regulatory implications**. Light-duty vehicles are primarily intended for personal use or small-scale commercial operations. This category includes passenger cars and light commercial vehicles (LCVs), such as vans and small pickups. In contrast, heavy-duty vehicles are designed for the transport of large volumes of goods or passengers. This category encompasses trucks and lorries (used for freight transport) as well as buses and coaches (used for passenger transport). Due to their size, function, and environmental impact, heavy-duty vehicles are generally subject to distinct regulatory frameworks, particularly concerning emissions standards and safety requirements. The automotive industry also produces two- and three-wheel vehicles, with specific legislation setting separate type-approval and safety standards.

Reflecting the complexity of the vehicle manufacturing process, the automotive sector encompasses a **wide range of economic operators**. OEMs are responsible for the design, engineering, and final assembly of vehicles. Tier 1 suppliers deliver fully assembled components, systems, or modules directly to OEMs – these are typically complex and integrated rather than individual parts. Tier 2 suppliers provide parts or materials to Tier 1 suppliers, while Tier 3 suppliers supply basic components or raw materials to Tier 2 suppliers.

The automotive industry corresponds to division 29 of section C in the NACE Rev.2 classification. It is further divided into the following groups:

- Manufacture of motor vehicle (29.1)
- Manufacture of bodies (coachwork) for motor vehicles; manufacture of trailers and semi-trailers (29.2)
- Manufacture of parts and accessories for motor vehicles (29.3)

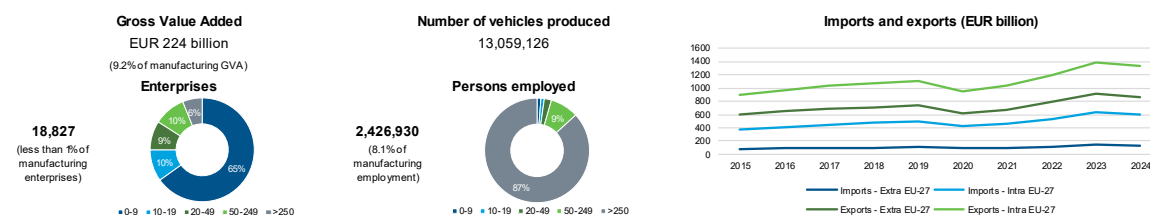
3.4.1.2 Size, value, employment

The automotive sector has long been a **cornerstone of the European economy**, characterized by strong upstream and downstream interconnections. It generates substantial demand for upstream industries – such as metalworking, chemicals, plastics, and textiles – while simultaneously driving activity in downstream sectors like vehicle maintenance and transportation services.

The significance of the automotive industry within the European economy is evident in key economic indicators (Figure 16). **In 2022, the sector contributed approximately EUR 224 billion in gross value added (GVA), making up 9.2% of total manufacturing GVA.** Nearly **19,000 enterprises** operated in the industry, with around 65% among them employing fewer than 10 people. Although the sector comprises less than 1% of all manufacturing firms, it accounts for 8% of manufacturing employment, **directly providing jobs to 2.4 million individuals**. The majority of these workers are employed by large

enterprises with over 250 employees. Beyond its direct impact, the automotive sector is estimated to support an additional 11 million jobs across the EU.⁴⁶⁹

Figure 16. The European Automotive Sector – Key Facts



Sources: The data on gross value added, the number of enterprises, and employment are sourced from Eurostat's *Structural Business Statistics dataset* (online code: *sbs_sc_ovw*). Trade data on imports and exports are drawn from the Eurostat dataset *International trade of EU and non-EU countries since 2002 by SITC* (online code: *ds-059331*). The number of vehicles produced is based on data from the European Automobile Manufacturers' Association, as published in *The Automobile Industry Pocket Guide 2023/2024* (<https://www.acea.auto/files/ACEA-Pocket-Guide-2023-2024.pdf>). All figures refer to the year 2022.

The automotive industry continues to demonstrate a **strong trade performance**, maintaining substantial surpluses. In 2024, the EU exported motor vehicles worth approximately EUR 265 billion to non-EU countries, while imports totalled EUR 132 billion – resulting in a trade surplus of EUR 133 billion. Lastly, the automotive industry stands out as a **leader in research and development**. Around 15% of its gross value added is reinvested into R&D activities, making it one of the most innovation-intensive industries in Europe.

3.4.2 Regulatory landscape, reporting and compliance activities

The current legal framework places substantial compliance and reporting obligations on automotive manufacturers, covering areas such as type-approval procedures, sustainability and emissions reporting, and the documentation of non-compliance. These procedures often involve extensive data collection, processing, and reporting. These legislations were selected following the criteria for the legal analysis laid down in the methodological section (see Section 3.1.1). The analysis below highlights several legislative acts that impose significant regulatory obligations on the automotive industry – many of which have broad applicability and involve tasks that could be automated to enhance efficiency and reduce costs.⁴⁷⁰ Annex A3.2 provides a detailed description of the legislative acts described below.

3.4.2.1 Legal and regulatory framework

Type-approval legislation

Regulation (EU) 2018/858 establishes the European Union's legal framework for the approval and market surveillance of motor vehicles, their trailers, and related systems and components. It applies specifically to vehicles in categories M (passenger transport), N

⁴⁶⁹ European Commission. (2024). *The future of European competitiveness: In-depth analysis and recommendations*. https://commission.europa.eu/document/download/ec1409c1-d4b4-4882-8bdd-3519f86bbb92_en?filename=The%20future%20of%20European%20competitiveness_%20In-depth%20analysis%20and%20recommendations_0.pdf

⁴⁷⁰ The Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (DG GROW) has compiled a comprehensive list of EU regulations and directives relevant to the automotive sector. The European Automobile Manufacturers' Association (ACEA) also regularly publishes its *Automotive Regulatory Guide*, with the 2023 edition available <https://www.acea.auto/publication/automotive-regulatory-guide-2023/>.

(goods transport), and their trailers in category O. It sets out technical requirements to ensure that all motor vehicles and trailers placed on the market meet safety and environmental standards.

In particular, **the regulation harmonises type approval procedures across the Union.** Type-approval refers to the process by which a vehicle, system, or component is certified as compliant with regulatory standards. For new approvals, Article 23 stipulates that manufacturers are required to submit an application to an approval authority in one Member state only. In addition to the application itself, manufacturers have to produce an information folder containing an information document – structured according to the template defined in the corresponding implementing acts – along with relevant data, drawings, and other supporting materials (Article 24).⁴⁷¹ Manufacturers must also specify the type of procedure selected for the vehicle's type approval as indicated in Article 22(1), and provide any additional information that approval authorities may request during an EU type-approval process. The information folder must be submitted in electronic format, although paper-based applications may be accepted at the discretion of the approval authorities.⁴⁷²

The regulation also requires automotive manufacturers to issue a certificate of conformity (CoC) in paper format for each vehicle placed on the market (Article 36(1)). The template for the CoC is defined in an implementing act.⁴⁷³ This certificate provides key information about the vehicle's main characteristics and technical performance and must be delivered to the buyer along with the vehicle. Starting from 5 July 2026, manufacturers will be allowed to issue the CoC in electronic format. From that date onward, they will also be required to provide the CoC to type-approval authorities as structured electronic data (Article 37(1)).

Additional reporting obligations are outlined as follows. Under Article 9(5), manufacturers must provide the Commission with data not included in the type-approval certificate but necessary for compliance verification. Recordkeeping duties extend to consumer complaints (Article 13(7)) and approval documents (Article 14(3)), which must be kept for up to 10 years. If any non-conformity is discovered after a product was placed on the market, Article 14(1) requires manufacturers to notify the relevant authority and outline corrective actions. Lastly, Article 14(4) obliges manufacturers to share conformity documentation in the local language with national authorities or the Commission.

CO₂ performance standards

Regulation (EU) 2019/631 establishes CO₂ performance standards for passenger cars and light commercial vehicles, with the associated reporting duties being further elaborated in the Implementing Regulation (EU) 2021/392. According to Article 9(1), manufacturers must collect real-world data on fuel and energy consumption that are linked to a vehicle's identification number (VIN). The data collected must be submitted to the Commission every year by 1 April, as stipulated by Article 9(3). In case of an error, they are required to notify the Commission by uploading a corrected data set, as indicated

⁴⁷¹ The templates for the information folders are laid down in Annexes I and I to the Commission Implementing Regulation (EU) 2020/683.

⁴⁷² The reporting requirements were materially increased by Regulation (EU) 2019/2144, which laid down more technical requirements for all new vehicles, like mandatory drowsiness detection systems. This Regulation increased the amount of information that must be reported on but did not formally change the reporting processes.

⁴⁷³ The template for the CoC is laid down in Annex VIII to the Commission Implementing Regulation (EU) 2020/683.

in the Articles 6(2) and 6(3). This effort aims to monitor real-world vehicle emissions and compare them to laboratory results, closing any discrepancies that might arise from test manipulations or variances in driving behaviour.

Regulation (EU) 2019/1242, as reinforced by Regulation (EU) 2024/1610, sets CO₂ performance standards for heavy-duty vehicles. Manufacturers must collect and report data required for the calculation of CO₂ emissions for each new vehicle and submit them to the Commission by 30 September each year (Article 13b). This aligns with the EU's broader goals of reducing emissions from the transport sector and increasing transparency and accountability among vehicle manufacturers.

Exhaust emissions and lifetime requirements

Regulation (EU) 2024/1257, known as the Euro 7/VII Regulation, introduces a unified framework for monitoring and reporting vehicle and engine emissions, as well as battery durability. It replaces the previous standards set by the Euro 6 regulation for light-duty vehicles and the Euro VI regulation for heavy-duty vehicles. While the emission limits for passenger cars under Euro 7 remain largely consistent with those under Euro 6, a notable change is the extension of particulate matter measurement to include particles with a diameter greater than 10 nanometres, instead of the previous threshold of 23 nanometres. For heavy-duty vehicles, the regulation imposes stricter emission limits and extends the required compliance period, meaning that vehicles must meet these standards over a longer distance or time.

In addition to exhaust emissions, Euro 7 introduces new requirements addressing non-exhaust sources such as tyre and brake wear and sets minimum performance standards for battery durability. All vehicles must be equipped with on-board monitoring systems capable of tracking emissions during real-world operation. These systems must be able to transmit data both through the on-board diagnostic interface and on-the-air, enabling authorities to assess roadworthiness and regulatory compliance.

The regulation will be implemented in phases. For new types of passenger cars and light vans, the requirements become binding on November 29, 2026, and will apply to all vehicles in these categories one year later. For buses, trucks, and trailers, the requirements take effect for new types on May 29, 2028, and for all vehicles one year thereafter. Separate implementation dates are set for tyre abrasion requirements, depending on the tyre class.

In terms of reporting obligations, Article 13(2) requires manufacturers to furnish data to the Commission upon request to verify compliance with the Regulation's Annexes. Article 14 mandates the submission of a formal declaration of compliance to the approval authority, certifying that the vehicle or component meets the standards set out in the Regulation.

REACH Regulation

Under REACH, the **general regulation on the registration, evaluation, authorization and restriction of chemicals**, any company that produces or imports a chemical substance in quantities above one tonne per year must register it, create or update a Chemical Safety Report (CSR) if volumes are above ten tonnes, keep Safety Data Sheets (SDS) up to date, notify the European Chemicals Agency (ECHA) of substances present in articles and or on the candidate-list in concentrations exceeding 0.1% weight per weight, pass hazard information along the supply chain, store records for ten years, and apply for authorization where required.⁴⁷⁴

⁴⁷⁴ ACEA has published guidelines to help automotive manufacturers and suppliers comply with the REACH Regulation, focusing on substance registration, communication, and risk management across the supply chain.

Corporate sustainability

The **Corporate Sustainability Reporting Directive** (CSRD) (Directive (EU) 2022/2464) sets out the obligation for large companies (balance sheet total exceeding EUR 20 million, net turnover exceeding EUR 40 million, or more than 250 employees) to report Environmental, Sustainability and Governance (ESG) data, extending and strengthening existing obligations under the Non-Financial Reporting Directive (NFRD) (Directive 2014/95/EU). The CSRD **adopts a “double materiality” approach, meaning that companies must document both how their activities impact people and the environment and how social and environmental issues affect the company.**

Reporting is conducted in accordance with the European Sustainability Reporting Standards (ESRS), which were developed by the European Financial Reporting Advisory Group (EFRAG) and adopted by the European Commission through delegated acts.⁴⁷⁵ The ESRS are grouped into several categories: two cross-cutting standards (covering general requirements and disclosures), and ten topical standards addressing environmental, social, and governance matters.

Articles 19a and 29a of the directive outline key reporting obligations for in-scope entities. Article 19a requires companies to disclose information on environmental, social, and governance (ESG) matters, following the double materiality approach. Article 29a complements this by mandating the use of standardized reporting frameworks to ensure consistency and comparability across disclosures.

In its initial version, the CSRD provided for a phased implementation, with large public-interest entities already subject to the NFRD expected to report in early 2025 on the financial year 2024. Large entities not subject to the NFRD and listed SMEs were required to start reporting later. The Commission has temporarily paused additional reporting requirements. A broader revision of the ESRS is currently underway, with a new version expected to be in place by the 2027 financial year.

In addition to the CSRD, **the Directive on corporate sustainability due diligence (CSDDD) (Directive (EU) 2024/1760) will entail extensive reporting requirements**, although its scope is more limited. It will only apply to companies with more than 1,000 employees and a net worldwide turnover of EUR 450 million. The reporting requirements are detailed in article 16 and include, among others, an annual statement on due diligence, risk management, complaints received, and efforts to combat climate change. The Commission is required to issue delegated acts to determine the content and criteria for reporting by April 2027. The Omnibus Package adopted in February 2025 simplified due diligence requirements.

Other legislation

In addition to obtaining approval for the whole vehicle, manufacturers must comply with various legal obligations related to individual parts and components. For instance, **Directive 2006/40/EC, which addresses emissions from air-conditioning systems in certain vehicle categories**, requires manufacturers to provide approval authorities with technical documentation detailing the installed systems and the refrigerant gases used (Article 4(2)). This requirement promotes transparency and enables authorities to assess whether vehicles meet environmental standards.

⁴⁷⁵ The ESRS were initially adopted on 31 July 2023 through Commission Delegated Regulation (EU) 2023/2772. An amending delegated regulation was adopted on 11 July 2025 (pending publication in the Official Journal), temporarily easing reporting requirements for companies scheduled to begin reporting for the 2024 financial year.

The **End-of-Life Vehicles (ELV) Directive (Directive 2000/53/EC)** aims to prevent and reduce waste from vehicles at the end of their life cycle by promoting reuse, recycling, and recovery. To achieve this, manufacturers must ensure that new vehicles are at least 85% reusable or recyclable by weight, and at least 95% reusable or recoverable by weight. Additionally, the directive prohibits the use of certain hazardous substances – such as lead, mercury, cadmium, and hexavalent chromium – except in specific cases where no viable alternatives exist, as outlined in Annex II.

By the same token, **Directive 2005/64/EC addresses the type-approval of motor vehicles regarding their reusability, recyclability, and recoverability**. Under Article 5(2), manufacturers must supply approval authorities with technical data necessary to perform the calculations stipulated in Annex I. This includes detailed information on the materials used in the construction of vehicles and their components to assess the potential for recycling or reuse. Moreover, Article 6(7) imposes a continuing obligation: manufacturers must inform the competent authority of any changes that could impact the relevance of the issued certificate of compliance, ensuring traceability and ongoing conformity even after approval is granted. In 2023, the Commission has proposed a new Regulation that would replace Directive 2000/53/EC and Directive 2005/64/EC and set stronger targets for reuse, recyclability, and recovery as well as expand the scope of application.

In conclusion, manufacturers within the EU must navigate a complex regulatory landscape that demands proactive data management, transparent communication with authorities, and rigorous recordkeeping. These obligations span the entire lifecycle of the product – from initial type-approval to post-market surveillance – and are essential for ensuring environmental sustainability, product safety, and regulatory compliance across the Union.

3.4.2.2 Reporting and compliance activities

Type-approval legislation

The existing legislation allows for **several type-approval procedures** (Article 22 of Regulation (EU) 2018/858). In **the step-by-step procedure**, each system, component, or separate technical unit are approved individually and incrementally, with the whole vehicle being approved once all individual approvals have been obtained. In **the single-step procedure**, the entire vehicle is approved in one operation. **Mixed type-approval** consists of a step-by-step approach in which some approvals are obtained separately, while others are included in the final approval of the whole vehicle. In all cases, manufacturers are required to produce and submit an information folder and follow the sequence of steps described below.⁴⁷⁶

Companies seeking type-approval must first undergo an initial assessment conducted by the national type-approval authorities. The outcome of this assessment is the issuance of a Conformity of Production (CoP) statement, confirming that the manufacturer has the capability to consistently produce products that comply with technical standards, performance criteria, and marking requirements. The CoP is subject to regular verifications through independent audits.

Once approved, the manufacturer makes available roughly a dozen or more prototype cars that will be used to test compliance with the existing legislation. In most Member States, the national type-approval authorities (TAA) do not have in-house testing facilities.

⁴⁷⁶ European Commission (2016). *The new EU type-approval framework for motor vehicles – Commission Proposal Explained*. Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs – Automotive and Mobility Industries Unit, <https://www.europarl.europa.eu/cmsdata/102200/PPT%20type%20approval.pdf>.

Compliance is, in these cases, verified by designated Technical Services who run the tests and compile a report detailing the results. This document is submitted to the type-approval authority for review.

If all requirements are met, the authority issues the type-approval certificate. Following the start of production, the authority may carry out audits and inspections to ensure continued compliance of both the production facilities and the vehicles produced. Every time a manufacturer sells a vehicle, it is required to provide the buyer with a Certificate of Conformity.

The overall type-approval procedure involves substantial data collection to compile the required information folder. This process is highly complex, demanding extensive documentation for both individual subsystems and the complete vehicle. The type-approval procedure also requires frequent exchanges of information. For example, manufacturers must submit specific data to designated technical services, typically via email. These services then prepare a report, which is either sent directly to the type-approval authority or, in some cases, returned to the manufacturer for onward submission. The formats used to share data vary depending on the content and may include PDF, Excel, or Word files.

CO₂ performance standards

Under article 9(1) of Implementing Regulation (EU) 2021/392, **if on-board fuel or energy consumption monitoring devices are installed, manufacturers are expected to collect real-world data together with the VIN.** The only exception to the rule concerns cases in which the owner of the vehicle has refused to make the data available. If the emissions data and the VIN cannot be collected directly by the manufacturer, they need to ensure that authorised dealers and repairs collect them on their behalf and transmit the data when the vehicle is brought in for services or repairs (Article 9(2)).

The data needs to be communicated to the Commission via a CSV text file that must meet the following conditions:

- Only the following parameters need to be included in the following order: VIN (17-character text), total fuel consumed (litres), total distance travelled (kilometres), distance travelled in charge depleting operation with engine off (kilometres), distance travelled in charge depleting operation with engine running (kilometres), distance travelled in driver-selectable charge increasing operation (kilometres), fuel consumed in charge depleting operation (litres), fuel consumed in driver-selectable charge increasing operation (litres), and total grid energy into the battery (kilo-Watt-hours).
- The fields are separated by a comma, “,”.
- The first row must contain the names of the variables, as shown in the fictitious example below (Figure 17).

Figure 17. Example of CSV file containing real-world emissions data

```
VIN,TotLifetimeFuel_l,TotLifetimeDist_km,ChargeDeplOprEngineOff_km,ChargeDeplOprEngineOn_km,ChargeIncrOpr_km,ChargeDeplFuel_l,ChargeIncrFuel_l,EnergyintoBattery_kWh
ABC12345678901234,2291,12496,,,,,
ABC12345678901235,2234,12343,325,2545,454453,345,345,345
```

Source: European Environment Agency.

The CSV file containing the information above needs to be submitted to the Business Data Repository (BDR) of the European Environment Agency using the Transport Reportnet 3 infrastructure. To this aim, manufacturers must designate a person of contact who is responsible for uploading the file. The EEA has prepared guidelines aimed at helping manufacturers use the Transport Reportnet 3 and submit the data following the procedure described in Figure 18.

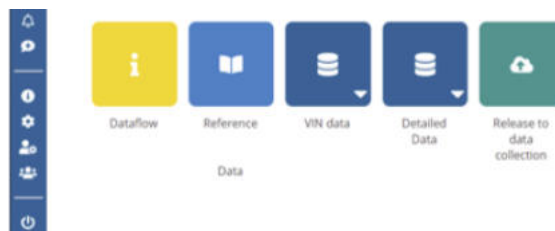
Figure 18. Submitting real-world emissions data – Step by step

Step 1: Log on to the digital infrastructure ((<https://transport.reportnet.europa.eu/>)).

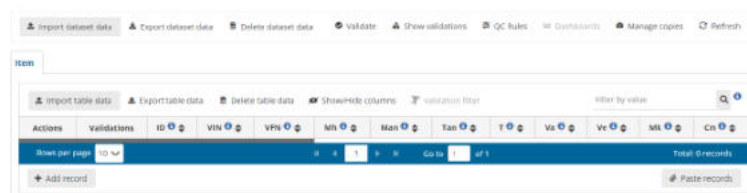
Step 2: Choose the relevant dataflow (*CO2 emissions – Cars – Manufacturers* or *CO2 emissions – Vans – Manufacturers*).



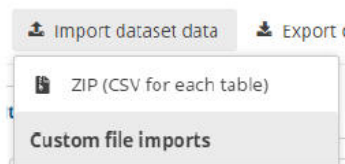
Step 3: Click on *Dataflow help* to obtain supporting documents.



Step 4: Upload the data by selecting *Detailed data* in the Dataflow.



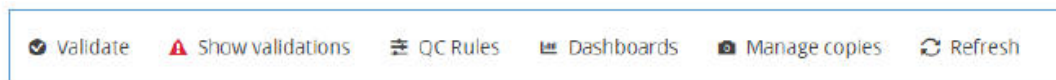
Select 'ZIP (CSV for each table)' in the *Import dataset data* button.



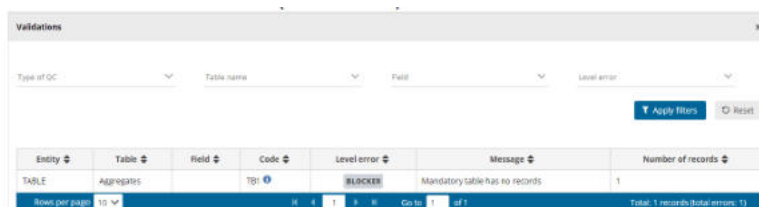
Indicate the path of the file to be uploaded.



Step 5: Check the quality of the dataset in order to validate it (several options are available).



Step 6: Review the results of the quality control (the show validations button lists the errors).



Step 7: Release the data by clicking on *Release to data collection* in the *Dataflow overview*.

Source: European Environment Agency. *How to use Transport Reportnet 3 for reporting of CO2 emissions from new passenger cars and light commercial vehicles (vans).*

Exhaust emissions and lifetime requirements

Under the Euro 7/VII Regulation, automotive manufacturers are required to **collect and monitor emissions performance data**. These systems must capture real-world data for a range of pollutants and performance metrics, including NO_x, PM, PN, CO, NH₃, as well as brake and tyre particle emissions and battery durability. OBM systems must also support over-the-air data transmission to enable remote compliance monitoring. The collected data may need to be shared with authorities to verify compliance with the Regulation. The regulation also establishes an obligation for manufacturers to deliver an environmental vehicle passport (EVP).

REACH Regulation

The European Chemicals Agency (ECHA) has developed several tools that facilitate compliance with REACH.⁴⁷⁷

The **registration process**, for example, involves several steps. First, manufacturers are required to **submit an inquiry** with ECHA. The goal of the inquiry is to assess the data that are available (and, therefore, if more testing is required) and if the substance is not already registered. The inquiry dossier needs to be prepared with IUCLID (International Uniform Chemical Information Database), a software designed to record, store, maintain, and exchange data on the characteristics of chemical substances. The inquiry is then submitted to ECHA using the REACH-IT online application.

The **registration follows a similar process**. The technical dossier is created by entering the required information in IUCLID. If necessary, the accompanying CSR can be generated using the application CHESAR, also developed by ECHA. Once all the validation steps have been completed in IUCLID, the dossier can be submitted to ECHA via REACH-IT.

Corporate sustainability

To meet CSRD/ESRS requirements, companies must **report on both qualitative and quantitative KPIs**, following structured patterns that include descriptions of policies, targets, action plans, transition plans, and potential financial impacts. Although EFRAG initially planned to develop sector-specific standards, this effort has been postponed in favour of prioritizing the implementation of the sector-agnostic ESRS. As a result, one of the key challenges for the automotive industry is that current reporting requirements do not fully reflect the sector's specific characteristics. Nevertheless, a recent study indicates that **nearly 50% of the KPIs expected under ESRS were already being disclosed by automotive firms as early as 2021**.⁴⁷⁸

To achieve compliance with the CSRD, **automotive manufacturers are required to collect and process data in the ten specific and two overarching categories**. These efforts will arguably span various divisions within a company and will be both time-consuming and labour-intensive. Furthermore, it is to be expected that data capabilities will need to be expanded – comprehensive ESG databases will need to be compiled, sourcing data from various IT systems.

Companies are also required to set targets that extend beyond environmental goals, which have so far been the dominant focus in the industry. The current legal framework

⁴⁷⁷ See the section on the Chemicals industry for detailed information on the available applications that facilitate compliance and reporting under REACH.

⁴⁷⁸ Hefter, K., Gebhardt, J., Weiss, C., Puetz, J., Leiendecker, J., & Haas, V. (2023). *How the Automotive Industry Can Prepare for CSRD/ESRS*. Boston Consulting Group. <https://media-publications.bcg.com/How-the-Automotive-Industry-Can-Prepare-for-CSRD-ESRS.pdf>.

expects automotive manufacturers to also set goals regarding pollution, biodiversity, as well as social and governance issues.

Furthermore, companies are required to publish a time-bound transition plan that reports on their strategy to achieve sustainability goals. In contrast to the siloed approach typically adopted to carry out compliance activities on technical and environmental issues, sustainability requirement will require collaboration between several departments across the entire organisation, from sustainability groups and human resources to compliance and finance.

Concerning the reporting format, sustainability KPIs are to be included in the annual management report in a dedicated section. The sustainability report is to be structured in four parts, in the following order: general, environmental, social, and governance information. Figure 19 illustrates a version of a sustainability statement, as published in Commission Delegated Regulation (EU) 2023/2772.

Figure 19. Sustainability statement (example)



Source: Commission Delegated Regulation (EU) 2023/2772.

3.4.2.3 Overview of existing challenges

One of the key challenges facing the automotive industry in regulatory compliance and reporting is **the sheer volume and complexity of legal obligations**. As one of the most heavily regulated sectors in the EU, automotive manufacturers must navigate a broad array of requirements. A significant portion of these obligations originates from UNECE (United Nations Economic Commission for Europe) regulations, developed under the World Forum for Harmonization of Vehicle Regulations (WP.29).⁴⁷⁹ These are frequently integrated into the EU legal framework through various regulations and directives. In addition to binding legal requirements, industry standards also contribute substantially to the compliance burden.⁴⁸⁰

In its review of EU legislative acts adopted between 2010 and 2022, **the European Automobile Manufacturers' Association (ACEA) identified over 50 regulations and 20 directives affecting the automotive sector** (see Figure 20). Legislative activity during this period was notably intense, with only a brief slowdown between 2015 and 2017. For example, in 2021 and 2022 alone, 20 legislative acts were adopted. This sustained pace

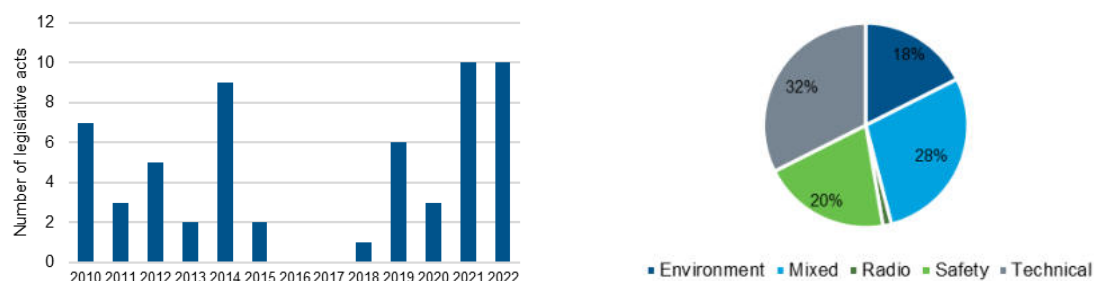
⁴⁷⁹ Further information on UNECE regulations can be found on the official website, <https://unece.org/wp29-introduction>. Information on the translation of UNECE regulations into EU law can be accessed at https://single-market-economy.ec.europa.eu/sectors/automotive-industry/legislation/un-regulations_en.

⁴⁸⁰ For examples on industry standards, see the following report on ISO standards for the automotive sector: International Organization for Standardization. (2016). ISO and road vehicles (2nd ed.), <https://www.iso.org/publication/PUB100292.html>.

of regulatory development requires automotive manufacturers to allocate substantial resources to monitoring legal changes and updating their compliance and reporting processes.

Moreover, the legislative acts span a wide range of areas. As Figure 20 illustrates, nearly one-third of the legislative acts adopted between 2010 and 2022 address technical aspects. A significant share focuses on safety (20%) and environmental issues (18%), with the remainder being cross-cutting and covering multiple domains.⁴⁸¹ In addition to these regulatory obligations, manufacturers must also report on financial performance, sustainability criteria, or cybersecurity. As a result, reporting responsibilities within companies tend to be siloed, with different teams managing reporting based on the subject matter. **Industry representatives have pointed out that no single unit can oversee the full spectrum of reporting activities.** Similarly, **there is no unified tool available to support all reporting needs.** Instead, companies rely on a variety of methods and systems depending on the type of reporting required. This makes regulatory reporting particularly complex and costly.

Figure 20. Overview of legislative activity (2010-2022)



Source: European Automobile Manufacturers' Association. (2023). *Automotive Regulatory Guide – 2023*. <https://www.acea.auto/publication/automotive-regulatory-guide-2023/>.

A related challenge lies in the **fragmented nature of the reporting process**, which involves multiple actors intervening at different stages, oftentimes using specific formats and templates. For instance, reports submitted by technical services during type-approval procedures can vary in both layout and content. As industry representatives have noted, harmonisation of this process is only likely if a single tool proves effective enough that it gains widespread adoption – or if legislation mandates it. The **lack of systematic harmonisation** across the member States also means that manufacturers may need to submit different datasets to up to 27 separate electronic platforms. Moreover, **reporting is not conducted on a regular schedule**. It is triggered by product updates, regulatory changes, or identified non-conformities, making it difficult to plan ahead and allocate resources efficiently.

Lastly, stakeholder interviews revealed that **the main challenge lies not in completing templates, but in collecting relevant data and verifying its accuracy**. Therefore, the use of basic tools like Word or Excel does not necessarily indicate that the underlying reporting process was simple or straightforward.

⁴⁸¹ Safety requirements relate to external, internal, active safety, operation, occupant protection. Environmental requirements include engine power, fuel consumption, substances, emissions, and noise emissions. Technical requirements concern vehicles, components, and separate technical units. Radio requirements relate to telecommunications and wireless devices. Mixed requirements include diverse and other requirements. Only legislative acts adopted before 2022 are included in the analysis.

3.4.3 Overview of reporting and compliance techniques

This section provides an overview of the techniques used for regulatory compliance and reporting purposes in the automotive industry. It starts with a description of the different techniques used, supported by examples, and concludes with an evaluation of their implementation in the industry.

3.4.3.1 Description of techniques and tools

Based on the required level of human oversight and automation, the techniques used for compliance and reporting in the automotive sector were categorized as standard, advanced, or fully automated. Each category is described below.

Standard techniques

Standard reporting techniques are characterized by manual data entry, substantial human oversight, and limited automation of repetitive tasks. These methods rely heavily on **spreadsheet-based tools, such as Excel or Google Sheets**. Depending on the type of information being submitted, **other formats like Word or PDF may be used** – for example, technical drawings of parts or components are more commonly shared as PDFs. Once collected, the data entries are reviewed and validated by compliance officers, who compare them against relevant regulatory thresholds.

Compliance reports are usually submitted via e-mail (and occasionally by post in physical format) or directly uploaded to official repositories. However, there is typically no integration between the manufacturer's IT infrastructure and public regulatory portals, entailing a fragmented and manual reporting process.

Illustrative example: International Material Data System (IMDS)

The International Material Data System (IMDS) is a global platform that collects, manages, and analyses data on materials used in vehicle components. Originally developed through a collaboration among several automobile manufacturers – including Audi, BMW, Daimler, DXC, Ford, Opel, Porsche, VW, and Volvo – the initiative has since expanded to include additional companies.

IMDS was initially set up to support compliance with existing environmental requirements and in particular the ELV Directive.⁴⁸² It does so by requiring supplies to submit detailed information about the parts they intend to sell, including a declaration of all substances contained via a Material Data Sheet (MDS). Reporting in the IMDS is generally a contractual obligation.

At its core, IMDS is a centralised database offering the following key functionalities:

- Declaration of all the materials used in parts and components
- Identification of hazardous substances
- Transparency across the value chain
- Support for recyclability and sustainability

IMDS has become an essential tool for achieving and demonstrating compliance with EU legislation:

- ELV directive: By tracking the composition of components, manufacturers can certify that restricted substances (i.e., lead, cadmium, mercury, and hexavalent chromium) do not exceed the legal thresholds. The detailed breakdown also allows OEMs to evaluate recyclability and recovery potential.

⁴⁸² <https://www.imds-professional.com/en/database-imds/>

- **REACH Regulation:** Under article 33(1), information on SVHC must be communicated throughout the supply chain. Whenever the REACH Candidate List is updated, IMDS automatically checks submitted data and alerts producers if SVHCs are replaced with wildcards or marked as confidential in the MDS they submitted. In such cases, producers must resubmit the MDS to ensure compliance. Although this process is largely manual, IMDS facilitates adherence to REACH requirements.
- **IMDS A2 – Advanced:** This premium tool includes a “REACH Report” feature that helps manufacturers assess compliance with REACH.
- **Waste Framework Directive (Directive 2008/98/EC):** Companies must report SVHCs in products by registering them in the SCIP database maintained by ECHA. IMDS supports direct submission to SCIP from within the platform.

While IMDS is accessible online, several commercial solutions offer enhanced reporting capabilities. For example:

- **Acquis Compliance** provides software that manages IMDS reporting, including validation and submission of MDSs.⁴⁸³
- **APA Engineering’s MDS Xpress** offers a user-friendly interface, identifies missing MDSs, and automates tasks such as tree structure design and supplier communication.⁴⁸⁴
- **Yordas Helix** enables real-time monitoring, predictive compliance tools, and cross-sector regulatory coverage to reduce supply chain risks.⁴⁸⁵

Advanced techniques

Advanced techniques involve the use of specialised software that partially automates the collection, processing, and reporting of compliance-related data. While these tools significantly reduce manual effort, human input and oversight remain essential, particularly in configuring workflows, validating data, and interpreting results.

Enterprise Resource Planning (ERP) software is designed to manage information across an entire organisation – from finance and manufacturing to human resources and other back-office operations. One of its key advantages is the ability to connect all business applications to a common database, ensuring data consistency and accuracy. Depending on the deployment type, several options are available:

- **Onsite ERP** – the software is installed on the company’s premises, who is in full control of the software. The costs of onsite ERP include a perpetual license (priced on a per-user basis) and yearly maintenance fees (typically, valued at 18% to 20% of the software list price).⁴⁸⁶ Onsite ERP remains common in industries that are subject to significant legal, regulatory and organisational constraints.
- **Cloud ERP** – the software is not installed on the company’s premises and instead accesses it via the internet. Cloud ERP typically involves a monthly or annual fee, with the software being referred to as software-as-a-service (SaaS). Licenses can be charged on a per user, per application, or per application set basis. The main

⁴⁸³ <https://www.acquiscompliance.com/compliances/imds-camds-elv-gadsl/>

⁴⁸⁴ <https://apaengineering.com/chemical-regulatory-compliance/software/imds/>

⁴⁸⁵ <https://www.yordasgroup.com/news/beyond-imds-tools-for-automotive-chemical-regulatory-compliance/>

⁴⁸⁶ SAP. *Cloud ERP deployment options*. <https://www.sap.com/mena/resources/cloud-erp-deployment-options>.

advantage of cloud ERP rests in its scalability, with companies being able to adjust their systems, applications, and resources to their needs.

- **Hybrid ERP** – the software combines on-premises and cloud-based functionalities. Hybrid ERP systems typically require substantial involvement of the in-house IT team, who must manage the local infrastructure and ensure inter-operability with the cloud-based services.

ERP systems play a key role in streamlining compliance and reporting processes.⁴⁸⁷

Firstly, they enhance data integrity by enforcing high standards for storing, managing, and analysing information. However, companies remain responsible for verifying the accuracy of the data entered into the system. Secondly, ERP systems offer robust data security features, including controlled access, encryption and recovery. Thirdly, ERP systems enable real-time monitoring of requirements and provide updates on legislative changes. They can also be configured to generate and highlight compliance-relevant metrics. Fourthly, ERP systems are adaptable to evolving regulatory frameworks. They also facilitate transparent data audit trails, by organising data in a structured and accessible manner. Transactions and changes to the data are recorded, improving traceability and accountability. Lastly, ERP systems can generate accurate and complete reports.

ERP solutions are increasingly incorporating artificial intelligence (AI) technologies. These innovations are expected to enhance operational efficiency – with manufacturers being able to interconnect diverse ecosystems, collect and process data from various sources, monitor trends in real-time, and generate automated compliance reports.

Illustrative example: SAP S/4 HANA

SAP S/4HANA Cloud, Public Edition is a modular ERP solution that can suit the needs of manufacturing companies of all sizes. Through its many functionalities, it allows users to achieve and demonstrate compliance with several EU regulatory requirements, including:

- REACH regulation. Products are systematically assessed to ensure they meet regulatory standards, while registrations and regulated substances are continuously monitored. All substance and regulatory data are stored in a centralized system, enabling efficient generation and management of compliance documentation. Additionally, safety data sheets can be quickly created and distributed to relevant recipients, streamlining communication and ensuring transparency.
- Sustainability reporting. Manufacturers can use pre-built ESG data models to estimate relevant metrics and include them in the sustainability reports. For example, the software includes a module that can calculate greenhouse gas emissions and automatically included them in the ESG report.

Other commercially available solutions include Oracle Fusion Cloud, QAD Adaptive, and Infor CloudSuite Industrial.

While an increasing number of commercial solutions emerge to streamline regulatory compliance and reporting activities, their effectiveness depends on the existence of common standards and secure data exchange frameworks. To streamline reporting and ensure interoperability across the automotive value chain, initiatives like Catena-X are emerging to establish trusted data spaces.

⁴⁸⁷ CEO Monthly. (2024). *How ERP adoption helps organisations stay compliant with regulations*. <https://www.ceo-review.com/how-erp-adoption-helps-organisations-stay-compliant-with-regulations/>.

Illustrative example: Catena-X

Catena-X is a collaborative data ecosystem specifically designed for the automotive industry to enable sovereign, standardized, and secure data exchange across the entire supply chain. It supports regulatory compliance at the EU level by developing interoperable frameworks, semantic data models, and open-source modules that facilitate traceability, reporting, and certification. Through initiatives like the Digital Product Passport⁴⁸⁸ and alignment with EU sustainability directives,⁴⁸⁹ Catena-X helps companies meet emerging regulatory requirements efficiently and transparently.

For example, Catena-X has developed an Environmental and Social Standards Kit to support compliance with the German Supply Chain Due Diligence Act. The approach is easily extendable to relevant EU legislation, such as the Corporate Sustainability Due Diligence Directive (CSDDD), the Ecodesign for Sustainable Products Regulation (ESPR), and the Corporate Sustainability Reporting Directive (CSRD).⁴⁹⁰

The solution enables companies to identify whether Environmental and Social Standards (ESS) incidents have occurred within their supply chain, thereby helping them demonstrate compliance with applicable regulations. Current practices often involve extensive manual effort, with experts conducting isolated investigations, limited data sharing, and inconsistent data quality. The ESS Kit addresses these inefficiencies by establishing standards for data sovereignty and security, and by defining a standardized data model with a clearly structured tracing process.

One of Catena-X's success stories is the adoption by Koller Kunststofftechnik GmbH, a manufacturer of injection-moulded parts, of a centralized solution for collecting and analysing production and energy data. By implementing the ONCITE DPS system from German Edge Cloud, the company enabled real-time data analysis through a unified dashboard. This allows for faster error detection and process optimization. Energy consumption is now tracked with greater precision, supporting targeted savings initiatives and more effective sustainable energy management.

Automated techniques

Automated techniques harness emerging technologies – such as robotic process automation, machine learning, artificial intelligence, or blockchain – to streamline regulatory reporting processes. They typically require little human oversight throughout the process, although manual quality checks may be necessary once the final report is produced

Robotic process automation (RPA) refers to technologies that rely on software robots to execute routine, rule-based tasks that are typically carried out by humans. RPA is often used to interact with digital systems and enter data into forms, extract information, or

⁴⁸⁸ Catena-X. (2025, August 19). *Catena-X and OPC Foundation join forces to enable seamless industrial data exchange for the Digital Product Passport*. <https://catena-x.net/news/catena-x-and-opc-foundation-join-forces-to-enable-seamless-industrial-data-exchange-for-the-digital-product-passport/>.

⁴⁸⁹ Catena-X. (2025, February 26). *Catena-X welcomes EU Omnibus proposal*. <https://catena-x.net/news/catena-x-welcomes-eu-omnibus-proposal/>.

⁴⁹⁰ Eclipse Tractus-X. (n.d.). *Environmental and Social Standards Kit – Adoption View*. <https://eclipse-tractusx.github.io/docs-kits/kits/environmental-and-social-standards-kit/adoption-view/>.

transfer data across documents.⁴⁹¹ Several commercial RPA solutions are available, both proprietary and open source.⁴⁹² A preliminary analysis of existing software solutions indicate that proprietary tools offer a greater range of information functionalities and are more likely to be integrated with AI tools. Their adoption is always associated with a licensing cost. Open-source tools, on the other hand, are shaped by the existing community of developers, although they are showing steady development and growth.

Illustrative example: RPA for automated report generation

A recent study examined the implementation of RPA for report generation within an automotive company.⁴⁹³ In the benchmark scenario, an engineer receives an automated email from the reporting system. Upon clicking the attachment, the system begins generating the report – a process that can take several minutes, during which the system remains unavailable for other tasks. With RPA, engineers simply connect to the reporting system, specify the required reports, and configure their email client to forward the automated messages to a robot. The robot then initiates the report generation by clicking the attachment and runs an Excel macro to compare the new report with previous versions. The final outputs are saved to the file system. RPA effectively decouples two different activities, eliminating idle time and significantly improving efficiency. While the maintenance efforts were estimated at 2 hours per week, the automation of the report generation is estimated to have saved 1.2 full-time employees (1.2 FTEs).

Artificial intelligence (AI) refers to a wide range of computer science techniques that can execute tasks that are typically the result of human intellect. **Machine learning** (ML) techniques, a subset of AI, employ algorithms to analyse large datasets, identify patterns, and uses them to make predictions, decisions, or classifications. A subset of ML techniques, **deep learning** (DL) algorithms use structures called neural networks to learn from large volumes of data. They are mostly applied to discover patterns in unstructured data, such as images, speech, or text. **Generative artificial intelligence** (Gen AI) is a type of artificial intelligence that can generate new content based on patterns identified in existing data. Examples of Gen AI include Chatbots or image generators. Such techniques are used in the detection of anomalies and predictive analytics, enhancing real-time compliance monitoring. These techniques can be used to automate activities such as data collection, validation, analysis, documentation, reporting, monitoring, and decision support.

Blockchain technology relies on distributed ledger technology that organizes data into blocks that are linked together in chronological order. Blockchain technologies have several features. Firstly, they are decentralized: instead of data being stored centrally, it is stored across a network of nodes. Secondly, once data is stored, it cannot be easily altered or deleted. Thirdly, all participants have access to the same data. Fourthly, transactions are validated using clearly defined rules. Lastly, cryptographic techniques are used to enhance the security of the data – only authorised parties are allowed to make

⁴⁹¹ Diligent Corporation (2021). *Meeting Climate Reporting Expectations With Robotic Process Automation*. <https://www.diligent.com/-/media/Project/Diligent/Master/Insights/White-Papers/PDF-Media-Files/US-Meeting-Climate-Reporting-Expectations-With-Robotic-Process-Automation.pdf>.

⁴⁹² Ribeiro, J., Lima, R., Eckhardt, T., & Paiva, S. (2021). *Robotic process automation and artificial intelligence in industry 4.0—a literature review*. *Procedia Computer Science*, 181, 51-58.

⁴⁹³ Wewerka, Judith, and Manfred Reichert. *Robotic process automation in the automotive industry—lessons learned from an exploratory case study*. International Conference on Research Challenges in Information Science. Cham: Springer International Publishing, 2021.

changes. In the automotive sector, blockchain technologies are used to trace information on different components used along the supply chain, ensuring compliance with various legal requirements.

Illustrative example: XCEED (eXtended Compliance End-to-End Distributed)⁴⁹⁴

Compliance standards apply to the vehicle as well as to individual parts and their components. The average supply chain in the automotive industry includes 30,000 parts per car and hundreds of suppliers.⁴⁹⁵ Demonstrating compliance is therefore a data-intensive process that requires tracking extensive information from many sources.

The XCEED project – a blockchain-based solution developed by Groupe Renault, Faurecia, Knauf Industries, Simoldes, and Coşkunöz in partnership with IBM – seeks to eliminate the need to rely on extensive databases and paper trails. Specifically, XCEED created data infrastructure that allows actors along the supply chain to share and track information in a secure and transparent manner. Access and visibility are governed by specific permissions to ensure confidentiality while users and transactions are verified and safely stored for transparency and accountability purposes.

XCEED enables rapid information exchange, with a capability of handling up to 500 transactions per second. Documents are shared in real-time through a streamlined, automated process, significantly reducing – if not eliminating – data inconsistencies. Regulators benefit from transparent, reliable, and continuously updated compliance data.

The platform supports compliance reporting across a wide range of legal requirements, including safety, environmental, cybersecurity, and vehicle recycling standards. It also enables certification for every vehicle component throughout its lifecycle – from design and production to aftersales. The outcome is an expedited certification process. Verification can be performed at any stage, and response times to regulatory inquiries are substantially reduced.⁴⁹⁶

The expected gains are significant: Renault anticipates that the technology will yield productivity improvements of 15% to 30% in the relevant services, along with a 10% reduction in costs related to managing non-quality and non-compliance, and a 50% decrease in the overall cost of non-compliance.⁴⁹⁷

XCEED is not, however, the only blockchain technology available in the automotive industry. In 2020, Toyota launched its Blockchain Lab with the goal of advancing

⁴⁹⁴ Renault Group. (2021, April 21). *XCEED, the new blockchain solution for the certification of vehicle compliance, is moving a step further in Europe*. <https://media.renaultgroup.com/xceed-the-new-blockchain-solution-for-the-certification-of-vehicle-compliance-is-moving-a-step-further-in-europe/>.

⁴⁹⁵ IBM. (2021). Renault: *Driving auto supply chains forward with blockchain*. <https://www.ibm.com/case-studies/renault>.

⁴⁹⁶ Knauf Automotive. (2021, May 7). *XCEED – Blockchain in the automotive industry*. <https://knaufautomotive.com/xceed-blockchain-in-the-automotive-industry/>.

⁴⁹⁷ Renault Group. (2021, July 6). *XCEED: A new blockchain solution for Renault plants in Europe*. <https://www.renaultgroup.com/en/magazine/technology/xceed-a-new-blockchain-solution-for-renault-plants-in-europe/>

technologies that improve traceability across the supply chain and throughout the production process.⁴⁹⁸ At the same time, BMW developed the PartChain project, integrating blockchain and cloud technologies to track the movement of parts across its intricate, multi-stage supply chain.⁴⁹⁹ Another key objective of the initiative is to establish full traceability of raw materials, ensuring transparency from origin to final assembly. To support industry-wide applications of blockchain and other emerging technologies, the Mobility Open Blockchain Initiative (MOBI) was established in 2018, comprising 120 automotive, mobility and technology companies. One of the stated objectives of MOBI is to create standards for trusted self-sovereign data and identities, verifiable credentials, and data spaces interoperability.⁵⁰⁰

3.4.3.2 Current use of reporting techniques

Compliance and reporting activities in the automotive industry encompass a highly complex and diverse set of processes, involving a wide range of economic operators – from suppliers and OEMs to technical services. **This diversity is reflected in the variety of reporting techniques employed.** A qualitative analysis of survey responses from industry associations indicates that standard and advanced tools are the most commonly used techniques in the sector, whereas automated tools are rarely, if ever, utilized (see Table 21).

Targeted interviews with industry representatives confirmed these findings and further highlighted the heterogeneity of tools and formats used in regulatory reporting. For instance, during type-approval procedures, manufacturers share information with technical services using formats such as PDF, Excel, or Word. The choice often depends on the nature of the information being communicated – drawings of parts, for example, are typically transmitted as PDFs. Reports issued by technical services also vary significantly in both layout and content. In the case of electronic certificates of conformity, on the other hand, the format is defined by existing legislation and consists of structured data in XML. However, the completion of these forms remains a manual task.

To carry out compliance and reporting activities under the type-approval legislation, automotive companies rely predominantly on advanced methods. They are mostly used for maintaining records as well as for collecting and processing data. Standard tools, whether paper-based or spreadsheet-based, are used less frequently, while automated tools are employed even more rarely. For preparing applications, advanced software remains the preferred choice, whereas automated tools are never used. When sharing data with authorities and reporting cases of non-compliance, both standard techniques (i.e., by post or via e-mail) and specialised software are used, although rarely. These patterns likely reflect the infrequent nature of such reporting. The submission of type-approval applications is more likely to occur via post or email. Automated tools are never used to share records, notify authorities, or submit applications.

⁴⁹⁸ Toyota Motor Corporation. (2020). *Toyota Blockchain Lab, accelerating blockchain technology initiatives and external collaboration*. <https://global.toyota/en/newsroom/corporate/31827481.html>.

⁴⁹⁹ BMW Group. (2020). *BMW Group expands use of blockchain technology in supply chain management*. <https://www.press.bmwgroup.com/global/article/detail/T0307164EN?language=en>.

⁵⁰⁰ Additional information on MOBI and its working groups is available at www.dlt.mobi.

Table 21. Use of reporting techniques for carrying out compliance and reporting activities

Reporting activity		Standard tools / Manual submission	Advanced tools / Submission to official websites	Automated tools
Type-approval requirements	Maintaining records	Supporting	Predominant	Marginal
	Collecting / processing data	Supporting	Predominant	Marginal
	Preparing applications	Marginal	Supporting	Not in use
	Sharing records with authorities	Marginal	Marginal	Not in use
	Notifying authorities	Marginal	Marginal	Not in use
	Submitting applications	Supporting	Not in use	Not in use
CO ₂ performance standards	Collecting real-world data	Inconclusive	Inconclusive	Inconclusive
	Processing data	Inconclusive	Inconclusive	Inconclusive
	Preparing reports	Inconclusive	Inconclusive	Inconclusive
REACH regulation	Registering substances	Supporting	Supporting	Not in use
	Preparing / updating SDSs and CSRs	Marginal	Supporting	Not in use
	Notifying ECHA	Supporting	Supporting	Not in use
	Communication along supply chain	Predominant	Predominant	Not in use
	Submitting authorisation requests	Never	Inconclusive	Inconclusive
Exhaust emissions	Notifying authorities	Supporting	Marginal	Marginal
	Sharing data with authorities	Predominant	Marginal	Marginal
	Submit declarations of conformity	Predominant	Supporting	Supporting
Sustainability reporting	Preparing sustainability reports	Predominant	Predominant	Not in use
	Reporting on ESG standards	Predominant	Predominant	Not in use
	Make data available during audits	Predominant	Predominant	Not in use

Source: The data used to compile this table were collected through a survey of industry associations.

Although manufacturers are required to report annually on CO₂ performance standards, stakeholder feedback was insufficient to determine the specific techniques used in the process.

To comply with reporting requirements under REACH, automotive manufacturers are, by and large, equally likely to use standard and advanced tools. The only exception concerns the preparation of SDSs and CSRs, for which standard tools are less likely to be employed. Automated tools do not appear to be used in carrying out reporting activities under the REACH regulation. The results for the preparation and submission of authorisation requests were inconclusive, potentially due to the low occurrence of such activities.

Reporting activities under the exhaust emissions legislation (current or anticipated) are more likely to involve standard tools and manual submission, especially when sharing data with authorities or submitting declarations of conformity. Advanced and automated tools are rarely used to notify authorities or share data with them – they are relatively more likely to be employed when sending declarations of conformity.

Standard and advanced tools are regularly employed to support reporting activities under current sustainability legislation. Among standard tools, spreadsheet-based solutions are more commonly used than paper-based methods. Automated techniques are not used at all.

Automotive manufacturers primarily use commercial software – standard or customized – **for compliance with type-approval legislation,** irrespective of the type of techniques used (Table 22). **In-house or hybrid solutions are less common,** though the

former appear as being in use for CO₂ performance reporting. For REACH reporting, a mix of commercial and hybrid software supports advanced techniques, while adoption for automated techniques is less common. All software types – commercial, in-house, and combinations – are used under exhaust emissions legislation. For sustainability reporting, standard commercial software is preferred when using advanced techniques, with limited use of other software types.

Table 22. Type of software used for compliance and reporting activities

		Standard commercial software	Tailored commercial software	Custom-made, in-house software	Mixed solutions
Type-approval requirements	Advanced techniques	Yes	Yes	Inconclusive	Inconclusive
	Automated techniques	Yes	Yes	No	No
CO ₂ performance standards	Advanced techniques	Inconclusive	Inconclusive	Yes	Inconclusive
	Automated techniques	Inconclusive	Inconclusive	Yes	Inconclusive
REACH regulation	Advanced techniques	Yes	Yes	No	Yes
	Automated techniques	No	Inconclusive	No	No
Exhaust emissions	Advanced techniques	Yes	Yes	Yes	Yes
	Automated techniques	Yes	Yes	Yes	Yes
Sustainability reporting	Advanced techniques	Yes	No	No	No
	Automated techniques	No	No	No	No

Source: The data used to compile this table were collected through a survey of industry associations.

3.4.4 Challenges, barriers, and qualitative cost-benefit analysis

The evaluation of the methods currently used in the automotive industry to ensure regulatory compliance and reporting is based on desk research and stakeholder input. It reviews key challenges in current practices and identifies barriers to the adopting of automated approaches. The analysis concludes with a qualitative cost-benefit analysis, offering a preliminary assessment of the relative gains of implementing advanced and automated reporting solutions.

3.4.4.1 Overview of challenges per reporting technique

Traditional paper-based reporting methods present several challenges. They are often time-consuming and require significant manual effort, particularly due to the need to enter data twice – once for the report and again into digital systems. Although less frequently mentioned, updating manual forms can be costly, especially when regulations change frequently. Paper formats also fall short in supporting compliance monitoring and data analysis. Additional drawbacks include a higher risk of errors and inconsistencies, potential for physical damage or loss, security vulnerabilities, and limited accessibility.

Spreadsheet-based methods also come with several limitations. A common issue is their tendency to contain errors and inconsistencies, often resulting from manual data entry or incorrect formulas. Furthermore, manually entering and updating data is particularly time consuming. Another important drawback is that spreadsheets cannot handle complex data structures effectively. Additional concerns include limited support for collaboration, weak data security and analytical capabilities, particularly when identifying compliance risks. While less frequently mentioned, risks such as data loss or corruption are still relevant.

Advanced – though not fully automated – techniques present a key challenge in integrating smoothly with existing systems, workflows, and reporting platforms. Moreover, they are not always tailored to the specific needs of the organisation, which can limit their effectiveness. Other common drawbacks include high software license fees and potential

reliance on vendors to ensure the system remains updated and secure. Concerns related to data privacy and ownership, and employee training are mentioned less frequently. Technical issues such as software bugs or system crashes are rarely reported.

The use of **automated techniques**, by contrast, are characterised as being complex to implement, requiring substantial infrastructure investments and ongoing maintenance. Ensuring interoperability with existing software and reporting systems is a key challenge, compounded by a lack of practical guidance and sector-specific implementation examples. Data security and regulatory uncertainty also emerge as relevant concerns. These issues likely reflect both the relatively recent adoption of such technologies and the limited familiarity stakeholders have with automated tools and processes.

3.4.4.2 Barriers to adoption

One of the key challenges in adopting automated reporting tools is the **lack of interoperability** between reporting platforms and the IT systems used by automotive companies. This issue is further exacerbated by varying reporting requirements across Member States, each of which may rely on distinct platforms. Even when a unified portal is used for reporting, the process often remains difficult to automate. As demonstrated by the submission of real-world CO₂ emissions data, reporting is still largely manual, with no capability for system-to-system data transfer.

Stakeholders also express **scepticism about the feasibility of a single automated tool** addressing the full spectrum of reporting requirements. Automation would likely need to be tailored to each domain, leading to duplication of effort and increased implementation costs. Customising tools to meet varied specifications raises deployment costs, reducing the overall cost-effectiveness of automated solutions.

Another significant barrier to the adoption of automated reporting solutions is the **perception that such tools do not fundamentally alleviate administrative burden**. While automation can streamline processes and simplify some administrative tasks, it does not affect the volume or complexity of reporting obligations imposed by current legislation.

Another important barrier to the adoption of automated tools for regulatory compliance and reporting relates **to trust in automated outputs**. While participants in the online survey expressed a general willingness to trust the outputs of automated techniques, they consistently underscored the need for continued human oversight. This oversight may range from minimal manual reviews to extensive validation processes. Interviews with industry representatives reinforced these concerns, noting that although automation can simplify certain reporting tasks, exclusive reliance on automated outputs could expose manufacturers to significant financial penalties and reputational risks in the event of errors. Consequently, **manual checks remain essential**, limiting the extent to which automation can deliver substantial labour cost savings.

Table 23. Barriers to adoption of reporting techniques for compliance and reporting activities

	Time-consuming	Errors	Human capital and digital skills	Scalability and tailoring	Financial costs
Standard techniques	High	High	Low	High	Low
Advanced techniques	Medium	Medium	Medium	Medium	Medium
Automated techniques	Low	Low	High	Low	High

Source: The data used to compile this table were collected through a survey of industry associations.

3.4.4.3 Qualitative cost-benefit analysis

Overview of costs

Since manually entering and updating data is a time-consuming process, labour costs are the most significant costs associated with **standard techniques** such as paper forms and spreadsheets. Stakeholders widely agree that these costs are substantial. They are followed by expenses related to employee training. Technical costs – such as maintenance, support, and security – are generally considered reasonable. In contrast, capital costs (e.g., hardware and IT infrastructure) and operating costs (e.g., licensing and subscription fees) are seen as negligible.

When it comes to **advanced techniques** that rely on specialised software, capital and operating costs are considered the main cost drivers. These are followed by technical expenses, particularly those related to maintenance and support. Labour costs are generally viewed as reasonable, as are expenses linked to training and organisational changes, which were consistently assessed as reasonable by stakeholders.

The most significant costs associated with **automated techniques** relate to capital and operating expenses, with all stakeholders unanimously agreeing that initial set-up costs are substantial. Technical costs – particularly those linked to maintenance and support – are also seen as important cost drivers. Training-related expenses are generally considered reasonable, though stakeholder views vary significantly regarding their actual scale. Labour costs, by contrast, are typically assessed as negligible. However, industry representatives added nuance to this assessment, noting that the additional effort required to verify automated outputs may offset the labour savings gained during data collection, processing, and report preparation.

Expected benefits

The benefits of adopting **advanced and automated techniques** for regulatory compliance and reporting are wide-ranging, and include improved accuracy (i.e., consistent and precise data entry and processing), greater efficiency (i.e., faster data collection, processing, and reporting), cost savings (i.e., reduction in manual labour), or scalability (i.e., better management of large volumes of data).

Stakeholders participating in the online survey highlighted the ability to deliver highly accurate results as the primary benefit of using advanced techniques for regulatory compliance and reporting. They also pointed to efficiency gains as a major advantage, while cost savings were seen as beneficial but less significant. Additional benefits include real-time monitoring and timely updates on regulatory changes. Scalability, data security, and the ability to generate transparent and accurate audit trails were considered less important.

Stakeholders participating in the online survey expressed less consensus when assessing the current or expected benefits of **automated techniques**. This divergence may reflect both limited familiarity with such tools and their relatively lower adoption in the automotive industry, making it harder for participants to confidently assess their advantages.

The primary advantage of automated compliance and reporting techniques lies in the significant efficiency gains they offer. Stakeholders indicated that the activities most suitable for automation typically involve labour-intensive and repetitive tasks, such as collecting and processing data, maintaining and updating records, and preparing applications and compliance reports. Submitting reports, identifying cases of non-compliance and risks, and the notifying relevant authorities have also been signalled out as suitable for automation, although at a smaller rate.

Closely following efficiency gains is the benefit of real-time monitoring, which stakeholders also regard as highly valuable. AI-powered chatbots or virtual assistants are seen as able to provide valuable assistance in tracking regulatory changes and deadlines. Answering

compliance-related questions and identifying of compliance risks are also cited as potential uses. Assisting with form completion is less cited.

Accuracy is another expected benefit, though it is rated less favourably than the previous two – consistent with earlier findings indicating that, while automated results are generally trusted, manual verification remains necessary. This reliance on manual checks may also explain why these techniques are perceived to have a less important impact on cost savings. On the other hand, scalability and data security are considered more substantial benefits. In contrast, the ability to track regulatory updates and generate transparent, accurate audit trails is viewed as relatively marginal.

Cost-benefit assessment

To evaluate the potential gains from implementing advanced and automated reporting technologies, Table 24 provides a comparative assessment against standard reporting techniques. The analysis focuses on the most relevant cost and benefit components.

Across the three reporting techniques assessed, different cost drivers vary in significance. Labour costs are the dominant expense for standard techniques, while they play a much smaller role in advanced and especially automated approaches. In contrast, capital and operating costs are considered as substantial for automated techniques but are viewed as negligible for standard methods and reasonable for advanced techniques. Technical costs, particularly those related to maintenance and support, are important for automated systems but decrease in relevance for standard and advanced techniques. Overall, automation offers potential efficiency gains and labour cost savings, though its adoption may require significant upfront investment and ongoing technical support.

Table 24. Qualitative assessment of the costs and benefits associated with different reporting techniques

Type of reporting technique	Costs				Relative Benefits			
	Capital & operating	Labour	Training	Technical	Accuracy	Efficiency	Cost savings	Real-time monitoring
Standard techniques	Negligible	Significant	Significant	Reasonable				
Advanced techniques	Reasonable	Reasonable	Reasonable	Reasonable	Significant	Significant	Reasonable	Reasonable
Automated techniques	Significant	Negligible	Reasonable	Significant	Reasonable	Significant	Negligible	Significant

Source: The data used to create this table were collected through a survey of industry associations.

In terms of expected benefits, both **advanced and automated techniques are assessed as highly valuable in terms of for their efficiency** in streamlining data collection, analysis, and reporting. The accuracy of the results also ranked high for advanced techniques but was assessed as relatively less important when using automated techniques. This may reflect the **need for additional manual checks, with stakeholders being reluctant to fully trust automated outputs**. Cost savings, while an important benefit when using advanced techniques, was assessed as less important for automated techniques. Real-time monitoring, however, was seen as a relatively more important benefit when using automated tools. Audit trails consistently received the fewest votes in both cases.

Taking stock of these results, **transitioning from standard to advanced reporting techniques is expected to result in higher capital and operating expenditures**, while reducing labour and training costs. Technical costs are not anticipated to change

significantly. The primary benefits of adopting advanced solutions lie in improved accuracy and greater operational efficiency.

In contrast, the adoption of **fully automated techniques would entail significantly higher capital and operational investments**, along with increased technical costs. Labour expenses, however, are expected to be comparatively lower. Overall, **the net cost savings are likely to be more modest when compared to advanced techniques**. While efficiency gains remain notable, the key advantage of automated solutions lies in their ability to enable real-time compliance monitoring. These tools also offer improvements in accuracy over standard methods, though this benefit is less pronounced when compared to advanced techniques.

Overall, **most stakeholders who participated in the online survey indicated that the expected benefits of automated techniques outweigh their associated costs**. The main driver for adoption is the anticipated reduction in administrative burden and time spent on compliance activities. Improved data accuracy and a lower risk of penalties due to non-compliance were also cited as important drivers for adoption. These findings suggest that efficiency and reliability are the key perceived benefits motivating the use of automated tools in regulatory processes.

3.4.5 Policy recommendations

The outlook for the further development of advanced and automated techniques for regulatory compliance and reporting is generally positive, as highlighted by participants in the online survey, with most considering that further investments should be a priority for the automotive sector.

Harmonise reporting requirements and digital data standards across EU

Reduce administrative burden and fragmentation in regulatory compliance by establishing unified, machine-readable data standards and harmonized reporting requirements for the automotive sector.

Operational actions:

- **Mandate the adoption of EU-wide digital data standards for all major automotive compliance processes** (e.g., type-approval, emissions, REACH, CSRD), ensuring that data is structured, machine-readable, and interoperable across Member States.
- **Require that all digital compliance tools and reporting platforms – public and private – adhere to these standards** as a precondition for regulatory acceptance and public funding.
- **Leverage Catena-X as a reference architecture and operational model** for secure, standardized, and interoperable data exchange across the automotive supply chain. Encourage the integration of Catena-X modules and semantic data models into national and EU-level reporting platforms.
- **Develop a Union-wide reporting portal for key compliance obligations**, enabling single submissions that can serve multiple authorities and reducing duplication of effort for manufacturers operating in multiple Member States.
- **Monitor implementation and effectiveness through annual reporting and stakeholder feedback**, with corrective actions for non-compliance or persistent fragmentation.

Support the development and uptake of advanced digital compliance tools

Facilitate the transition from manual and spreadsheet-based reporting to advanced, integrated digital compliance solutions, with a focus on interoperability and supply chain transparency.

Operational actions:

- **Provide targeted funding and incentives** (e.g., grants, tax credits, or public procurement preferences) for the development and adoption of advanced compliance software, including ERP systems and sector-specific solutions.
- **Prioritize tools that are interoperable with Catena-X and other recognized data ecosystems**, ensuring seamless data exchange and traceability across the supply chain.
- **Encourage the use of Catena-X open-source modules and semantic data models** in the design of new compliance tools, particularly for traceability, sustainability reporting, and digital product passports.
- **Require regular interoperability testing and certification** for all compliance tools, with results published in a public registry to promote transparency and continuous improvement.
- **Support pilot projects and demonstration cases** that showcase the benefits of integrated, Catena-X-compatible compliance solutions, with a focus on reducing administrative burden, improving data quality, and enabling real-time reporting.

3.5 Healthcare

3.5.1 Overview of the sector

3.5.1.1 Definition

The healthcare sector in the European Union is defined according to Section Q of the Statistical Classification of Economic Activities in the European Community (NACE Rev. 2). Section Q broadly encompasses “**Human health and social work activities**” and is subdivided into three key divisions: Division 86 (human health activities, including hospital, medical, and dental practice activities), Division 87 (residential care activities), and Division 88 (social work activities without accommodation).⁵⁰¹

For the purpose of regulatory and compliance analysis, the core healthcare sector is most accurately represented by Division 86 – Human health activities, which forms the statistical backbone of Eurostat’s health expenditure and workforce reporting.⁵⁰² Division 86 includes a range of healthcare service providers, both public and private, and is composed of the following three groups:

- **Hospital activities (NACE 86.1):** Including inpatient and outpatient hospital services, emergency and trauma units, surgical centres, and intensive care departments. These establishments typically offer comprehensive diagnostic, therapeutic, and surgical services with specialised medical staff and equipment.

⁵⁰¹ European Commission: Eurostat, *NACE Rev. 2 – Statistical classification of economic activities in the European Community*, Publications Office, 2008

⁵⁰² Eurostat, *Healthcare expenditure statistics – methodology*: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Healthcare_expenditure_statistics_-_methodology

- Medical and dental practice activities (NACE 86.2): Encompassing services provided by general practitioners, medical specialists, dental surgeons, and related outpatient healthcare professionals in private practices or ambulatory care settings.
- Other human health activities (NACE 86.9): Covering a wide spectrum of ancillary healthcare services not classified under hospitals or medical practices. This includes diagnostic laboratories, pathology services, imaging centres, physiotherapy and rehabilitation clinics, midwifery and home nursing services, and other paramedical professions.

In parallel, a large share of the EU's regulatory framework on medicines and medical products applies not only to healthcare providers but also to the pharmaceutical manufacturing sector, which is classified separately under NACE Section C – Manufacturing. Specifically, **Division 21 (Manufacture of basic pharmaceutical products and pharmaceutical preparations)** covers the production of active substances, medicines, and vaccines placed on the EU market.⁵⁰³

3.5.1.2 Size, value, employment

The healthcare sector is one of the **largest and most economically significant components of the EU economy**. In 2022, current healthcare expenditure in the EU totalled approximately EUR 1,648 billion, representing 10.4% of the EU's Gross Domestic Product (GDP).

Expressed in per capita terms, healthcare expenditure amounted to roughly EUR 3,685 per EU inhabitant in 2022. When adjusted for purchasing power differences, this translates into substantial investment per resident, though national figures vary considerably depending on income levels and health system structures.⁵⁰⁴

Healthcare spending as a share of GDP varies widely across EU Member States. In 2022, Germany recorded the highest healthcare expenditure relative to GDP at 12.6%, followed by France at 11.9% and Austria at 11.2%, underscoring the strong role of public investment in healthcare in these countries. At the other end of the spectrum, Luxembourg reported the lowest ratio at 5.6%, with Romania and Ireland also below the EU average, spending 5.8% and 6.1% of GDP on healthcare, respectively.⁵⁰⁵

Healthcare financing in the EU remains dominated by public mechanisms. In 2022, 81.3% of all current healthcare expenditure was funded through government schemes and compulsory contributory health insurance schemes or medical savings accounts. Government schemes alone financed 30.0%, while compulsory schemes accounted for 51.3% of spending. The remaining 18.7% came from other sources, including household out-of-pocket payments (14.3%) and voluntary health insurance schemes, which represented smaller but variable shares across Member States.⁵⁰⁶

In 2022, hospitals accounted for the largest share of current healthcare expenditure in the EU 22.0% was distributed across residential long-term care facilities, ancillary service

⁵⁰³ Eurostat, *Statistical Classification of Economic Activities in the European Community (NACE Rev. 2)*, Section C – Manufacturing, Division 21: Manufacture of basic pharmaceutical products and pharmaceutical preparations.

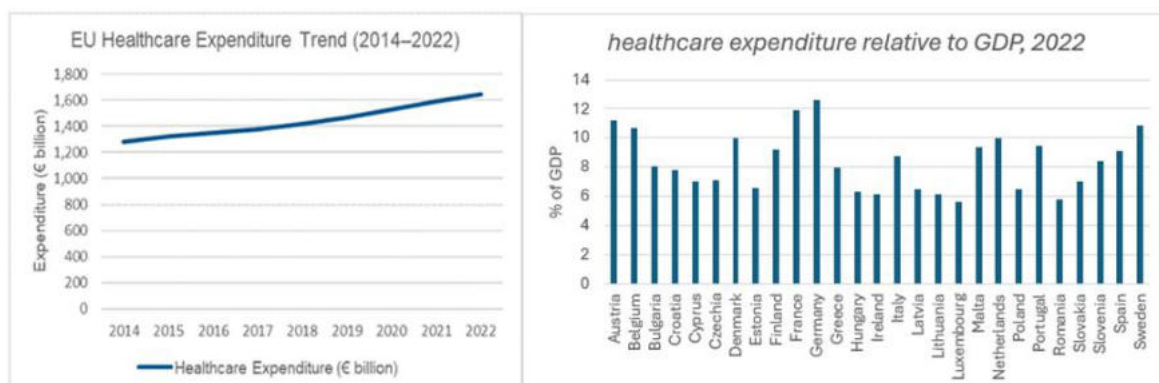
⁵⁰⁴ Eurostat, *Healthcare expenditure statistics – overview*: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Healthcare_expenditure_statistics_-_overview

⁵⁰⁵ Eurostat (2024), *Healthcare expenditure statistics by function, provider and financing scheme*. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Healthcare_expenditure_statistics_by_function%2C_provider_and_financing_scheme

⁵⁰⁶ Ibid.

providers, preventive care providers, and other entities involved in the delivery of healthcare.⁵⁰⁷

Figure 21. The European Healthcare Sector – Key Facts⁵⁰⁸



Employment in the healthcare sector also underscores its economic significance. In 2023, about 8.5% of the total EU workforce was employed in human health, reflecting a steady rise from 7.3% in 2008. Notably, more than three-quarters of health workers in the EU are female, emphasizing the sector's critical role in women's employment across the region.⁵⁰⁹

3.5.2 Regulatory landscape and reporting and compliance activity

3.5.2.1 Legal and regulatory framework

This overview gathers the main reporting and compliance duties that apply to healthcare actors operating in the European Union. It outlines, in broad terms, the obligations that hospitals, laboratories, and pharmaceutical or medical device companies must fulfil under the Medical Devices Regulation (MDR), the In Vitro Diagnostic Medical Devices Regulation (IVDR), the Clinical Trials Regulation (CTR), the EU Pharmacovigilance Framework, and the Paediatric Medicines Regulation. The full legal review is available in Annex 3.4.

Medical Devices Regulation (MDR, Regulation (EU) 2017/745)

The Medical Devices Regulation establishes a strengthened EU framework for the safety and performance of medical devices. It introduces more rigorous clinical evaluation, market surveillance, and transparency requirements, aiming to enhance patient safety and restore trust following past device-related scandals.

In Vitro Diagnostic Medical Devices Regulation (IVDR, Regulation (EU) 2017/746)

The In Vitro Diagnostic Medical Devices Regulation (IVDR, Regulation (EU) 2017/746) sets out comprehensive rules for placing in vitro diagnostic devices on the EU market. It expands the scope of regulated products, introduces stricter classification and conformity assessment procedures, and reinforces performance evaluation and oversight mechanisms—especially for higher-risk diagnostics

⁵⁰⁷ Eurostat, Healthcare expenditure statistics – overview. https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Healthcare_expenditure_statistics_-_overview.

⁵⁰⁸ The data on healthcare expenditure trends (2014–2022) and healthcare expenditure as a percentage of GDP (2022) are sourced from Eurostat's Healthcare Expenditure Statistics (online codes: *hlth_sha11* and *hlth_sha11_hf*).

⁵⁰⁹ European Commission (2023). *State of Health in the EU: Synthesis Report 2023*. https://health.ec.europa.eu/system/files/2023-12/state_2023_synthesis-report_en.pdf.

Clinical Trials Regulation (CTR, Regulation (EU) No 536/2014)

The Clinical Trials Regulation aims to harmonise and streamline the assessment and conduct of clinical trials across Member States. It establishes a centralised EU submission process via the Clinical Trials Information System (CTIS) and enhances transparency, data sharing, and public trust in clinical research.

Pharmacovigilance Framework (Consolidated Directive 2001/83/EC)

The EU Pharmacovigilance Framework governs the continuous monitoring of authorised medicinal products, with the objective of identifying, assessing, and mitigating risks related to their use. It strengthens obligations for data collection, signal detection, and risk minimisation across the medicinal product lifecycle.

Deforestation Regulation (Regulation (EU) 2023/1115)

The Regulation on Medicinal Products for Paediatric Use addresses the lack of appropriately authorised medicines for children in the EU. It requires the development and approval of Paediatric Investigation Plans (PIPs), enhances transparency on paediatric research, and integrates paediatric-specific requirements into EU authorisation and safety processes.

3.5.2.2 Reporting and compliance activities

Medical Devices Regulation (MDR) – Regulation (EU) 2017/745

Healthcare providers involved in the procurement and use of medical devices must ensure that only devices **bearing a valid CE mark and meeting MDR requirements are introduced into their facilities**. This verification typically includes checking the Declaration of Conformity provided by the manufacturer and, where applicable, the certificate issued by a Notified Body. In practice, such checks are carried out through internal procurement procedures and, where available, national databases or relevant EUDAMED modules. All related documentation — including certificates, batch numbers, and supporting technical files — must be properly archived to demonstrate compliance and ensure traceability.

In clinical settings, **traceability requirements are particularly important for Class III and implantable devices**. Healthcare professionals are required to record the Unique Device Identifier (UDI) in patients' health records and, where relevant, in national implant registries. This facilitates effective monitoring and ensures that any future recalls, safety notices, or adverse events can be traced back to specific devices and patients. Institutions must also ensure that their internal systems allow device identifiers to be securely logged, stored, and retrieved during inspections.

Post-market surveillance and vigilance reporting are further pillars of MDR compliance. While the primary responsibility for incident reporting rests with manufacturers, healthcare providers play a key role in detecting and notifying serious incidents (such as those leading to death, serious deterioration of health, or conditions that could have resulted in such outcomes). These reports are submitted through national vigilance systems, which then forward the information to the European database (EUDAMED). Providers must also cooperate in the implementation of Field Safety Corrective Actions (FSCAs), which may involve withdrawing devices, informing patients, or adapting clinical workflows. Internally, institutions are expected to log all incidents — even those not formally reportable — as part of their broader risk management and audit procedures.

In some cases, hospitals and clinics may engage in the reprocessing and reuse of single-use devices, provided this is permitted under national derogations or EU-level authorisation. In such instances, providers must demonstrate that reprocessing meets strict MDR quality and safety standards. This involves detailed documentation of

sterilisation, inspections, and validation cycles. Where required, reprocessed devices must be relabelled and assigned new UDIs to ensure ongoing traceability.⁵¹⁰

In Vitro Diagnostic Medical Devices Regulation (IVDR) – Regulation (EU) 2017/746

Diagnostic laboratories and medical testing facilities across the EU are subject to a significantly enhanced set of compliance obligations under the IVDR. For in-house devices—i.e., diagnostic tests developed and used within the same healthcare institution—laboratories must comply with Article 5(5) of the IVDR, which imposes rigorous conditions. These include **maintaining full documentation of the design, manufacturing, and performance evaluation processes, conducting a risk assessment for each in-house device, and justifying its use** by demonstrating that no equivalent CE-marked IVD is available on the market. Institutions must ensure that in-house devices meet relevant general safety and performance requirements, even though they are exempt from conformity assessment by a Notified Body. Laboratories are also required to monitor the performance of these tests over time and keep records available for inspection by national competent authorities.

For CE-marked IVDs acquired from external suppliers, laboratories must ensure traceability by recording and managing Unique Device Identifier (UDI) data for Class B, C, and D diagnostics. This includes integrating UDI information into laboratory information management systems (LIMS) and internal inventory control processes. Institutions are expected to maintain up-to-date technical documentation for all commercial IVDs in use and to verify that manufacturers' instructions for use and performance claims are consistent with clinical and operational requirements. Compliance activities further include active participation in post-market performance follow-up (PMPF), especially for Class C and D devices. This entails continuously reviewing device performance, managing user feedback, and documenting adverse events or performance deficiencies over time.

All diagnostic laboratories must also participate in the vigilance system established under the IVDR. Serious incidents, such as those involving false negative or false positive results that could result in significant patient harm, must be reported without delay to the competent authority, typically within 15 calendar days. Field safety corrective actions (FSCAs), such as test recalls or instructions for modification, must be communicated to users and regulatory bodies in a structured format, using the vigilance reporting requirements defined by the European Commission.⁵¹¹

Clinical Trials Regulation (CTR) – Regulation (EU) No 536/2014

Under the Clinical Trials Regulation, sponsors, investigators, and healthcare institutions involved in the conduct of clinical research must comply with a harmonised set of **obligations related to authorisation, transparency, safety reporting, and post-trial documentation**. All interventional clinical trials of medicinal products in the EU must be submitted through the Clinical Trials Information System (CTIS), which acts as the single-entry point for trial applications, substantial modifications, and result reporting. Healthcare professionals involved in these trials are responsible for ensuring that all clinical trial

⁵¹⁰ European Union (2017), Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC, OJ L 117, 5.5.2017. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017R0745>.

⁵¹¹ European Union (2017), Regulation (EU) 2017/746 of the European Parliament and of the Council of 5 April 2017 on in vitro diagnostic medical devices and repealing Directive 98/79/EC and Commission Decision 2010/227/EU, OJ L 117, 5.5.2017. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32017R0746>.

applications are complete, accurate, and submitted within the regulatory timelines. This includes uploading protocol documents, investigator brochures, subject information sheets, and responses to Requests for Information (RFIs) issued during the assessment procedure. Investigators and sponsors must also manage the process of obtaining informed consent in accordance with national requirements and Good Clinical Practice (GCP), documenting both the process and the signed consents in a manner that ensures traceability and auditability.

Once a trial is authorised, **compliance activities continue throughout the trial's lifecycle**. Trial sites must implement and document quality assurance measures, monitor protocol adherence, and maintain complete and up-to-date Trial Master Files (TMFs). Investigators must promptly record and report serious adverse events (SAEs) and suspected unexpected serious adverse reactions (SUSARs), which are transmitted electronically through designated safety reporting channels. These safety data contribute to pharmacovigilance assessments and must be submitted according to defined timelines and formats. In addition to safety reporting, healthcare professionals are expected to register the start of participant recruitment, report the end of recruitment, and submit annual safety reports (ASRs) and Development Safety Update Reports (DSURs) as applicable. Throughout the trial, updates on trial status, protocol amendments, and any temporary halts or early terminations must also be reported through CTIS.

After trial completion, healthcare institutions must ensure that clinical trial results, including summary results and layperson summaries, are submitted to CTIS and made publicly available. This transparency requirement applies regardless of the trial outcome and includes structured data uploads as well as document-based submissions.⁵¹²

Pharmacovigilance Framework (Consolidated Directive 2001/83/EC)

Under the EU Pharmacovigilance Framework, as established in Directive 2001/83/EC and further implemented through Commission Implementing Regulation (EU) No 520/2012, healthcare professionals and institutions play a central role in the continuous monitoring of the safety of authorised medicinal products. Healthcare providers are expected to **identify, document, and report suspected adverse drug reactions (ADRs), particularly serious or unexpected ones, through national reporting systems or directly into EudraVigilance**, the European Medicines Agency's central pharmacovigilance database. This includes ensuring that all relevant clinical and demographic details are recorded accurately and submitted within the required timeframes, which vary depending on the severity of the reaction. Pharmacists, doctors, and nurses working in hospitals or outpatient settings must also be familiar with the national pharmacovigilance procedures and reporting portals, as they are often the first point of contact for adverse reaction detection.

In addition to frontline reporting, hospitals and larger healthcare institutions are often involved in structured pharmacovigilance programmes and collaborations with marketing authorisation holders (MAHs). These may include providing access to anonymised patient records for follow-up, supporting targeted safety studies (such as post-authorisation safety studies or registries), and participating in signal detection or validation processes.

Healthcare professionals are also expected to contribute to the traceability of biological medicinal products by recording and retaining information such as the brand name and

⁵¹² European Union (2014), Regulation (EU) No 536/2014 of the European Parliament and of the Council of 16 April 2014 on clinical trials on medicinal products for human use, and repealing Directive 2001/20/EC, OJ L 158, 27.5.2014. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32014R0536>.

batch number, especially in the case of vaccines or other biologics. This traceability supports more effective post-market safety investigations.

Hospitals must also collaborate with MAHs in the event of safety signals, field safety corrective actions, or product withdrawals, and are expected to implement risk minimisation measures communicated via Direct Healthcare Professional Communications (DHPCs) and safety alerts. Finally, larger institutions may be required to provide input into Periodic Safety Update Reports (PSURs) or Risk Management Plans (RMPs), particularly when requested by authorities as part of additional monitoring requirements.⁵¹³

Paediatric Regulation (Regulation (EC) No 1901/2006)

Healthcare professionals, clinical investigators, and pharmaceutical sponsors involved in the development, prescription, or monitoring of medicinal products for children must ensure compliance with the Paediatric Regulation, which aims to improve the quality and availability of medicines for paediatric populations across the EU. For clinical research activities, investigators must contribute to paediatric data generation by **adhering to approved Paediatric Investigation Plans (PIPs)**, which outline how medicines should be studied in children and are required before applying for marketing authorisation for new products or indications. Sponsors must submit paediatric trial protocols, amendments, and results via the **Clinical Trials Information System (CTIS)** and ensure that trials involving minors follow the ethical and legal protections stipulated under both the Clinical Trials Regulation and national laws.

Once authorised, marketing authorisation holders must label and inform healthcare professionals clearly on the paediatric use of medicines, including specific dosages, formulations, and safety profiles validated through PIPs. Hospitals and paediatricians are expected to monitor and report adverse events occurring in children via national pharmacovigilance systems or directly into EudraVigilance, with attention to paediatric-specific reactions or unexpected outcomes. In addition, healthcare professionals involved in research or procurement may consult the European Medicines Agency's Paediatric Medicines database, which consolidates information on authorised paediatric medicines, waivers, and rewards.

Pharmaceutical companies are required to notify healthcare professionals about any modifications to paediatric indications and to comply with obligations regarding labelling and product information to reflect authorised paediatric uses. Clinicians must ensure that such updated information is applied in prescribing and treatment practices. For off-label paediatric use, institutions must document rationale, treatment outcomes, and, where applicable, submit observational data contributing to broader paediatric pharmacovigilance efforts. These recurring obligations reflect the shared responsibility of both regulators and healthcare professionals in safeguarding paediatric patient

⁵¹³ European Union (2001), Directive 2001/83/EC of the European Parliament and of the Council of 6 November 2001 on the Community code relating to medicinal products for human use, OJ L 311, 28.11.2001. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32001L0083>; and European Union (2012), Commission Implementing Regulation (EU) No 520/2012 of 19 June 2012 on the performance of pharmacovigilance activities provided for in Regulation (EC) No 726/2004 of the European Parliament and of the Council and Directive 2001/83/EC of the European Parliament and of the Council, OJ L 159, 20.6.2012. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012R0520>.

populations while encouraging the responsible development and monitoring of child-appropriate medicines.⁵¹⁴

3.5.2.3 Overview of existing challenges

Healthcare actors have highlighted several issues that complicate regulatory compliance under current conditions.

One of the most prominent challenges is the **administrative and operational workload** resulting from recent regulatory reforms. The introduction of the Medical Devices Regulation (MDR) and In Vitro Diagnostic Medical Devices Regulation (IVDR) has significantly expanded compliance tasks, particularly around clinical evaluation, technical documentation, and post-market surveillance. Healthcare providers are now responsible for traceability procedures and incident reporting, usually via national vigilance systems that feed into EU-level platforms such as EUDAMED. Manufacturers, meanwhile, face continuous performance monitoring, recertification, and expanded reporting duties. These activities require specialised staff, training, IT infrastructure, and internal coordination—resources often lacking in smaller hospitals, SMEs, or public-sector institutions.⁵¹⁵ According to Authorised Representatives, device registration in EUDAMED remains highly time-consuming. They are legally obliged to verify whether manufacturers have correctly registered devices, but the system does not notify them when changes occur. This forces repeated manual checks across hundreds of devices, significantly increasing the administrative workload.⁵¹⁶

Technical and institutional fragmentation further complicates compliance. Regulatory data must be captured and submitted across multiple platforms—such as EudraVigilance, the Clinical Trials Information System (CTIS), and EUDAMED—yet many health organizations still depend on legacy or non-interoperable IT infrastructures. This significantly hampers efficient data exchange, traceability, and audit readiness. The European Commission's study on the interoperability of Electronic Health Records (EHRs) in the EU highlights that fragmented national implementations, and outdated systems continue to limit harmonised data sharing and coordination between Member States.⁵¹⁷ According to interviewees, manufacturers also face overlapping documentation requests from different actors—including Notified Bodies, Authorised Representatives, competent authorities, and EU-level databases—which often require duplicating the same information in different formats. This duplication was identified as one of the most pressing challenges.⁵¹⁸

Legal and procedural uncertainty is another critical issue. Despite the harmonising intent of the MDR, IVDR, and CTR, Member States have applied certain provisions

⁵¹⁴ European Union (2006), Regulation (EC) No 1901/2006 of the European Parliament and of the Council of 12 December 2006 on medicinal products for paediatric use and amending Regulation (EEC) No 1768/92, Directive 2001/20/EC, Directive 2001/83/EC and Regulation (EC) No 726/2004, OJ L 378, 27.12.2006. <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32006R1901>

⁵¹⁵ Nüssler, A. (2023). The new European Medical Device Regulation: Friend or foe for hospitals and patients? BMC Health Services Research, 23, 274. <https://doi.org/10.1186/s12913-023-09921-7>; and, Muehlematter, U. J., Schöb, D., & Bossuyt, P. M. (2023). The challenges for manufacturers of the increased clinical evaluation in the European Medical Device Regulations: A quantitative study. International Journal of Medical Informatics, 178, 105222. <https://pubmed.ncbi.nlm.nih.gov/37198369/>

⁵¹⁶ Based on an interview conducted with an authorised representatives' association.

⁵¹⁷ European Commission (2021). Interoperability of Electronic Health Records in the EU. Available at: <https://digital-strategy.ec.europa.eu/en/library/interoperability-electronic-health-records-eu>.

⁵¹⁸ Based on an interview conducted with an authorised representatives' association.

differently—especially during transitional periods—leading to inconsistencies in clinical evidence expectations, conformity assessments, and in-house test exemptions. This variability affects both manufacturers and healthcare providers, who must adapt compliance procedures depending on national interpretations. For example, interviewees noted that non-EU manufacturers, particularly SMEs, often struggle even to understand their compliance obligations, let alone manage them digitally.⁵¹⁹

3.5.3 Overview of reporting and compliance techniques

3.5.3.1 Description of techniques and tools

Standard tools for data compliance in the healthcare sector

In the context of EU healthcare regulations, standard compliance tools refer to basic, non-specialised instruments used by organisations to manage and submit required regulatory information. This may take the form of spreadsheets or other documents to track compliance activities, compile technical files or risk assessments, or manual data entry to prepare and submit information to national authorities or EU platforms such as EudraVigilance, CTIS, or national systems.

Advanced tools for data compliance in the healthcare sector

Advanced tools encompass a variety of healthcare-specific compliance software suites. These tools are typically cloud-based and offer centralized environments for managing internal compliance processes, such risk assessment, incident reporting, or audit documentation.

Key feature of many compliance software tools is their emphasis on **interoperability with other health IT systems and regulatory databases**. Standards-based integration is crucial for meeting data submission and reporting obligations under various EU regulations. The adoption of HL7's Fast Healthcare Interoperability Resources (FHIR)⁵²⁰ standard has been identified as a strategic enabler of structured and interoperable data exchange across healthcare institutions in Europe.⁵²¹

A number of commercial software platforms provide targeted support for healthcare compliance with EU regulations. For example, **Rimsys**,⁵²² a cloud-based regulatory information management (RIM) software, helps medical device manufacturers manage regulatory data including Unique Device Identifiers (UDI), product registrations, and EUDAMED submissions. It allows automated data preparation and supports machine-to-machine transfer. Similarly, the software **RAMS**⁵²³ provides a digital compliance platform that supports MDR and IVDR implementation by enabling product classification, regulatory planning, deadline tracking, and access to a global regulatory knowledge base.

⁵¹⁹ Ibid.

⁵²⁰ HL7's Fast Healthcare Interoperability Resources (FHIR) is an international standard for exchanging healthcare information electronically, designed to improve interoperability between different health IT systems.

⁵²¹ European Parliament (2022). *The European Health Data Space. Study requested by the Committee on Industry, Research and Energy*. [https://www.europarl.europa.eu/RegData/etudes/STUD/2022/740054/IPOL_STU\(2022\)740054_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/STUD/2022/740054/IPOL_STU(2022)740054_EN.pdf).

⁵²² For further information: <https://www.rimsys.io/>

⁵²³ For further information: <https://www.emergobyul.com/software/regulatory-affairs-management-suite>

Another emerging solution is Formwork by **OpenRegulatory**⁵²⁴, an AI-powered electronic quality management system designed for lean medical device teams. It offers real-time checks, audit-proof e-signatures, and MDR-ready documentation templates within a user-friendly interface. For clinical trials, **ShareCRF**⁵²⁵ is a flexible electronic data capture system that allows sponsors to digitally manage multi-site trial data. It helps secure data collection, participant confidentiality, and remote monitoring.

Figure 22. Rimsys interface showing UDI compliance functionalities⁵²⁶

The screenshot displays the Rimsys interface for the 'Inanna™ SofStent Coil Assist Stent'. The top navigation bar includes 'All Products', 'Tags', 'EU Basic UDI', a search bar, and an 'Option+S' button. The main header shows the product name and a 'Part Number: 4894'. Below this, a horizontal menu lists various sections: Details, Fields, UDI, Configurations, Registrations, Projects, Essential Principles, Standards, Certificates, and Selling Status. The 'UDI' section is currently active, showing a sidebar with regulatory regions: Universal UDI, FDA UDI, EU MDR/IVDR, Saudi Arabia SFDA, China NMPA, South Korea MFDS, and Singapore HSA. The main content area is titled 'UDI Details' and contains several form fields: 'Issuing Entity' (dropdown menu), 'Primary DI Number' (text input), 'Has a secondary DI been issued by a different issuing entity?' (Yes/No radio buttons), 'Issuing Entity of Secondary DI' (dropdown menu), 'Secondary DI Number' (text input), 'Has a previous DI been issued?' (Yes/No radio buttons), 'Issuing Entity of Previous DI' (dropdown menu), 'Previous DI Number' (text input), 'Relationship Type (with previous DI)' (dropdown menu), 'Issuing Entity of Unit of Use DI' (dropdown menu), 'Unit of Use DI Number' (text input), and 'Is the device subject to direct marking or permanent marking requirements?' (Yes/No radio buttons).

Automated tools for data compliance in the healthcare sector

Automated solutions make use of artificial intelligence, robotic process automation, and interoperability standards like HL7 FHIR to streamline how data is recorded, submitted, and monitored.

AI-powered tools are particularly promising. They can help spot risks, fill documentation gaps, and generate alerts when compliance issues arise. For example, AI can monitor adverse event data to support post-market surveillance or help hospitals ensure their reports to EUDAMED (for medical devices) or CTIS (for clinical trials) are complete and submitted on time. One notable example is **Nuance DAX**,⁵²⁷ an AI-powered digital scribe already used in some US and European hospitals. It records doctor–patient conversations and automatically creates clinical notes, helping reduce documentation time and improve compliance-ready records. Such systems rely on Natural Language Processing (NLP) to structure clinical information in ways that align with regulatory requirements, and early

⁵²⁴ For further information: <https://openregulatory.com/formwork/>

⁵²⁵ For further information: <https://www.sharecrf.com/>

⁵²⁶ Rimsys interface showing UDI compliance functionalities. The dashboard highlights centralised UDI data management, with automated updates, country-specific formatting, and direct submission features to EUDAMED and other regulatory databases.

⁵²⁷ For further information: <https://www.nuance.com/healthcare/dragon-ai-clinical-solutions/dax-copilot/infographic/move-beyond-scribes-to-automatically-document-care.html?srsId=AfmBOOpKy0UxrdNMFzYaP7JVkPCC8nH3Mgr1IBkz0gyfcPSfPoR3-PPa>

evaluations suggest they can significantly reduce documentation workload for healthcare professionals.⁵²⁸ Electronic Health Record (EHR) systems are also integrating AI features to assist with compliance. Platforms like **Oracle Health**⁵²⁹ increasingly offer embedded dashboards that flag missing data, alert staff about overdue safety updates, and help generate regulatory reports in structured formats.

For highly repetitive and rule-based tasks, robotic process automation (RPA) is emerging as a solution to manage uploads to EU regulatory platforms. In principle, robotic scripts can automatically prepare XML files for vigilance reporting in EUDAMED or transfer structured data to systems such as CTIS or EudraVigilance. While large-scale use in hospitals remains limited, pharmaceutical companies have piloted these technologies; for instance, Pfizer has applied automation to accelerate the production of clinical study reports and to process case safety reports during the COVID-19 vaccine rollout.⁵³⁰

Finally, blockchain technologies are being explored for their potential to provide secure and tamper-proof audit trails. Pilot projects have proposed blockchain-enabled EHR auditing systems that log who accessed patient data, when, and for what purpose, thereby enhancing GDPR compliance and supporting regulatory audits.⁵³¹

While still experimental in healthcare, such systems could eventually provide pharmaceutical companies and hospitals with transparent and immutable compliance records across multiple actors.

3.5.3.2 Current use of reporting techniques

Healthcare compliance and reporting rely on a range of techniques, from paper records and spreadsheets to advanced digital tools and emerging automated systems. Standard methods remain dominant, while EHRs, compliance software, and interoperability standards increasingly support reporting. Automated solutions are still marginal (Table 25).

The comparative overview of current compliance and reporting techniques across the five healthcare regulations highlights the **persistence of spreadsheets alongside the gradual emergence of more specialised compliance software**. Paper-based tools were consistently rated as marginal, with stakeholders indicating they are rarely, if ever, used for core compliance tasks. Where they do appear, it is mainly for ad hoc exchanges of documentation with competent authorities, rather than as a systematic approach to regulatory reporting. This confirms that the digital transition away from paper in healthcare compliance is already well advanced.

⁵²⁸ Mehta, R., et al. (2024). "Ambient artificial intelligence in healthcare: scoping review." *Oxford Open Digital Health*, 3(1), <https://pmc.ncbi.nlm.nih.gov/articles/PMC10990544/pdf/ocae022.pdf>.

⁵²⁹ For further information: <https://www.oracle.com/health/>

⁵³⁰ <https://www.blueprism.com/resources/case-studies/how-pfizer-has-scaled-rpa-to-deliver-breakthroughs-that-change-patients-lives/>

⁵³¹ Ullah, F., He, J., Zhu, N., et al. (2024). "Blockchain-enabled EHR access auditing: Enhancing healthcare data security." *Computers in Biology and Medicine*, 170, 107774. <https://www.sciencedirect.com/science/article/pii/S2405844024104380>

Table 25. Use of reporting techniques for carrying out compliance and reporting activities

Technique	Standard - Paper	Standard - Excel	Advanced compliance software	Fully automated tools
Overall use	Marginal	Likely predominant	Likely predominant	Marginal
Medical Devices Regulation	Marginal	Likely Predominant	Supporting	Marginal
In Vitro Diagnostic Medical Devices Regulation	Marginal	Supporting	Supporting	Marginal
Clinical Trials Regulation	Marginal	Marginal	Predominant	Marginal
Pharmacovigilance Framework	Inconclusive	Inconclusive	Inconclusive	Inconclusive
Paediatric Medicines Regulation	Inconclusive	Inconclusive	Inconclusive	Inconclusive

Source: The data used to compile this table were collected through a survey of industry associations.

Spreadsheets remain used and can be considered a likely predominant tool, particularly under the Medical Devices and In Vitro Diagnostic Regulations. Stakeholders confirmed their use for storing technical documentation, preparing post-market surveillance (PMS) records, and submitting incident reports. However, the authorised representatives' association underlined that **reliance on Excel is increasingly problematic**, and its role is therefore gradually declining. These findings suggest that Excel persists not because of its suitability but because of regulatory requirements that have not yet adapted to more advanced reporting options.

Specialised compliance software has also reached the level of a predominant technique. Stakeholders reported using commercial or in-house systems to manage technical documentation, coordinate PMS activities, and prepare incident reports. In-house tools, in particular, play a notable role. Interviewees explained that these tools are commonly used to track certificate expiry dates, manage technical files, monitor device registrations, and coordinate audits. The authorised representatives' association noted that their organisation has completely phased out Excel for managing device tracking and certification processes and now relies on a dedicated in-house compliance system that tracks status changes, keeps historical records, and significantly reduces human oversight.

Fully automated tools are still used only to a limited extent across all frameworks, regardless of the type of compliance and reporting activity. Survey evidence and desk research show that automation is currently applied only in isolated cases, typically for structured internal processes rather than for direct regulatory submissions. Still, stakeholders clearly see potential for automation in healthcare compliance. The tasks most frequently identified as suitable for automation include uploading device or incident data to EUDAMED (three stakeholders), compiling pharmacovigilance reports (three stakeholders), and monitoring supply disruptions under MDR and IVDR (three stakeholders). These are repetitive, data-driven processes that currently absorb significant administrative time if performed manually. Some stakeholders also pointed to the preparation and submission of clinical trial reports in CTIS, as well as device registration, PSUR preparation, and trend reporting, as promising candidates for automation.

Taken together, the findings show a healthcare sector that has largely moved away from paper, still relies heavily on spreadsheets, and is progressively investing in specialised compliance software. Both Excel and dedicated compliance systems can now be considered the predominant techniques for regulatory compliance, albeit for different reasons: spreadsheets persist due to entrenched regulatory practices, while compliance software reflects a broader digitalisation trend. Automation is not yet widely deployed but is recognised as highly promising for structured, repetitive reporting activities. The overall

conclusion is that the sector stands at an intermediate stage: most organisations operate between spreadsheet-based workarounds and partially integrated compliance platforms.

3.5.4 Challenges, barriers, and qualitative cost-benefit analysis

3.5.4.1 Overview of challenges per reporting technique

Standard tools

Paper-based techniques are widely seen as inefficient and burdensome in healthcare compliance. Stakeholders underlined that they remain highly time-consuming, requiring repetitive manual work and duplication of effort, such as re-entering information into digital systems. They also pointed to difficulties with storing, retrieving, and archiving records, which can undermine long-term traceability and auditability.

Spreadsheets (e.g. Excel, Google Sheets) were also associated with important limitations. Stakeholders highlighted the risk of errors and inconsistencies in manual entry, the time required for updates, and the lack of interoperability with official reporting systems such as CTIS, EUDAMED or EudraVigilance. Issues related to data loss, corruption, and cybersecurity were also raised, along with concerns about readability.⁵³²

Advanced tools

Stakeholders emphasised the burden of high licence and subscription costs and noted persistent difficulties in integrating such tools with internal systems such as clinical, HR, inventory, or reporting platforms. Although commercial platforms are marketed as being interoperable with EU regulatory systems like EUDAMED, CTIS, or EudraVigilance, **stakeholders underlined that in practice such integration remains complex**. Other limitations include uncertainty about whether outputs are always accepted by regulators, limited staff uptake, insufficient customisation for MDR, IVDR, or CTR-specific workflows, and ongoing concerns about data security risks. One stakeholder also raised concerns about vendor lock-in, the absence of on-premise deployment options, and the rapid pace of product evolution, which demands highly flexible IT departments to keep up.⁵³³

These barriers suggest that, although advanced software can reduce manual work and support compliance, its value depends heavily on alignment with organisational workflows and regulatory acceptance.

Automated tools

Automation-based tools are promising but difficult to adopt in practice. Stakeholders emphasised the **high cost of setup and infrastructure as a primary challenge**, alongside the technical complexity of integrating such tools into existing IT systems. Other challenges included limited in-house expertise to manage or validate automated outputs, cybersecurity risks, and ongoing uncertainty about the regulatory acceptability of machine-generated evidence. Additional concerns were raised about vendor lock-in and the lack of on-premise deployment options, with several stakeholders stressing that rapidly evolving products require flexible IT departments to adapt effectively.⁵³⁴

Concerns were also expressed about trust, transparency, and explainability—particularly for AI-based systems. Healthcare compliance is a legally sensitive domain, and regulatory authorities may be reluctant to accept AI-generated outputs without clear documentation,

⁵³² Based on stakeholder survey responses.

⁵³³ Ibid.

⁵³⁴ Ibid.

traceability, and human oversight.⁵³⁵ Stakeholders generally reported that, while automation is seen as promising, it is not yet trusted enough to replace manual review.⁵³⁶

3.5.4.2 Barriers to adoption

The analysis of survey responses highlights five main categories of barriers that affect the use of compliance and reporting tools in the healthcare sector. Resource intensity refers to the workload and time required to maintain, update, and operate the tools. Losses and risks capture the potential for errors, data loss, corruption, or security breaches that undermine compliance reliability. Human capital and digital skills reflect the level of staff expertise, training, and acceptance needed to operate the tools effectively. Scalability and tailoring describe the ability of a tool to integrate with other systems, adapt to specific regulatory workflows, and keep pace with evolving requirements. Finally, financial costs encompass expenses linked to licences, infrastructure, maintenance, and development.

Table 26. Barriers to adoption of reporting techniques for compliance and reporting activities

Barrier	Standard tools - Paper	Standard tools - Spreadsheets	Advanced tools – specialised software	Fully automated tools
Resource intensive	High	High	Medium	Medium
Losses and risks	High	High	Medium	Medium
Human capital and digital skills	Low	Low	Medium	High
Scalability and tailoring	High	High	Medium	High
Financial costs	Low	Low	High	High

Source: The data used to compile this table were collected through a survey of industry associations.

The comparison across tools illustrates how barriers shift as organisations move from manual to digital and automated solutions. Paper-based and spreadsheet methods remain the most burdensome in terms of labour and error risks, making them increasingly unsustainable for meeting the demanding reporting obligations under MDR, IVDR, CTR, pharmacovigilance and paediatric medicines regulations. While they involve limited financial outlays, their lack of scalability and reliability creates systemic risks for organisations that continue to rely on them.

By contrast, specialised compliance software offers a more structured way of managing documentation and reporting, reducing error risks and improving traceability. However, these tools demand ongoing investment in licences, training, and system integration, which can be a barrier for smaller organisations. Their effectiveness depends heavily on successful interoperability with regulatory portals and internal hospital or industry IT systems.

Fully automated tools have the potential to deliver significant efficiency gains by reducing manual workload and enhancing real-time monitoring. Yet they are associated with the highest financial costs, steep human capital requirements, and persistent uncertainties about regulatory acceptance. This underlines the importance of trust, transparency, and oversight when automation is introduced into legally sensitive compliance areas.

⁵³⁵ EPRS. (2023). Regulating AI: Implications for the Health Sector. [https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739263/EPRS_BRI\(2023\)739263_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/BRIE/2023/739263/EPRS_BRI(2023)739263_EN.pdf)

⁵³⁶ Based on stakeholder survey responses.

Overall, the findings suggest a trade-off between accessibility and sophistication: while simpler tools are affordable but inefficient, advanced and automated systems are more powerful but also more costly and complex to implement.

3.5.4.3 Qualitative cost-benefit analysis

Overview of costs

Compliance costs remain a major concern for healthcare organisations. This is particularly evident in the medical devices sector. For example, according to a 2024 MedTech Europe survey, the average **medical device manufacturer is facing a 50% increase in compliance costs** under the MDR over a five-year cycle. These rising costs stem not only from expanded regulatory requirements—such as post-market performance studies and recertification—but also from the resources needed to establish and operate compliance reporting systems. Labour, infrastructure, licensing, maintenance, and integration with national and EU-level platforms all form part of the cost profile, which varies significantly depending on whether organisations rely on standard, advanced, or automated techniques.

Paper-based compliance remains highly labour-intensive. Maintaining documentation, preparing reports, and ensuring regulatory correspondence all depend on manual input, which absorbs significant staff time. Stakeholders emphasised that labour costs dominate the cost profile of paper-based systems.

Spreadsheets reduce some of the storage burden associated with paper but remain **strongly dependent on manual data entry** and updates. Labour costs are therefore still substantial. Participants also underlined the potential for costly errors, losses, or data corruption—making risk management a notable expenditure item. Integration with regulatory submission systems represents another challenge: spreadsheets are often incompatible with portals such as EUDAMED or CTIS, forcing organisations to invest additional time or resources in data transfer.

The shift to specialised compliance software alters the cost profile considerably. Here, the main expenses move away from labour and towards **licensing and subscription fees**, which most stakeholders rated as significant. Hardware and maintenance also become important, reflecting the need for IT infrastructure and ongoing support. While these tools can reduce errors and streamline reporting, they often require investments in training and human capital, since staff must become familiar with new workflows. Integration and interoperability also emerge as a cost factor, especially where healthcare organisations need to connect compliance software with clinical, HR, or inventory systems.

Automation introduces the highest upfront and ongoing costs. Hardware, licensing, and integration dominate the expenditure profile, as automated solutions such as robotic process automation or AI-based monitoring require advanced infrastructure and tailored interoperability. Research and development costs were also flagged as significant, underscoring that automation is still relatively novel and often requires adaptation to specific regulatory workflows. While labour costs decrease relative to paper or spreadsheet methods, they are not eliminated: staff are still needed to supervise, validate, and interpret automated outputs. Training and human capital costs therefore remain relevant, particularly given the expertise required to manage and trust automation in a legally sensitive domain like healthcare.

Taken together, the survey suggests a clear shift in the nature of compliance costs across reporting methods. Paper and spreadsheets are dominated by labour costs and the inefficiencies of manual data entry, with errors and rework creating additional expenditure. Specialised compliance software reduces these inefficiencies but introduces licensing and infrastructure as major cost drivers. Fully automated solutions promise long-term efficiency gains but require substantial upfront investment in infrastructure,

interoperability, and research, making them most accessible to larger or better-resourced organisations.

Expected benefits

The introduction of advanced and automated compliance tools in the healthcare sector promises significant benefits, ranging from greater efficiency and cost savings to improved accuracy and security in regulatory reporting. **Consultation findings indicate that advanced software solutions are most closely associated with gains in accuracy and speed**, while fully automated approaches are linked to efficiency and scalability.

Specialised compliance software is valued for reducing the time and effort required to prepare and submit reports. By replacing paper-based processes and spreadsheets with templates tailored to MDR, IVDR, and CTR obligations, these tools standardise documentation, cut down on repetitive work, and improve accuracy—an important factor for inspections and audits. **They also contribute to long-term cost savings by reducing errors**, easing preparation for Notified Body reviews, and supporting training and safety updates through scheduling and tracking functions.

Looking ahead, **stakeholders see potential in predictive dashboards and smart automation to detect risks earlier**, track training requirements dynamically, and streamline post-market surveillance. Automated tools are particularly valued for reducing administrative burden, enhancing audit readiness through digital trails, and ensuring faster compliance with reporting deadlines. While cost savings and reduced reliance on consultants are expected, automation is viewed less as a cost-cutting measure than as a way to combine efficiency with greater regulatory assurance.

Fully automated systems are primarily associated with scalability and the elimination of manual uploads to EU portals such as CTIS or EUDAMED. Automated data validation and digital audit trails are seen as especially valuable, though their benefits depend on successful integration with existing hospital IT and EHR systems. Future applications—such as blockchain-based traceability or RPA bots for repetitive submissions—are considered promising but not yet widely implemented.

Overall, advanced tools are regarded as the most practical and immediately beneficial option, professionalising compliance and improving reliability. Automated tools are perceived as a longer-term investment, offering scalability and integration potential but still requiring caution and significant adaptation before their benefits can be fully realised.

Cost benefit assessment

The results highlight a clear contrast between the two categories of digital compliance tools. Specialised compliance software shows a balanced profile: while labour costs remain moderate due to training and system use, the main financial burden stems from licensing and integration expenses. Stakeholders strongly emphasised gains in efficiency and accuracy, particularly through built-in templates, automated documentation functions, and faster preparation for audits. However, real-time monitoring and seamless integration with EU-level platforms remain underdeveloped, limiting the full potential of these tools.

Table 27. Qualitative assessment of the costs and benefits associated with advanced reporting techniques

Type of reporting technique	Costs				Relative Benefits			
	Labour cost	Financial cost	Technical cost	Human capital	Efficiency	Accuracy	Audit readiness and traceability	Real-time monitoring
Advanced tools – specialised software	Reasonable	Reasonable	Reasonable	Reasonable	Significant	Significant	Reasonable	Reasonable
Fully automated tools	Reasonable	Significant	Significant	Significant	Significant	Reasonable	Significant	Significant

Source: The data used to create this table were collected through a survey of industry associations.

Fully automated systems, by contrast, present a far more demanding cost structure, with significant financial, technical, and human capital investments required. Integration with hospital IT systems and EU portals was highlighted as particularly costly, alongside the need for retraining staff to manage and validate automated outputs. Nonetheless, the benefits profile is correspondingly strong: automation is seen as delivering major gains in efficiency, accuracy, and traceability, notably by eliminating manual uploads to EU portals such as CTIS and EUDAMED, reducing recurring administrative submissions, and strengthening audit readiness with tamper-proof digital trails.

The survey suggests that specialised compliance software currently represents the most practical and cost-effective solution, balancing moderate costs with substantial gains in efficiency and accuracy. Fully automated tools, while offering the highest potential benefits, remain experimental and resource-intensive, making them better suited to large, technologically advanced organisations. For smaller or resource-constrained actors, automation is not yet a feasible mainstream option.

3.5.5 Policy recommendations

Survey findings and interviews highlight strong support for adopting advanced and automated compliance tools in healthcare, alongside recognition of persistent barriers. The following recommendations propose concrete operational actions to facilitate a fair, effective, and secure digital transition.

Public investment in advanced and automated tools

Ensure that healthcare providers of all sizes can access affordable, reliable digital compliance and reporting systems, especially for structured tasks such as safety reporting, clinical trial submissions, and EUDAMED registrations.

Operational actions:

- **Establish a joint EU–national funding mechanism** to support the development and maintenance of shared compliance platforms for smaller healthcare providers. These platforms should cover core regulatory obligations (e.g. EUDAMED uploads, pharmacovigilance reporting, and clinical trial submissions) and be accessible through secure web interfaces.
- **Destine existing EU funding, including from EU4Health and the Digital Europe Programme, to support shared compliance platforms for smaller**

healthcare providers and to pilot automated modules (e.g. templates for adverse event reporting) assessing their cost-effectiveness. These platforms should cover core regulatory obligations and be accessible through secure web interfaces.

- **Require periodic evaluation of uptake, usability, and impact** through national health digitalisation strategies, adapting funding and support measures to ensure equitable access

Reducing duplication and streamlining reporting

Simplify regulatory reporting by reducing overlaps across national, EU, and global systems and implementing a “once-only” principle for data submission.

Operational actions:

- **Align data requirements and submission templates** during the legislative drafting process for MDR, IVDR, CTR, pharmacovigilance, and paediatric medicines, ensuring mutual recognition of datasets across frameworks.
- **Promote the development of interoperable data spaces** and shared databases that allow authorities to reuse existing data rather than requesting re-entry from providers.
- **Pilot a unified “compliance dashboard”** at EU level integrating reporting across EUDAMED, CTIS, and pharmacovigilance platforms, reducing administrative burden for users.
- **Mandate the publication of standardised data models and APIs** for all new regulatory platforms, allowing providers to connect their internal systems without duplication

Strengthening infrastructure, interoperability, and standards

Build a secure, interoperable digital infrastructure to enable seamless data exchange between hospital systems, national databases, and EU-level platforms.

Operational actions:

- **Mandate EU-wide technical and semantic interoperability standards** for compliance data exchange (e.g. FHIR-based formats, metadata, and identifiers) as a precondition for certification of national and EU systems.
- **Allocate EU funding** for hospital IT upgrades that enable secure connectivity to EUDAMED, CTIS, and pharmacovigilance systems.
- **Require regular conformance testing and certification of compliance tools** against agreed standards, with public disclosure of results to promote transparency and market trust.
- **Integrate data security and privacy safeguards** (in line with the European Health Data Space framework) into all interoperability and infrastructure investments.

Building digital skills and human capital

Develop the digital and regulatory literacy of healthcare professionals to ensure the safe, effective, and confident use of advanced and automated compliance tools.

Operational actions:

- **Introduce dedicated training programmes** for compliance officers, regulatory staff, and clinical personnel through national health authorities and professional bodies, focusing on automation use, data integrity, and system validation.

- Use **EU funding** to finance reskilling and continuous professional development on digital compliance systems.
- **Encourage partnerships** between healthcare institutions, IT providers, and regulators to co-design practical training curricula and certification schemes.

3.6 Chemicals

3.6.1 Overview of the sector

This section offers an overview of the European chemicals industry. It begins by outlining the sector's core activities, followed by an analysis of key economic indicators.

3.6.1.1 Definition

Ranked as the **fourth largest manufacturing sector in Europe**, the chemical industry produces a **diverse range of products**, from petrochemicals to consumer chemicals. It corresponds to Division C20 of the NACE Rev. 2 classification, which includes the following subgroups:

- C20.1: Basic chemicals, fertilizers and nitrogen compounds, plastics, and synthetic rubber in primary forms (further divided into seven classes)
- C20.2: Pesticides and other agrochemical products
- C20.3: Paints, varnishes and similar coatings, printing ink, and mastics
- C20.4: Soap and detergents, cleaning and polishing preparations, perfumes, and toilet preparations (subdivided into two classes)
- C20.5: Other chemical products (subdivided into four classes)
- C20.6: Man-made fibres

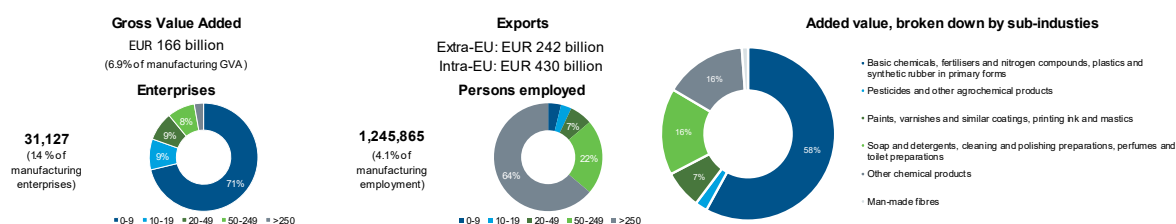
A key distinction within the chemical industry lies between **upstream and downstream activities**. Upstream refers to the early stages of production, primarily involving the manufacture of basic substances such as petrochemicals, basic inorganic and organic chemicals, and industrial gases. In contrast, downstream activities encompass the later stages of production, typically focused on formulation, specialty chemicals, and consumer-oriented products. Downstream sectors include paints, coatings, and adhesives, soaps, detergents, and cosmetics.

3.6.1.2 Size, value, employment

The chemicals industry is a **major contributor to the EU economy**, supplying both consumer goods and critical inputs to other sectors. **In 2022, it generated nearly EUR 166 billion in value added**, representing 6.9% of total manufacturing value added. The sector exhibits significant cross-border integration, with exports to non-EU countries reaching EUR 242 billion and intra-EU exports totalling EUR 430 billion (see Figure 23).

Over 30,000 enterprises were active in the chemical sector in 2022, with more than 70% employing fewer than ten people. The industry **provided jobs for approximately 1.2 million individuals**, accounting for 4% of total employment in the manufacturing sector. Nearly two-thirds of these jobs were in large firms (i.e., employing more than 250 employees), while a substantial share (22%) was concentrated in medium-sized enterprises.

Figure 23. The European Chemicals Sector – Key Facts⁵³⁷



Sources: The data on gross value added, the number of enterprises, and employment are sourced from Eurostat's *Structural Business Statistics dataset* (online code: *sbs_sc_ovw*). Trade data on imports and exports are drawn from the Eurostat dataset *International trade of EU and non-EU countries since 2002 by SITC* (online code: *ds-059331*). All data refer to 2022.

Most enterprises in the chemical industry are engaged in the production of soap and detergents, cleaning and polishing products, perfumes, and toiletries – accounting for 40% of all firms. This is followed by manufacturers of basic chemicals, fertilizers and nitrogen compounds, plastics, and synthetic rubber in primary forms, which represent 27% of the total. These two segments also concentrate the majority of employment, making up 23% and 45% of the workforce, respectively. This ranking remains consistent when examining the distribution of added value across sectors.

3.6.2 Regulatory landscape, reporting and compliance activities

The chemical sector is subject to extensive legal obligations under EU legislation, with companies required to comply with a wide range of regulations and directives – from substance registration and emissions reporting to sustainability disclosures. Meeting these requirements entails substantial efforts in data collection, processing, and reporting to official institutions. This section provides an overview of legislative acts that impose reporting obligations on the chemical industry, selected according to the criteria outlined in the methodological section (see Section 3.1.1). Priority was given to legislation with broad applicability and tasks that are potentially automatable.⁵³⁸ A detailed description of the legislative acts covered is provided in Annex A3.4.

3.6.2.1 Legal and regulatory framework

Registration, evaluation, authorisation and restriction of chemicals (REACH)

The regulation on the registration, evaluation, authorisation and restriction of chemicals (REACH) (Regulation (EC) No 1907/2006) establishes specific obligations for economic operators manufacturing, importing, or using chemicals. The regulation aims to ensure the safe use of chemical substances and reduce risks to human health and the environment.

To this end, economic operators must **register** all substances manufactured or imported (on their own or in preparations) in quantities that exceed 1 tonne per year (Article 6). Under Article 14, a Chemical Safety Report (CSR) is required during the registration process if manufacturers or importers place on the market more than 10 tonnes of a substance per year per registrant. Registrants are also required to keep the CSR available

⁵³⁷ The trade data focusing on the SITC two-digit categories 51 (Organic chemicals), 52 (Inorganic chemicals), 53 (Dyeing, tanning and colouring materials), 55 (Essential oils and resinoids and perfume materials; toilet, polishing and cleansing preparations), 56 (Fertilizers – other than those of group 272), 57 (Plastics in primary forms), and 59 (Chemical materials and products, n.e.s.).

⁵³⁸ The Directorate-General for Internal Market, Industry, Entrepreneurship and SMEs (DG GROW) has compiled a list of EU regulations and directives relevant to the chemicals sector, https://single-market-economy.ec.europa.eu/sectors/chemicals/chemicals-legislation_en.

and up to date. The registration obligation extends to substances contained in articles, provided that the total quantity is above 1 tonne per year and the substance is intended to be released under normal or reasonably foreseeable conditions of use (Article 7.1). Before registering a substance, however, manufacturers and importers must **submit an inquiry** dossier to ECHA to determine whether the substance has already been registered and to facilitate data sharing (Article 26). Article 22 stipulates the **obligation to update a registration** whenever new relevant information becomes available and share it with the agency.

REACH also establishes several **notification obligations**. Producers or importers are required to notify the ECHA if a substance in articles meets the criteria in Article 57, is identified in accordance with Article 59.1 (Article 7.2), is present in those articles in quantities exceeding one tonne per producer or importer per year and above a concentration of 0.1% weight by weight. They must also notify ECHA when substances are used for product and process-oriented research and development (PPORD) to benefit from a temporary exemption from registration (Article 9).

Downstream users are subject to additional reporting obligations. According to Article 38, downstream users must notify ECHA of unsupported uses, that is uses outside of the supplier's exposure scenario. They must also notify the ECHA of any classification differences relative to that provided by the supplier. Additionally, downstream users using a substance from the Authorisation List must inform ECHA of their use within three months of first supply (Article 66).

Manufacturers must submit **applications for authorisation** to the European Chemicals Agency (ECHA) if they wish to continue using or placing on the market substances listed in Annex XIV of REACH, also known as the Authorisation List (Article 66). These substances are typically classified as Substances of Very High Concern (SVHC) due to their hazardous properties, such as being carcinogenic, mutagenic, toxic for reproduction, or persistent and bio-accumulative.

Lastly, entry 78 of Annex XVII introduces **reporting obligations for certain synthetic polymer microparticles** (SPM). Specifically, manufacturers, industrial downstream users, and suppliers placing SPM on the market for consumer or professional use for the first time must submit annual reports to ECHA, providing estimates of SPM emissions to the environment.

The regulation also sets an obligation to assemble and keep available the information for a period of 10 years after they were last manufactured, imported, supplied, or used the substance or preparation.

While not directed at official institutions, Article 31 requires suppliers of a substance or preparation to provide recipients with a Safety Data Sheet (SDS). Under article 32, suppliers have a duty to communicate information down the supply chain for substances that do not require an SDS under Article 31. Producers are, however, required to share sufficient information with consumers and professional customers on substances included in the Candidate List to ensure safe use of articles (Article 33). Lastly, any actor in the supply chain is required to communicate information to the next actor or distributor up the supply chain (Article 34).

Legislation on the classification, labelling and packaging (CLP)

The Regulation on classification, labelling and packaging of substances and mixtures (CLP) (Regulation (EC) No 1272/2008) establishes legal obligations for companies to classify, label, and package chemicals prior to placing them on the market. The aim is to ensure that consumers and workers have access to clear and sufficient information when purchasing or using hazardous products. It also aligns EU rules with the classification and labelling criteria agreed at the UN level.

Under article 40, manufacturers have one month to **inform the ECHA of the classification and labelling (C&L) of a substance** referred to in Article 39 and which they have placed on the market. The information communicated to ECHA will thereafter be included in the C&L inventory if the substance has not already been registered under REACH. This obligation applies if the substance is hazardous, irrespective of tonnage. Companies are also required to notify the Agency when a modification to the classification and labelling has been adopted pursuant to the review in article 15(1). If the notification results in different entries for the same substance, the notifiers and registrants are expected to come to an agreement and inform the ECHA accordingly (Article 41).

Article 45 establishes an obligation for suppliers to **provide information about certain hazardous substances to national appointed bodies**. The latter then make this information available to poison centres who can provide assistance to citizens and medical professionals in case of an emergency. The requirements for poison centre notifications (PCN) have been harmonised by Annex VIII to the CLP Regulation, adopted in March 2017.

Lastly, as stipulated by Article 37, manufacturers, importers, and downstream can **submit a proposal for harmonised classification and labelling (CLH) of substances to the ECHA for industrial chemicals**. The objective is to ensure consistency in the classification of substances of high concern and improve safety.

The Biocidal Products Regulation (BPR)

The Biocidal Products Regulation (BPR) (Regulation (EU) No 528/2012) provides the legal framework for the marketing of biocidal products. These products are designed to protect humans, animals, materials, or articles from harmful organisms such as pests, bacteria, or fungi. The regulation aims to ensure a high level of protection for human health and the environment, while also facilitating the free movement of biocidal products across Member States.

According to Article 4, companies are required to **obtain approval for the active substances** used in their products. Article 6 outlines the specific data requirements for substance approval applications. In addition, Article 17 stipulates that **biocidal products themselves must be authorised before being placed on the market**.

The regulation also imposes several **notification obligations**. Article 47 requires authorisation holders to inform the competent authority of any newly identified adverse effects, evidence of resistance development, or lack of product efficacy. Article 95 mandates that any entity intending to place an active substance on the market must submit a dossier, a letter of access, or a reference to a dossier with expired data protection. Article 58 addresses consumer-facing obligations, including labelling and the provision of information upon request.

To reduce unnecessary testing and associated costs, Article 62 introduces a **data-sharing obligation**, requiring companies to share results from tests and studies where applicable. Manufacturers placing biocidal products on the market are required to maintain comprehensive documentation, either in paper or electronic format. This includes SDS, specifications of active substances and ingredients, detailed records of manufacturing operations, results of internal quality control procedures, and identification of production batches (Article 65.2).

The Fertilising Products Regulation (FPR)

The FPR applies to a wide range of fertilizing products and aims to ensure that they meet specific environmental and safety standards before being marketed in the EU.

The reporting obligations for manufactures are specified in Article 6. They are required to **prepare a technical file** before the product is placed on the market and **undergo a conformity assessment** (Article 6.2), **keep the dossier for five years** (Article 6.3),

maintain complaint and recall registers (Article 6.4), **label and identify each batch** (Article 6.7), **notify authorities of risks** (Article 6.8), and **supply the full dossier on request** (Article 6.9). They also need to **issue an EU declaration of conformity** and **affix the CE mark** (Article 16).

Corporate sustainability reporting (CSRD)

The **Corporate Sustainability Reporting Directive (CSRD)**, formally known as **Directive (EU) 2022/2464**, establishes enhanced sustainability disclosure requirements for large companies within the European Union. These requirements apply to entities that exceed thresholds in total assets (EUR 20 million), net turnover (EUR 40 million), or workforce size (more than 250 employees).

To ensure consistency and comparability across disclosures, the directive mandates the use of the European Sustainability Reporting Standards (ESRS), developed by the European Financial Reporting Advisory Group (EFRAG) and formally adopted by the European Commission. Their objective is to ensure consistency and comparability across reports. The ESRS framework comprises two cross-cutting standards (i.e., general disclosures), alongside ten topical standards that cover specific environmental, social, and governance (ESG) issues.

Articles 19a and 29a of the directive outline the core reporting obligations. Article 19a requires companies to **disclose ESG-related information in line with the double materiality approach** – companies must report on how their activities affect sustainability matters, and how sustainability matters create risks or opportunities for the company. Article 29a **mandates the use of the ESRS framework to standardize reporting practices**.

The **implementation of the CSRD is being phased in over time**. Companies already subject to the NFRD are expected to begin reporting in 2025 for the 2024 financial year. Other large entities and listed small and medium-sized enterprises (SMEs) will follow in subsequent phases. In the meantime, the European Commission has temporarily paused the introduction of additional reporting requirements. A revised version of the ESRS is currently under development and is expected to apply to reporting for the 2027 financial year.

Other legislation

The **Waste Framework Directive (WFD)** (Directive 2008/98/EC) sets the legal framework for waste management across the EU. Under the WFD, companies that supply articles containing substances of very high concern (SVHCs) on the Candidate List in a concentration exceeding 0.1% weight per weight are required to **submit information on these articles to ECHA**.

Under the **Prior Informed Consent (PIC) Regulation** (Regulation (EU) No 649/2012), when exporting a chemical listed in Annex I, exporters based in an EU Member State must notify the designated national authority of their country at least 35 days before the first export each year. In addition to the export notification requirement, exports of certain chemicals can only occur if the importing country outside the EU has granted a valid explicit consent. Under PIC, exporters must also submit an annual report to their Designated National Authority detailing the exact quantities of Annex I chemicals exported to or imported from non-EU countries, along with any additional information requested for regulatory compliance.

3.6.2.2 Reporting and compliance activities

Registration, evaluation, authorisation and restriction of chemicals (REACH)

Before initiating the registration process, **companies must submit an inquiry to the ECHA** and confirm whether the substance they intend to register has not already

undergone this procedure. Specifically, they must **submit an electronic dossier** that contains a complete and clear description of the manufactured or imported substance. As all REACH-related dossiers, the inquiry dossier must be generated using the latest version of IUCLID. The dossier is then submitted manually via REACH-IT, ECHA's dedicated online platform.

Illustrative example: IUCLID

IUCLID 6 (International Uniform Chemical Information Database) is a software application developed by the European Chemicals Agency in collaboration with the Organisation for Economic Co-operation Development (OECD). It allows users to create, store, maintain, and exchange data on chemical substances, supporting the preparation of compliance dossiers under various EU legislative acts, including REACH, CLP and PCN, BPR, WFD, and the Drinking Water Directive (DWD).⁵³⁹

One of IUCLID's key advantages is its ability to be customized and configured to suit different contexts and requirements. It also employs harmonized electronic formats, ensuring that the information recorded complies with international standards. IUCLID also offers users the possibility to validate the data before submission, making sure that the dossier is complete.

IUCLID can be downloaded free of charge from ECHA's website and installed directly on a computer. The desktop interface provides access to a range of functionalities (see Figure 24). Through the Substance section, users can enter substance data, which can then be edited, reused, and exchanged. The Dossier function generates a fixed snapshot of the substance dataset, representing the version that will be submitted to ECHA. Import and bulk export features enable data exchange as .i6z files.

Figure 24. IUCLID 6 – Desktop interface



Source: ECHA. *Introduction to IUCLID*.

To register substances, **companies must submit a registration dossier to ECHA**. The required information varies depending on the quantity of the substance, as outlined in Annexes VI to X of the REACH Regulation. The data is compiled in a technical dossier and includes details on substance identity, physicochemical properties, fate, (eco)-toxicological characteristics, and use and exposure information. For substances exceeding ten tonnes per year, a Chemical Safety Report (CSR) must also be submitted, documenting the outcomes of the chemical safety assessment. In most cases, registration is subject to a fee. To notify ECHA of substances in articles listed on the Candidate List, **producers and importers must prepare a dossier that mentions the identity and classification of the substance**, a brief description of its use, and the tonnage range. When updating a previous registration due to an increase in the tonnage

⁵³⁹ Under Article 11 of the Drinking Water Directive (Directive 2020/2184), economic operators have an obligation to notify the ECHA if they intend to propose a modification to the European positive lists – the official list of substances that are explicitly authorized for use in the manufacturing materials in contact with drinking water. They must also submit applications to add new entries and amend or remove existing entries. The submission will start in 2026.

band, which requires providing additional information, the inquiry procedure must be followed.

Illustrative example: REACH-IT

REACH-IT is a dedicated platform that enables economic operators, Member State competent authorities, and ECHA to securely exchange data and communications. Manufacturers and importers are required to use REACH-IT to submit and manage their dossiers under the REACH and CLP regulations.

To use REACH-IT, companies must first register online. Upon registration, they can manage the company information and authorised users. To submit a dossier, they have to first log in and use the specialised platform to send their data.

There are three modalities for submitting a dossier using REACH-IT: uploading a IUCLID dossier, preparing it online within the application, or accessing webforms (see Figure 25). Clicking on the *Upload IUCLID dossier* button then allows users to select the type of dossier they would like to submit (i.e., registration, inquiry). A step-by-step submission wizard will then provide information on how to finalise the submission process.

Figure 25. REACH-IT – Homepage



Source: ECHA). *Discover REACH-IT*.

The *Prepare online in REACH-IT* button can be used to prepare IUCLID dossiers for C&L notifications (see below) and registration dossiers in the case of a joint submission. The *Webform applications* concerns the submission of alternative chemical name requests, downstream user reports (for authorisation, registration, and notification for substances in articles).

To **notify substances in articles to ECHA**, companies need to **generate a dossier** reporting the identity of the substance, the tonnage range of the substance in the notified article, and a brief description of the use(s) of substance in the article(s) and of the uses of the article(s). The notification must be submitted within six months of the inclusion of the substance in the Candidate list. For the **PPORD notification**, producers or importers must **report the substance identity**, including its hazard classification, **information on the PPORD activities**, as well as the **quantity of the substance**.

To **report unsupported uses to ECHA under Article 38(1)**, downstream users must **submit a report** that includes a brief, general description of their use and the associated conditions of use. If testing on vertebrates is required, a testing proposal must also be provided. This report should be submitted within six months of receiving the supplier's safety data sheet. **Notification of use of a substance included in the Authorisation List under Article 66** needs to include the authorisation number and contact information. Downstream users may also provide information on the typical annual volume, the number of staff using the substance, a brief description of the use, and involvement in potential substitution activities.

Applications for authorisation to continue using substances listed in Annex XIV of REACH must include **details of the intended use, a CSR, and an analysis of alternatives as well as a socio-economic assessment**. Applicants need to show whether the specific use of the substance is safe and that there are either no alternatives to its use or the socio-economic benefits outweigh the costs.

Starting from October 2025, manufacturers, industrial downstream users, and suppliers will need to **report estimates on annual emissions of SPM into the environment**. The reports must be submitted every year by 31 May for emissions occurring during the previous year.

Illustrative examples: QSAR toolbox and CHESAR

The QSAR Toolbox is a free software application developed to support companies in generating hazard assessments for chemical substances. As a computational tool, it offers an alternative to animal testing by predicting the toxicity of substances before they are manufactured.

The process begins with the identification of the target substance and the selection of the property to be assessed. Next, the profiling module is used to characterize the substance and identify structurally or mechanistically similar analogues. The data module provides access to extensive experimental datasets and chemical inventories.

A critical step is the grouping module, where substances are organized into toxicologically relevant categories based on shared characteristics. In the data gap filling module, users can estimate missing data for the target substance using information from analogues. Finally, the reporting module enables users to export the results in a structured format suitable for regulatory submissions (Figure 26).

Figure 26. Example of a data matrix report generated an excel file

Target chemical		Analogues			
Target chemical		Analogue #1	Analogue #2	Analogue #3	Analogue #4
Structure		Analogues			
Chemical ID		Chemical ID			
CAS number		66-25-1	112-64-7	123-68-6	110-62-9
Chemical name		Water	Undecanal	Propanal	Valeral
Other identifier					
Status		CCCCC=O	CCCCCCCCC=O	CCCC=O	CCCCC=O
Estimated phys-chem properties					
Boiling point	°C	132	4.25	59.8	109
Melting point	°C	1.8	4.25	0.325	1.31
Molecular weight	Da	100	170	58.1	86.1
Vapor Pressure (Antoine method)	mm Hg	10.3	0.0038	0.254	329
Water solubility	mg/L	3.03E+03	14.3	4.3E+04	9.72E+03
Profiling results					
Profiles used for grouping/subcategorization					
Aldehydes (Acute toxicity)	Aldehydes (Acute toxicity)	Aldehydes (Acute toxicity)	Aldehydes (Acute toxicity)	Aldehydes (Acute toxicity)	Aldehydes (Acute toxicity)
Protein binding by QAS (subcategorization)	Schiff base formation => Schiff base	Schiff base formation => Schiff base	Schiff base formation => Schiff base	Schiff base formation => Schiff base	Schiff base formation => Schiff base
Organic functional groups (subcategorization)	Aldehyde	Aldehyde	Aldehyde	Aldehyde	Aldehyde
Endpoint specific	Aldehydes	Aldehydes	Aldehydes	Aldehydes	Aldehydes
Acute aquatic toxicity MOA by QAS	Class 3 (unspecific reactivity)	Class 3 (unspecific reactivity)	Class 3 (unspecific reactivity)	Class 3 (unspecific reactivity)	Class 3 (unspecific reactivity)
Acute aquatic toxicity classification by Verhaar	Class 3 (unspecific reactivity)	Class 3 (unspecific reactivity)	Class 3 (unspecific reactivity)	Class 3 (unspecific reactivity)	Class 3 (unspecific reactivity)
Endpoint	54.5/63.3 %	37.5/46 %	40/45.7 %	35.3/46.7 %	40/42.3 %
Measured and predicted data					
Data used for prediction					
substance	endpoint	value	unit	species, duration, test type, type of method, assay, assay test	value
Aquatic Toxicity	LC50	3.0	mg/L	Test organisms: <i>Daphnia magna</i>	3.0
Physical Chemical Properties				species, duration, test type, type of method, assay, assay test	

Source: <https://qsartoolbox.org/features/reporting/>.

Another application developed by ECHA to support companies in their reporting activities is CHESAR. This tool facilitates the creation of chemical safety assessments, chemical safety reports, and exposure scenarios. For instance, CHESAR offers functionalities to import substance data from IUCLID and automatically generate CSRs and ESs. While users still need to enter some data manually, the software allows for the automatic calculation of certain measures.

Classification, labelling and packaging (CLP)

To submit a C&L notification to ECHA, **manufacturers and importers must prepare an IUCLID dossier**. This dossier should include detailed information on the substance's identity – such as its name, identifiers, molecular and structural formula, composition, nature, and the amount of additives. It must also provide the substance's classification in accordance with CLP criteria, a justification for “no classification” where applicable,

specific concentration limits and M-factors, and the relevant labelling information. Notifications are submitted electronically, either via the REACH-IT portal or through ECHA's system-to-system submission service based on the XML schema.

A PCN notification must include the following information: unique formula identifier, the full chemical composition, the classification and labelling, toxicological information, product category, and information identifying the submitter, the product, and the type of use (i.e., consumer, professional, or industrial). The dossiers must be prepared in IUCLID and are submitted via REACH-IT or system-to-system.

A CLH dossier must include a CLH report prepared in the format specified by ECHA. This report should cover the substance's identity, the proposed harmonised classification and labelling, a scientific justification for the proposal, and detailed summaries of relevant studies. The submission is made electronically via ECHA's webform.

Illustrative example: CLH submission webform

The submission of a CLH report occurs via a webform specially designed for this purpose. The webform can be used to submit CLH pre-intentions, CLH intentions (and any updates), CLH proposals, and RCOM documents. The webform consists of several parts: submission related information; identification of dossier submitter; contact person; declaration; existing or proposed substance identity in Annex VI; information on the substance; further information on the substance and relevant legislative act; classification and labelling table (Figure 27).

Figure 27. Example of a section from the CLH webform

Classification and labelling table
Please provide the preliminary classification as foreseen at the moment by ticking the relevant tick box(es) in the corresponding sections.

Hazard	Current entry in Annex VI	Proposed future entry (if there is a current entry, please include also those hazard classes that are not addressed in the proposal)
		☐ Physical hazards
		☐ Health hazards
		☐ Environmental hazards
		☐ Additional hazards

SCLs, labelling elements, notes, ATEs ☐ Yes ☐ No

Expected date of submission: DD/MM/YYYY

Other remarks:

Upload the CLH dossier and any relevant documents here. If the file is confidential, please indicate this in the name of the file. Non-confidential documents will be shared with stakeholder organisations. Confidential documents will be shared only with MSCAs, the Committees, the European Commission and ECHA.

Attachments:

Maximum file size is 100 MB Maximum file size is 100 MB
(If you would like to submit more than one document, please create a zip archive where you include all files and upload the zip file as attachment. Maximum file size is 100 MB.)

Please note that if an access to documents application pursuant to Regulation (EC) No 1049/2001 on public access to documents is received regarding this information, in order to define its position ECHA will have to first perform an assessment of the content of this information. Therefore, you are invited to indicate below any reason for which disclosure of this information should be denied. Please note that ECHA can only examine reasons appearing in Article 4 of the Regulation (EC) No 1049/2001 on public access to documents as valid.

I have the following reasons enumerated in Article 4(1) or 4(2) of Regulation (EC) No 1049/2001 regarding public access to documents why the information submitted as confidential cannot be disclosed to persons requesting access to documents (please explain below in the commenting field those reasons; a reason could be that the protection of your commercial interests, including intellectual property, would be undermined):

Reason:

Source: ECHA. CLH submission webform.

The Biocidal Products Regulation (BPR)

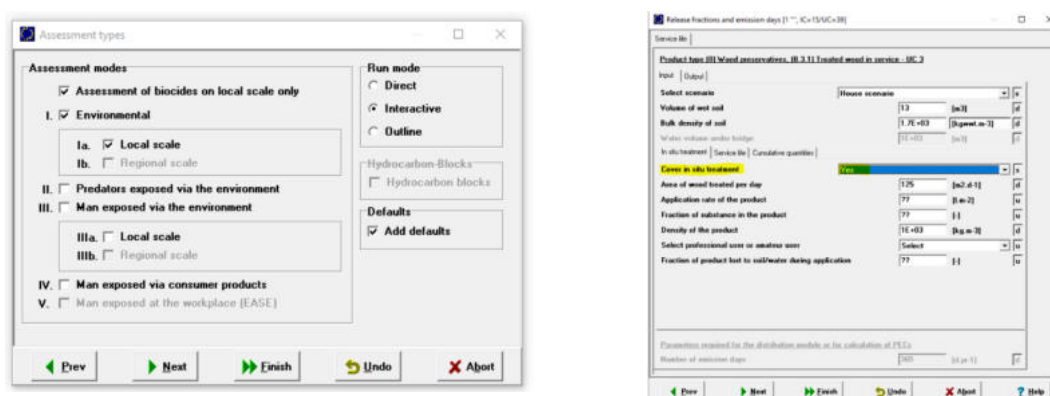
Under the BPR, **companies are required to apply for active substance approval** by submitting a dossier to ECHA. Article 6 lists the data requirement for the application. Companies first submit a dossier to ECHA, who performs a validation check and forwards it to a designated Member State competent authority to conduct a completeness check and an evaluation within one year. The outcome of this assessment is then submitted to ECHA's Biocidal Products Committee, which issues an opinion within 270 days. The European Commission, based on ECHA's opinion, is responsible for approving the substance for specific uses. The approval is valid for a period not exceeding 10 years, after which companies must apply for renewal.

Illustrative example: EUSES

EUSES (European Union System for the Evaluation of Substances) is a software designed by ECHA to assist companies when preparing their environmental exposure assessments, as requested under REACH and BPR. For instance, in version 2.2.0, users can access the SimpleTreat Model (4.0) to evaluate the fate and the distribution of a substance in a sewage plant.

To run an assessment in EUSES, users should first open the application, navigate to the File menu, and start a New Study. They then need to define the type of the assessment and the settings of the model. For example, Figure 28 below illustrates the recommended settings for assessing a biocidal substance. Note that the application also offers three types of run modes: direct, requiring minimum input from the user, intermediate modes, which shows intermediate results and allowing for default parameters and intermediate results to be overwritten, and outline mode, for experienced users. Next, users input the data, select the appropriate emission scenarios, and configure the model parameters. Once these steps are complete, the calculation can be executed.

Figure 28. Defining the settings for the assessment of a biocidal substance



Source: ECHA. How to use EUSES 2.2.

The dossier needs to follow the IUCLID format and is submitted using the tool R4BP 3. Approval of biocidal products also involves the submission of a IUCLID dossier via the submission tool R4BP 3.

Illustrative example: R4BP 3

The R4BP 3 constitutes the central hub used for the submission of applications to ECHA and the National Authorities. It is designed to simplify compliance with legislative requirements under BPR and assures a secure information channel between economic operators and the relevant authorities. To initiate a new submission, users should click on the *New Application* button (Figure 29) in the main interface and follow the subsequent steps.

Figure 29. R4BP 3 – Interface

The screenshot shows the ECHA R4BP 3 interface. At the top, there's a navigation bar with 'ECHA' logo and tabs for 'TASKS', 'MESSAGES', 'CASES', 'ASSETS', and 'NEW APPLICATION'. Below this, the 'Tasks due to users' section displays a table with columns: 'Task name', 'Product/Substance name', 'Action subelement', 'Due date', and 'Task type'. It lists tasks like 'No-OP certificate (NOC)', 'Initial assessment', and 'Potassium (K) Neonicotinoid (Potassium Sorbate)'. To the right, a 'GOODS TO REVIEW' section contains a list of tasks with status indicators (e.g., 'Not used', 'Used'). Below the main table, the 'My messages' section shows a table with columns: 'From', 'To/Subject', 'Product/Substance name', and 'Received'. It lists messages from 'Agency' and 'HSECA-Bulgaria'.

Source: ECHA (2017). *Introduction to R4BP 3 and SPC editor*.

The Fertilising Products Regulation (FPR)

To comply with the Fertilising Products Regulation, companies must ensure their products meet safety, quality, and environmental standards, maintain technical documentation, and perform conformity assessments using the appropriate module. They must prepare an EU Declaration of Conformity, affix CE marking, and provide mandatory labelling and safety information.

Corporate sustainability

To comply with the CSRD, **chemical companies must undertake extensive data collection and align their reporting with ESRS standards**. A central element of this process is the double materiality assessment, which involves identifying potential material impacts, risks, and opportunities (IROs) for each ESG topic relevant to a company. Determining whether these IROs are material will arguably require the development of a structured methodology. This process demands significant cross-functional collaboration, engaging many teams across an organisation, including sustainability, finance, procurement, and risk management.⁵⁴⁰

To this aim, companies may have to **implement data management systems capable of handling large volumes of sustainability data, moving beyond spreadsheet-based techniques**. They will also need to prepare for audits by embedding internal controls, defining KPIs in line with ESRS, and ensuring governance oversight at board level.

Other legislation

Under the WFD, the information supplied by companies on SVHC on the Candidate List to ECHA includes information on the identity of the article, the name, concentration range and location of the substance in the article, and other information that enables safe use and information on proper management of the article after it becomes waste. The notification dossier is prepared using IUCLID and is submitted using ECHA's submission portal, manually or via system-to-system transmission. The data is then included in the SCIP database on Substances of Concern in articles as such or in complex objects (Products) and becomes available to consumers and waste operators.

⁵⁴⁰ European Chemical Industry Council (CEFIC). (2025, March). *Corporate Sustainability Reporting Directive (CSRD) handbook for the chemical sector*. Brussels: CEFIC. <https://cefic.org/resources/corporate-sustainability-reporting-directive-csrd-handbook-for-the-chemical-sector/>.

The export notification required by the PIC Regulation must include details on the substance's identity, export information (destination, quantities, intended use, exporter/importer details), safety precautions, summaries of physico-chemical and (eco-)toxicological properties, its uses in the EU, and any regulatory restrictions with their justification.

3.6.2.3 Overview of existing challenges

Recent legislative developments in the chemicals sector were initially embraced by industry, as they aimed to replace a fragmented patchwork of national reporting systems with a more harmonized and streamlined framework. However, industry representatives argue that, rather than simplifying reporting, **these legislative measures have aimed to create comprehensive databases**, obliging manufacturers to submit large volumes of information – sometimes, without clear relevance or tangible benefits for society or research. **This has significantly increased the administrative burden on companies**, forcing them to allocate additional resources to data collection and processing, and ultimately undermining their competitiveness.

Moreover, despite the centralization of reporting through ECHA, **many member states still require companies to submit information to national authorities**, often in varying formats and with different content requirements. This duplication significantly increases the administrative burden, compounded by the fact that many reporting processes involve fees. The financial impact is particularly pronounced in sectors where product formulations change frequently, as companies must update their reports regularly, driving up compliance costs.

Furthermore, **some reporting obligations are perceived as particularly demanding**. For example, microplastics reporting requires manufacturers to consider all potential pathways through which their activities could release particles into the environment – from production and supply chain to the final use of the product. These requirements not only add administrative burden but also involve complex calculations that many small- and medium-sized companies lack the technical expertise to perform.

Similarly, **downstream users encounter significant challenges** when compiling safety data sheets due to the complexity and variability of the information they receive. When purchasing ingredients – whether individual substances or mixtures – from multiple suppliers, they often receive documentation in different languages and formats, ranging from paper-based to electronic. To create their own exposure scenarios, they must process this diverse data and extract all relevant details, a task that is both time-consuming and highly labour-intensive.

In summary, **companies in the chemicals sector face three interrelated types of challenges**. First, regulatory complexity persists due to **overlapping EU and national requirements** and frequent legislative updates. Second, **technical and operational hurdles** arise from the use of sophisticated tools such as IUCLID and the need for specialized expertise to meet demanding obligations like microplastics reporting. Finally, these factors combine to create a **significant administrative and financial burden**, as duplicated reporting, recurring updates, and compliance costs weigh heavily on businesses – particularly SMEs and downstream users.

3.6.3 Overview of reporting and compliance techniques

This section explores the approaches used to achieve regulatory compliance and meet reporting obligations in the chemicals sector, outlining the main techniques currently applied and evaluating the extent to which they have been effectively implemented across the industry.

3.6.3.1 Description of techniques and tools

Compliance and reporting techniques in the chemicals sector can be broadly classified as standard, advanced, or automated, depending on the level of automation and the degree of human input required. Each category is described in greater detail below, based on observed practices and industry evidence.

Standard techniques

Standard techniques rely heavily on manual work, with minimal or no automation of repetitive tasks such as data collection or processing. These approaches typically use spreadsheet-based tools such as Excel or Google Sheets and, in some exceptional cases, even paper documentation.

Interviews with industry representatives revealed several examples of spreadsheet-based solutions being used for compliance and reporting purposes. Excel files are not only employed to store data and populate templates or tables but often include built-in calculation modules to estimate exposure levels and other key parameters. For instance, **one industry association representing downstream users of chemicals has developed sophisticated Excel templates that allow members to upload their data and subsequently transfer it into IUCLID.**

Spreadsheets are also commonly used to format data for submission to authorities. In practice, **companies extract information from various internal systems** and manually compile it into Excel files, which are then adjusted to meet regulatory requirements.

Advanced techniques

Advanced techniques enable the digitalisation of certain tasks, although human input and oversight are still required throughout the reporting process. These techniques are typically implemented through commercial software or, in some cases, custom in-house applications developed by local IT teams.

Enterprise Resource Planning (ERP) systems are a type of advanced software that support regulatory compliance and reporting activities. They enable organizations of all sizes to integrate core business processes across departments such as finance, human resources, manufacturing, supply chain, and procurement. Typically composed of modular applications connected to a centralized database, ERP systems facilitate real-time access to accurate and consistent data. Additionally, ERP systems often include features for compliance tracking and may support the automated generation of documents, such as SDSs.

Illustrative example: SAP S/4HANA for product compliance

SAP S/4HANA is a commercial ERP solution that supports the automation of various reporting and compliance obligations. Its dedicated module, *SAP S/4HANA for Product Compliance*, enables organizations to generate safety data sheets (SDSs) and product labels that align with current regulatory standards, including EU-specific requirements. The system also includes functionality to automatically track regulated substances, monitoring quantities that are purchased, produced, sold, imported, or exported. Furthermore, it can assess product compliance against applicable legislation, particularly in relation to registration requirements and quantity thresholds.

Most companies also adopt software applications that integrate Environment, Health, and Safety (EHS) systems. An EHS system is a structured framework of policies, procedures, processes, and practices designed to manage environmental, health, and safety risks within an organization. In practice, EHS modules are often linked to various business functions – such as procurement, inventory and logistics, or

production – and automate compliance checks, enabling real-time monitoring and reporting.

While most commonly used ERP systems are sector-agnostic and applicable across various industries, **there are also specialized commercial solutions tailored for companies in the chemical sector.** These solutions address industry-specific requirements such as Safety Data Sheet (SDS) authoring and distribution, chemical inventory and hazard management, substance volume tracking and regulatory reporting, and poison centre notifications (PCNs).

Illustrative example: LISAM

LISAM Systems provides specialized software solutions for EHS compliance, with a strong focus on the chemical industry. Its applications enable companies to create, manage, and distribute SDSs and labels, ensuring compliance with regulations such as REACH and CLP. LISAM supports comprehensive chemical inventory management, hazard communication, and automated regulatory reporting, including volume tracking. The system integrates regulatory content libraries and offers multilingual document generation, making it easier for organizations to maintain compliance across multiple jurisdictions.

In addition to sector-specific tools, specialized software solutions are designed to meet specific sustainability reporting requirements. Corporate Sustainability Data Systems (CSDS), for example, are integrated platforms that collect, manage, analyse, and report data related to an organization's environmental, social, and governance (ESG) performance. These systems typically connect to other data sources—such as ERP and operational platforms—and offer features like dashboards, scenario modelling, and automated reporting. One of their key strengths is the ability to process and consolidate large volumes of data, ensuring accuracy and consistency across complex reporting frameworks.

Illustrative example: Workiva

Workiva is a commercial software that provides a unified, cloud-based platform to help organizations comply with the CSRD. The platform integrates financial and sustainability data in a single, controlled environment, ensuring audit-ready, assurance-compliant reporting.

Key Capabilities

- **Integrated Data Management:** Automates data collection from multiple sources using 60+ pre-built connectors and APIs.
- **Double Materiality & Gap Analysis:** Supports ESRS-aligned assessments to identify material topics, impacts, risks, and opportunities.
- **Carbon Management:** Enables GHG emissions tracking with audit-ready calculations aligned to the GHG Protocol.
- **Governance & Controls:** Provides real-time audit trails, risk alignment, and automated evidence for assurance readiness.
- **AI-Powered Efficiency:** Uses AI to benchmark sustainability strategies, draft ESRS-aligned narratives, and streamline reporting.
- **XBRL Tagging:** Ensures digital tagging for ESRS and Article 8 taxonomies, meeting EU reporting requirements.

Automated techniques

AI-powered tools enhance regulatory compliance and reporting by digitizing data collection, improving accuracy, and providing predictive insights. It further integrates

multiple data sources, can extract key information from unstructured documents, and ensures compliance through automated reporting.

An example of AI applications to regulatory reporting is the use of intelligent chatbots that provide **real-time guidance on disclosure requirements and share best practices from the industry**. These tools not only improve efficiency but also help organizations make informed decisions and demonstrate compliance more effectively. Another important capability is automated document analysis, where AI reviews uploaded policies, extracts relevant details, and integrates them into reports, reducing manual effort and minimizing errors.

For example, some companies receive SDS from their suppliers in PDF format and use automated tools to extract the relevant information and import it into their internal systems. These solutions are often developed in collaboration with IT providers, with the main objective of leveraging all the data stored in their existing platforms and making it usable for compliance and operational purposes.

Blockchain can enhance compliance in the chemicals sector by providing a secure, transparent, and tamper-proof way to track data across the supply chain. Each transaction or update—such as the registration of substances, safety data sheet (SDS) information, or proof of regulatory checks—can be recorded on a distributed ledger, ensuring that all stakeholders have access to the same verified information. This improves traceability of chemicals from production to end use, reduces the risk of data manipulation, and simplifies audits by creating an immutable compliance record.

3.6.3.2 Current use of reporting techniques

The choice of the techniques used for compliance and reporting purposes in the chemicals sector is determined by both legal requirements and the specific operational features of the firms carrying them out. **Stakeholder interviews indicate that small and medium-sized enterprises (SMEs)**, which typically manage a more limited portfolio of substances and products, **tend to rely more heavily on standard tools**. In contrast, **larger firms are more likely to adopt advanced tools** such as commercially available software solutions. Even though these companies may internally develop and use automated systems, such tools do not appear to be employed for compliance and reporting. Table 28 below presents a qualitative assessment of the frequency with which these techniques are used across different legislative frameworks.

Table 28. Use of reporting techniques for carrying out compliance and reporting activities

Reporting activity	Standard tools	Advanced tools	Automated tools
Registration, evaluation, authorisation and restriction of chemicals (REACH)	Predominant	Supporting	Marginal
Classification, labelling, and packaging (CLP)	Predominant	Supporting	Marginal
The Biocidal Products Regulation (BPR)	Predominant	Supporting	Marginal
The Fertilising Products Regulation (FPR)	Inconclusive	Inconclusive	Inconclusive
Corporate sustainability reporting (CSRD)	Predominant	Supporting	Marginal

Source: Authors' elaboration based on feedback from industry representatives.

Reporting activities under **REACH** are typically carried out using **standard tools**, supported by the software developed by ECHA and other commercial solutions. Smaller companies, which often have less experience using IUCLID, rely almost exclusively on standard tools – such as Excel templates – to support dossier preparation. While some specialized software incorporates AI features, there is currently limited evidence that these tools enable full automation of the reporting process and that their use is widespread.

Since reporting under CLP is done through **IUCLID** dossiers, **standard tools are commonly used to prepare submissions, supported by ECHA software**. However, the submission procedure itself can vary, **enabling different levels of automation**. C&L notifications may be submitted manually via REACH-IT or through a System-to-System (S2S) service. The latter relies on XML schema definition (XSD) files and enables companies to integrate their own IT systems for dossier submission, significantly increasing automation. S2S also supports bulk submissions, which are then made available in REACH-IT and the C&L Inventory. The available evidence does not allow for an assessment of how widely this option is actually used within the industry.

PCNs are submitted through the **ECHA Submission Portal (PCN Portal)**, the EU's central platform that forwards notifications to national poison centres. Access requires a valid ECHA account and an associated company profile. While the process is largely manual, **an S2S option is available, enabling integration with a company's IT systems and offering a higher degree of automation**. As with C&L notifications, there is insufficient evidence to determine how widely this option is actually implemented across the industry.

By contrast, **CLH dossiers** are submitted via a dedicated webform on ECHA's website. Although this method also requires a REACH-IT account – some fields require information from it – the **process offers limited flexibility and remains predominantly manual**, making it less suitable for automation.

Evidence on the methods currently employed to ensure compliance under **BPR** is limited. However, given the reporting requirements (IUCLID dossier) and the submission channel (ECHA's R4BP 3 tool), **it is reasonable to assume that standard methods**, supported by advanced tools, are the predominant approach. For FPR, the diversity of activities involved, and the limited stakeholder feedback received do not provide sufficient basis to identify the techniques commonly used.

A recent report commissioned by the European Chemical Industry Council (CEFIC) on the implementation of sustainability reporting in the chemical sector indicates that most companies continue to rely heavily on spreadsheet-based solutions, with 84% of surveyed companies reporting their use.⁵⁴¹ Advanced tools are less prevalent: carbon calculation tools are employed by 55% of companies, while integrated ERP systems are used by 45%. There is, however, **evidence of a growing trend toward more sophisticated solutions**, such as Corporate Sustainability Data Systems (CSDS), which enable more efficient management of the large volumes of data required for sustainability reporting. Currently, only 19% of companies use CSDS, but 33% plan to adopt such tools in the future. The use of AI remains marginal, with 7% of companies currently applying it to sustainability reporting and 10% indicating plans for future implementation.

Overall, **reporting practices in the chemicals sector remain heterogeneous**, with a strong reliance on advanced tools and on traditional tools such as spreadsheets, particularly for sustainability reporting. Automation potential varies significantly across regulatory frameworks: C&L notifications and PCNs offer options for S2S integration, enabling higher efficiency, whereas CLH submissions remain largely manual. These differences highlight both the progress made and the challenges that persist in achieving streamlined, technology-driven reporting across the sector.

⁵⁴¹ European Chemical Industry Council (CEFIC). (2025, March). *Corporate Sustainability Reporting Directive (CSRD) handbook for the chemical sector*. Brussels: CEFIC. <https://cefic.org/resources/corporate-sustainability-reporting-directive-csrd-handbook-for-the-chemical-sector/>.

3.6.4 Challenges, barriers, and qualitative cost-benefit analysis

This section evaluates current reporting practices in the chemicals sector, based on desk research and stakeholder input. It outlines key challenges and barriers to automation and concludes with a qualitative assessment of the potential benefits of adopting advanced reporting solutions.

3.6.4.1 Overview of challenges per reporting technique

Traditional spreadsheet-based techniques, while widely adopted, present several significant limitations that affect both efficiency and accuracy. First, Excel **files are often used not only to store data but also to perform calculations**. This dual function introduces a critical challenge: ensuring that the underlying data is always up to date so that the resulting outputs are accurate. Any delay or inconsistency in updating the data can compromise the reliability of the analysis.

Second, these methods frequently **result in a duplication of effort**. Reports are typically submitted in Word format, with regulatory authorities often requesting access to the underlying data. This necessitates extensive manual transfer of information from Excel to Word, a process that is both time-consuming and resource-intensive, leading to inefficiencies and misallocation of resources.

Third, industry associations create Excel templates to assist their members in calculating specific metrics required for compliance reporting. However, **this process must be repeated** each time a new report is generated or when a formulation changes, further increasing the administrative burden and extending reporting timelines.

Finally, **spreadsheet-based approaches rely heavily on manual input**, which significantly increases the risk of human error. Even minor inaccuracies can propagate through calculations, resulting in erroneous reporting and potential compliance risks.

The use of **advanced tools** for regulatory reporting introduces its own set of challenges, particularly for small and medium-sized enterprises (SMEs). **Many companies encounter difficulties in meeting technical and reporting requirements** due to the complexity of these tools and the obligations they are designed to address. For instance, IUCLID, despite significant improvements, remains a **highly complex application** that requires substantial training for effective use. Industry representatives note that simply downloading and installing the software, along with reviewing existing guidance materials, is insufficient to ensure efficient and accurate reporting.

In addition to the steep learning curve, **firms must conduct additional quality assurance checks** to confirm that the outputs generated by these tools are logical and consistent with regulatory expectations. This adds another layer of complexity and resource demand, particularly for organizations with limited technical expertise.

Moreover, companies often receive input data from multiple clients in diverse formats and with varying levels of completeness. To leverage advanced tools effectively, a degree of standardization in data structure and content is essential. Without such harmonization, firms face significant challenges in consolidating and processing information.

Although **automated tools** are rarely used in practice, their potential adoption raises several important challenges. First, **automation inherently relies on strict standardization of processes and data formats**. However, regulatory reporting often necessitates a certain level of flexibility and granularity to accommodate variations in products, processes, and regulatory requirements. Striking the right balance between automation and adaptability remains a significant challenge.

Second, even in a fully automated environment, quality assurance remains indispensable. It is highly unlikely that a report generated entirely through automated mechanisms would be submitted to a regulatory agency without prior verification. Companies remain

concerned about the possibility of errors, which could lead to financial penalties, compliance breaches, or reputational damage. Consequently, additional checks and validations would still be required, potentially offsetting some of the efficiency gains promised by automation.

3.6.4.2 Barriers to adoption

Despite the potential benefits of automation, several significant barriers hinder its widespread adoption. First, achieving **harmonization across the supply chain is essential**. For automated systems to function effectively, all companies involved would need to adopt compatible tools and processes to ensure seamless communication and data exchange. This presents a major coordination challenge, compounded by a first-mover problem: individual firms may be reluctant to invest in automation unless others in the supply chain do the same.

Second, **interoperability with regulatory authorities' systems is critical**. Automated tools must be fully compatible with the IT infrastructure used by public institutions to ensure that data can be transmitted and processed without errors or delays. Without such interoperability, the efficiency gains promised by automation cannot be realized.

Third, **companies require legal certainty regarding the acceptance of outputs generated by automated tools**. Without clear regulatory guidance and assurances that automated reports will be recognized as compliant, firms are unlikely to commit resources to developing or implementing such systems.

Another important consideration is that companies typically operate interconnected IT systems, where even minor changes in one node can affect the entire infrastructure. **This means that the cost of implementing modifications** – however small – **can escalate significantly**, making it difficult to predict automation costs in advance.

Some activities are also unlikely to be automated. A prime example is registration under REACH. To prepare the required documentation, economic operators must conduct a series of tests – such as toxicology and ecotoxicity assessments. The results are then entered into a standardized template and submitted to the relevant authority, leaving little room for automation in the process.

Finally, **structural differences within the industry exacerbate these challenges**. Large companies often have the resources to develop centralized databases and sophisticated IT systems, enabling them to integrate automation more easily. In contrast, smaller firms typically lack the financial and technical capacity to make similar investments. They continue to rely on manual data collection and formatting, which limits their ability to adopt automated solutions and risks widening the gap between large and small players in the industry.

These barriers highlight that the transition to automated reporting is not merely a technical challenge but also an organizational, legal, and economic one, requiring coordinated efforts across industry stakeholders and regulators.

Table 29. Barriers to adoption of reporting techniques for compliance and reporting activities

	Time-consuming	Errors	Human capital & digital skills	Scalability & tailoring	Financial costs
Standard techniques	High	High	Low	Medium	Low
Advanced techniques	Medium	Low	High	Medium	Medium
Automated techniques	Low	Low	High	Low	High

Source: The data used to create this table were collected through a survey of industry associations.

3.6.4.3 Qualitative cost-benefit analysis

Overview of costs

Implementing each technique entails different categories of costs. For analytical purposes, these were classified into four groups: (i) capital and operational costs, such as IT infrastructure setup and license acquisition; (ii) labour costs; (iii) training costs; and (iv) technical costs, including maintenance and support services.

For **standard, spreadsheet-based techniques**, the primary cost drivers are labour and technical expenses. Stakeholders emphasized that manually entering data into Excel files and converting them into other formats is a time-intensive process requiring substantial workforce resources. Additionally, maintaining and updating spreadsheets incurs significant costs, as templates must be regularly revised to reflect regulatory changes and new reporting requirements. By contrast, capital and operating costs remain relatively low, and the process demands minimal training. Excel templates are generally designed to be user-friendly, requiring only the completion of predefined fields and the transfer of data into the required format.

For **advanced techniques**, the most significant cost component relates to workforce training. This is particularly critical for SMEs, which often lack extensive experience with IUCLID. They may also dispose of limited resources to digitalise workflows, including the development of centralised databases. Although capital and operating costs are higher than for standard techniques, they remain reasonable. ECHA-developed software is available free of charge. Companies typically only need to acquire ERP or EHS licenses. Labour and technical costs are comparatively lower, as several tasks are automated under advanced approaches.

Automated techniques require substantial investments, which are key drivers of capital and operational costs. Companies need to develop comprehensive databases (if not already in place) and standardize workflows and reporting processes. Training costs can also be significant, particularly for SMEs that may lack specialized personnel. Maintenance and support services, however, are expected to entail only moderate expenses. Labour costs, on the other hand, are expected to be much lower.

Expected benefits

The benefits of adopting advanced and automated reporting techniques were also grouped into four categories: (i) improved accuracy, ensuring consistent and precise results; (ii) enhanced efficiency, with a more streamlined and less time-consuming reporting process; (iii) cost savings; and (iv) real-time monitoring capabilities. These expected benefits were evaluated relative to standard techniques.

The primary benefit of adopting **advanced techniques** is enhanced accuracy, as data is extracted directly from other systems, reducing opportunities for manual manipulation and error. Another important advantage is the reduction in costs, driven by reduced manual input and partial automation of repetitive tasks. Efficiency gains are also notable, although they may be partially offset by the need to format compliance reports and ensure interoperability across systems. Additionally, many ERP systems offer integrated functionalities for real-time monitoring, providing a further advantage over standard techniques.

Automated techniques offer clear advantages over standard approaches, particularly in terms of efficiency and time savings. Stakeholders noted that reporting activities add limited value to companies and the industry overall, providing only occasional insights into business performance or societal impact. AI-powered tools, for instance, were singled out as potentially being used to summarise reports and extract relevant information from SDSs, a task that is particularly time-consuming. Blockchain-based solutions are seen as

a promising approach to improving transparency and streamlining communication across the supply chain.

Automated solutions also improve accuracy and enable real-time monitoring. Furthermore, they have the potential to deliver significant cost reductions, although the extent of these savings is difficult to quantify. For example, most companies currently employ at least one full-time staff member dedicated to managing SDSs – a largely manual task that could be automated using AI-powered tools capable of extracting and entering data from structured documents. However, the total labour savings remain uncertain, as additional verification work may still be required. Incomplete or missing data in SDSs pose a compliance risk, exposing companies to substantial fines from enforcement authorities.

Cost-benefit assessment

Table 30 presents a comparative assessment of the costs and benefits of adopting advanced and automated techniques relative to standard techniques.

Adopting advanced techniques would require significantly higher investments in IT infrastructure, including maintenance and support. Workforce training would also represent a substantial cost. However, these investments are expected to deliver notable improvements in accuracy and efficiency. By automating manual and repetitive tasks, organizations can optimize resource allocation, potentially achieving overall cost savings despite the higher initial outlay.

Table 30. Qualitative assessment of the costs and benefits associated with different reporting techniques

Type of reporting technique	Costs				Relative Benefits			
	Capital & operating	Labour	Training	Technical	Accuracy	Efficiency	Cost savings	Real-time monitoring
Standard techniques	Negligible	Significant	Negligible	Significant				
Advanced techniques	Reasonable	Reasonable	Significant	Reasonable	Significant	Reasonable	Reasonable	Reasonable
Automated techniques	Significant	Negligible	Significant	Reasonable	Significant	Significant	Reasonable	Significant

Source: The data used to create this table were collected through interviews of industry associations.

Similarly, adopting automated reporting methods would involve substantial expenses for IT infrastructure and workforce training. However, the main benefits lie in significantly improved accuracy and efficiency.

That said, the magnitude of costs and benefits varies by company size. Larger firms may have already invested in digitalizing workflows and building comprehensive databases, and their staff are often more experienced with advanced and automated techniques. This reduces the relative cost of automation compared to SMEs. However, larger companies may need to automate a greater number of processes, which can increase overall implementation costs. In terms of benefits, automation is most economically advantageous for companies managing extensive portfolios of substances and products. Consequently, automation is more likely to be profitable for larger firms.

3.6.5 Policy recommendations

While the industry views automation as an opportunity for future growth and streamlined reporting, there are several ways in which policymakers can facilitate the development and adoption of advanced and automated tools.

Harmonisation of reporting requirements and regulatory acceptance

To enable companies to fully leverage the benefits of automation, reporting requirements should be harmonized across EU Member States. Since automated techniques require substantial investments, they can only become cost-effective if applied to a broad range of reporting obligations and integrated with existing reporting platforms. This calls for harmonized reporting standards and the adoption of digital formats as the norm.

Equally important, companies must have confidence that outputs generated by automated systems will be accepted by the relevant authorities. Regulatory acceptance was consistently identified as a critical driver for the adoption of automation.

Operational Actions:

- **Conduct an EU-wide audit of reporting requirements across Member States** to identify inconsistencies, redundancies, and opportunities for harmonization. This process should involve a comprehensive mapping of national reporting obligations – both in terms of content and format – against EU-level requirements. The findings can then inform the development of standardized digital formats and interoperable data models to support streamlined, cross-border compliance.
- **Establish a regulatory framework that supports the use of automated tools** in compliance and reporting processes. This should involve close collaboration with ECHA and national authorities to develop formal guidelines and validation protocols for accepting outputs generated by automated solutions. To build trust and encourage adoption, pilot certification or pre-validation schemes could be introduced, accompanied by clear, transparent acceptance criteria that are consistently applied and enforced.

Investments in the digital infrastructure and digital literacy in the industry

Industry representatives acknowledged ECHA's investments in developing an IT infrastructure that facilitates compliance and reporting under REACH and CLP. The IUCLID software is already highly advanced, enabling automatic report generation once the relevant data is entered into the system. It further allows for automated checks and validation of the dossiers, leaving data collection as the primary task for businesses. Building on this progress, the industry would welcome similar tools to support compliance and reporting under other legislative frameworks.

However, many of the existing tools remain inaccessible to SMEs due to their complexity. Representatives from downstream user industries emphasized the need for additional training and capacity-building initiatives to help companies enjoy the benefits of these tools.

Industry representatives have also expressed interest in AI-powered solutions developed internally by the European Commission and made accessible to authorized industry users. Such an approach would enable companies – including SMEs that lack the resources to build their own infrastructure – to access state-of-the-art technologies while ensuring the security of their data.

Operational Actions:

- **Support the development of IUCLID-like digital tools** for additional legislative frameworks to streamline regulatory compliance and reporting. This effort should involve close collaboration between ECHA and the European Commission to ensure that new tools are interoperable, modular, and aligned with existing digital infrastructure. Incorporating feedback from industry stakeholders – particularly SMEs – is essential to ensure that these tools are user-friendly and do not impose significant adoption or operational costs.
- **Develop targeted digital literacy and training programs for SMEs and downstream users** to help bridge the digital divide and enable smaller companies

to effectively use advanced compliance tools. These programs should focus on building the necessary skills to navigate and benefit from existing digital solutions. Partnering with industry associations, chambers of commerce, and academic institutions can enhance the reach of these programs, while also ensuring the delivery of multilingual, modular, and tailored training content.

- **Promote the development of secure, AI-powered compliance solutions that are accessible to all industry stakeholders.** This could take the form of a centralized, EU-funded, AI-driven platform available exclusively to registered users located or conducting economic activities in the EU. To ensure trust and effectiveness, the initiative should be supported by a robust governance framework that guarantees data security and confidentiality. At the same time, the AI system should be trained on EU legislation, standards, and reporting requirements to ensure full regulatory alignment and relevance.

Reassessing the content of reporting

Automation can help reduce reporting and compliance costs by using robotic tools to perform repetitive, low value-added tasks. However, industry representatives consider that a full reliance on automated solutions is unlikely. Instead, they argue that a reassessment of reporting obligations would have a greater impact on reducing administrative burdens. Reporting should be guided by the principle that information serves to inform policymaking and research, rather than simply to populate comprehensive databases.

Operational actions:

- **Conduct an internal audit** to identify reporting requirements that do not directly contribute to policymaking, risk assessment, or scientific research. As part of this process, map each data element to its specific regulatory purpose, ensuring that reporting is driven by the utility of the data rather than the goal of compiling exhaustive databases. This reassessment should be informed by broad stakeholder consultations, engaging policymakers, industry representatives, and academic experts.
- **Conduct a comprehensive review of existing reporting obligations to identify overlaps or similarities** that could allow for consolidation. Streamlining these requirements would help reduce administrative burden by eliminating duplicated efforts. Developing harmonized reporting templates or enabling shared data submissions across regulatory frameworks would further reduce the costs of compliance and reporting in the chemicals sector.

3.7 Energy

3.7.1 Overview of the sector

This section provides an overview of the EU energy sector, starting with the sector's core activities and concluding with an analysis of key economic indicators.

3.7.1.1 Definition

The energy sector encompasses a **broad set of activities related to the production, transformation, distribution and trade of energy products**, including electricity, gas, steam, and air conditioning. It falls under Section D of the NACE Rev.2 classification (Electricity, gas, steam and air conditioning supply). The section is further divided into the following groups:

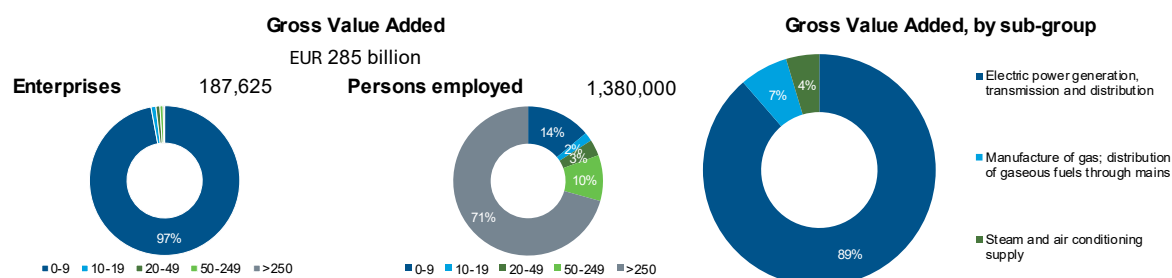
- 35.1 – Electric power generation, transmission and distribution;

- 35.2 – Manufacture of gas, distribution of gaseous fuels through mains;
- 35.3 – Steam and air conditioning supply.⁵⁴²

3.7.1.2 Size, value, employment

In 2022, the energy sector generated approximately **EUR 285 billion in gross value added** (GVA), representing nearly 3% of the EU's total GVA (Figure 30). Around **190,000 enterprises operated in the sector**, the vast majority of which employed fewer than 10 people. In total, the sector **directly employed 1.4 million individuals**, with two-thirds working in large enterprises of at least 250 employees. Nevertheless, small firms (fewer than 10 employees) and medium-sized firms (50–249 employees) accounted for a significant share of employment.

Figure 30. The EU Energy Sector – Key Facts



Sources: The data on gross value added, the number of enterprises, and employment are sourced from Eurostat's *Structural Business Statistics dataset* (online code: *sbs_sc_ovw*), specifically for the NACE Rev. 2 two-digit sector D35 (Electricity, gas, steam and air conditioning supply).

When broken down by group, electric power generation, transmission, and distribution accounted for 89% of the industry's GVA, while the manufacture and distribution of gas and steam and air conditioning supply represented significantly smaller shares, at 7% and 4%, respectively.

3.7.2 Regulatory landscape, reporting and compliance activities

This section presents an overview of the regulatory landscape in the energy sector by providing a detailed legal analysis of five legislative acts, their related compliance and reporting activities, as well as the underlying challenges companies encounter in fulfilling their regulatory requirements.

3.7.2.1 Legal and regulatory framework

This overview gathers the main reporting and compliance duties that apply to energy companies operating in the EU. It describes, in broad strokes, what businesses must file, monitor and disclose under the Energy Efficiency Directive (EED), the Regulation on Wholesale Energy Market Integrity and Transparency (REMIT), the EU Emissions Trading System Directive (EU ETS), the European Pollutant Release and Transfer Regulation (E-PRTR), and the Regulation on Submission and Publication of Data in Electricity Markets. The full legal review is available in Annex 3.5

*Energy Efficiency Directive (EED) (Consolidated Directive 2012/27/EU)*⁵⁴³

⁵⁴² We exclude the following extractive and manufacturing industries related to the energy sector: mining of coal and lignite (B05), extraction of crude petroleum and natural gas (B06), extraction of peat (B0892), support activities for petroleum and natural gas extraction (B091), and manufacture of coke and refined petroleum products (C19).

The Energy Efficiency Directive aims at lowering energy consumption in the EU by establishing a legislative framework that promotes energy efficiency. Under the EED, **obligated parties** (i.e., energy distributors or retail energy sales companies) **must enrol in national energy-efficiency schemes**, as designated by the Member States in an objective and non-discriminatory manner (Article 7a). Member states must put in place an independent review of energy savings by verifying a representative share of the energy efficiency improvement measures. Member States must publish the results each year.

Large companies must also have an **independent energy audit** at least every four years (Article 8). These audits should include a detailed analysis of energy consumption and the identification of areas where savings are possible (as described in Annex VI). Audits may be conducted by in-house experts or external energy auditors, provided the Member State has systems to ensure audit quality. Firms that operate a certified energy or environmental management system may skip the stand-alone audit if the system already includes one that meets the Directive's minimum criteria (Art. 8(6)). The EED also highlights the special role SMEs play in this system. Member States are required to **develop programmes to encourage SMEs to undergo energy audits**, as well as helping them implement recommendations from these audits. This may happen through national support schemes (Article 8(2)).

Regulation on Wholesale Energy Market Integrity and Transparency (REMIT) (Regulation (EU) No 1227/2011)⁵⁴³

The Regulation on Wholesale Energy Market Integrity and Transparency (REMIT) provides the legal framework for monitoring wholesale energy markets across the EU. Its purpose is to enhance transparency, prevent market abuse, and foster competition and trust among market participants.

REMIT establishes a series of obligations for market participants. Under Article 4, they are expected to **disclose inside information** in an effective and timely manner through Inside Information Platforms (IIPs). Article 8 of the regulation further calls on market participants to **share detailed transaction and facility information** with the Agency for the Cooperation of Energy Regulators (ACER), mostly via third parties called registered reporting mechanisms (RRMs). Article 9 stipulates that market participants engaging in transactions that need to be reported to the Agency under Article 8 have to register with the national regulatory authority.

EU Emissions Trading System Directive (EU ETS) (Directive 2003/87/EC)

The EU Emissions Trading System (EU ETS) is designed to reduce greenhouse gas (GHG) emissions by making polluters pay for their emissions. Based on a “cap and trade” principle, it sets emission limits for installations and operators, with the caps being annually reduced to meet the Union's climate target. Companies receive a mix of free and auctioned allowances, which can then be traded on the market. These operations are recorded in the Union Registry.

⁵⁴³ A revised Energy Efficiency Directive was adopted on 13 September 2023, with Directive 2012/27/EU being repealed with effect from 12 October 2025 (Art. 38 of Directive (EU) 2023/1791 on energy efficiency). The former legislative act was analysed as relevant data was obtained under the previous Energy Efficiency Directive.

⁵⁴⁴ REMIT was amended in May 2024 through Regulation (EU) 2024/1106 and Regulation (EU) 2024/1789. These revisions strengthened existing transparency and market integrity requirements, aligning them more closely with EU financial market legislation. The scope of REMIT was expanded to include additional wholesale energy products and participants. The amendments also enhanced the powers of ACER and national regulatory authorities.

The system currently covers sectors such as energy, manufacturing, and aviation in all EU Member States, the EFTA countries (Iceland, Liechtenstein, and Norway) and Northern Ireland for electricity generation.⁵⁴⁵ Overall, the EU ETS applies to the GHG emissions from almost 10,000 installations. In the energy sector, the EU ETS specifically targets carbon dioxide emissions from electricity and heat production.

The ETS Directive provides the legal basis for the EU Emissions Trading System and establishes several compliance and reporting obligations for economic operators. Under Article 7, operators must **inform the competent authority** if they plan any changes to the nature or functioning of an installation that may necessitate an updating of the GHG emissions permit. Under article 12(3), Member States must ensure that the operator of each installation **surrenders a number of allowances** that match the previous year's verified emissions. Member States are also responsible for ensuring that **each operator of an installation annually reports their emissions** to the competent authority in line with the monitoring guidelines and as verified by an accredited verifier (Article 14).

*European Pollutant Release and Transfer Regulation 2006 (E-PRTR) (Regulation (EC) No 166/2006)*⁵⁴⁶

Under the European Pollutant Release and Transfer Regulation (E-PRTR), large industrial facilities are required to report significant pollutant emissions and resource usage. The goal of the regulation is to increase transparency in data reporting and enable better identification and monitoring of sources of industrial pollution.

Article 5 of the E-PRTR requires operators of large energy facilities to **file yearly pollutant-release reports**. Under Article 6, they also have to **declare certain on-site waste disposals** as releases to land, while Article 9 establishes the responsibility of facility operators to **comply with reporting requirements** set in Article 5 and assure the quality of the reported data.

The E-PRTR has recently been revised to broaden the scope of industrial reporting. The updated regulation will require the Industrial Emissions Portal to publish detailed information on energy, water, and raw material consumption, along with information on operators' activities. Reporting requirements will be aligned with the revised Industrial and Livestock Rearing Emissions Directive (IED 2.0) and allow for adjustments to the list of pollutants. Quantification methods will also be harmonised, with the goal of enhancing data quality. The European Commission is responsible for developing the implementing rules, including standardised reporting formats for resource use and newly covered sectors.

Regulation on Submission and Publication of Data in Electricity Markets (Commission Regulation (EU) No 543/2013)

The Regulation on submission and publication of data in electricity markets obliges Transmission System Operators (TSOs) to calculate and transmit a broad set of grid and market indicators to the European Network of Transmission System Operators for Electricity (ENTSO-E). The objective of the regulation is to enhance

⁵⁴⁵ Following the 2023 revision of the ETS directive, a new system, names ETS2, was established covering emissions from fuel combustion in buildings, road transport, and other sectors (predominantly small sectors not covered by the current ETS system).

⁵⁴⁶ The European Parliament and the Council have recently adopted a new regulation on reporting of environmental data from industrial installations and establishing an Industrial Emissions Portal (Regulation (EU) 2024/1244). This new regulation repeals Regulation (EC) No 166/2006 as of 1 January 2028. Under Article 21 of Regulation (EU) 2024/1244, the reporting provisions under Regulation (EC) No 166/2006 will continue to apply for the year 2026. For this reason, the analysis focuses on the reporting obligations under Regulation (EC) No 166/2006.

transparency in the electricity market by ensuring that participants can easily access data on electricity generation, transportation, and consumption.

Specifically, the indicators cover total load (Article 6), the unavailability of consumption units (Article 7), year-ahead forecast margins (Article 8), infrastructure plans (Article 9), the unavailability of transmission infrastructure (Article 10), cross zonal capacities (Article 11), the use of cross zonal capacities (Article 12), congestion management measures (Article 13), forecast generation (Article 14), unavailability of generation and production units (Article 15), actual generation (Article 16), and balancing (Article 17). The data are to be reported within tight publication deadlines.

The regulation mandates the **establishment of a central information transparency platform** operated by ENTSO-E. According to Article 3, this platform must be freely accessible to the public via the internet and available in at least the English language. The data provided should be up-to-date, easily accessible, downloadable, and retained for a minimum of five years. All updates must be time-stamped, archived, and made publicly available.

3.7.2.2 Reporting and compliance activities

Energy Efficiency Directive (EED) (Consolidated Directive 2012/27/EU)

The EED requires that enterprises carry out an **independent energy audit every four years**. While SMEs are excluded from the scope of the energy audit obligations, the Directive states that Member States should encourage SMEs to nevertheless undergo energy audits through programmes and that they are allowed to set up support schemes for this purpose

While the Directive does not prescribe a common EU reporting template, **Annex VI establishes minimum criteria for energy audits**. A survey of supervisory authorities responsible for enforcing Article 8 (Article 11 in the revised regulation) from 16 Member States shows that, **in most countries, energy audits are submitted to the relevant authorities via digital platforms**. Some countries (e.g., Denmark, Malta) use Excel templates, while in Luxembourg, accredited auditors submit the data to the national authority.⁵⁴⁷ In nine countries, the submission of the full energy audit report is compulsory, while in the remaining seven member states reporting to the national authority is optional. In this case, enforcement is achieved through random checks.⁵⁴⁸

The audit must be **conducted by a qualified or accredited expert or implemented and supervised by an independent authority under national law**. According to the requirements specified in Annex VI, the energy audit must:

- be grounded in up-to-date, measured, and traceable operational data on energy consumption and (for electricity) load profiles;
- contained detailed review of the energy consumption profile of buildings or groups of buildings, industrial operations or installations (including transportation);
- include life cycle cost analysis whenever possible instead of simple payback periods.

The audit must also be **proportionate** and **sufficiently representative**, so that it gives a reliable picture of overall energy performance and of the most significant opportunities for

⁵⁴⁷ Johnsson, Simon, Mariana Andrei, and Maria Johansson. "Harmonizing energy audit reporting: Addressing data loss and policy challenges in the EU member states." *Energy* 319 (2025): 135040.

⁵⁴⁸ Ibid.

improvement.⁵⁴⁹ Auditors must provide validated calculations for proposed measures and clear information on potential savings.

The EED also requires that the **data used in the energy audits are storable** to facilitate historical analysis and monitor performance.

Regulation on Wholesale Energy Market Integrity and Transparency (REMIT) (Regulation (EU) No 1227/2011)

REMIT sets core duties for market conduct and transparency. Article 4(1) requires **timely disclosure of inside information** on the capacity and use of facilities for production, storage, consumption, or transmission of electricity, hydrogen, or natural gas or related to the capacity and use of LNG facilities.

The inside information and corresponding reports are **disclosed via IIPs**. IIPs have to ensure that the information can be easily accessed by the public either through a website or a clear application programming interface. As of 7 July 2025, 14 platforms have been successfully assessed, while a second assessment phase is ongoing for another platform.⁵⁵⁰

Under Article 4(4) of REMIT, publishing inside information on approved platforms also counts as effective disclosure, although it may not always qualify as timely and public in the sense of Article 4(1). For example, ENTSO-E and ENTSG have developed central transparency platforms that meet the requirements of Article 4(4). ENTSG, GIE, and other European Transmission System Operator platforms (TSO) have also applied to be considered IIPs and have set the objective of meeting the IIP requirements set under Article 4 REMIT.

To share data with IIPs, energy companies can use browser-based reporting tools where they manually enter the data by means of a web portal. Alternatively, market participants can **share data directly via web service API**.⁵⁵¹

Article 5 prohibits the manipulation (including attempts) of wholesale markets. Supervision now extends to trading by or via algorithms. Under the new Article 5a, a market participant that uses algorithms must report this to ACER and must maintain effective systems and risk controls to prevent disruptions and incorrect trading orders. Market participants are required to keep information on trading strategies, trading parameters or limits, as well as compliance and risk controls for at least five years, allowing competent authorities to retrieve these data when needed.

Under Article 8, market participants must **provide wholesale transaction information to ACER**. The information reported shall include the precise identification of the wholesale energy products bought and sold, the price and quantity agreed, the dates and times of execution, the parties to the transaction and the beneficiaries of the transaction as well as any other relevant information.

⁵⁴⁹ Several Member States have included in their national legislation a specific de minimis energy consumption that needs to be covered by the audits. See Ricardo Energy & Environment (2018), *Development of recommendations on the implementation of certain aspects of Article 8 and Annex VI of the Energy Efficiency Directive*, Final Report for DG ENERGY ENER/C3/2013-484/13/FV2017-427/01. https://energy.ec.europa.eu/publications/development-recommendations-implementation-certain-aspects-article-8-and-annex-vi-energy-efficiency_en.

⁵⁵⁰ The list of IIP platforms is available at <https://www.acer-remit.eu/portal/list-inside-platforms>.

⁵⁵¹ For example, the EEX Transparency Platform allows for effective and timely disclosure of information either via their browser-based Inside Information Messenger or directly through web service API. For more information, visit <https://www.eex-group.com/en/services/transparency-services>.

Before engaging in wholesale energy trading, companies must **register with the national regulator** of the Member State where they are established or reside, through the Centralised European Register of Energy Market Participants. If they are not based in the EU, market participants need to register in a Member State where they are active. Market participants provide company details, ownership structure, contact persons, and activities.

The **reporting of trades and orders is done via Registered Reporting Mechanisms (RRMs)**, which are ACER-accredited third parties. At the moment, the REMIT portal reports more than 100 RRM⁵⁵². Data has to be **submitted in a structured format**, as detailed in the Manual of Procedures (MoP) and Transaction Reporting User Manual (TRUM).⁵⁵³ The reporting format is usually XML based on ACER's schema, with the data uploaded via secure electronic channels. Templates and reporting guidance are made available by ACER, helping economic operators align their internal systems with technical rules and maintain consistent submissions over time.

EU Emissions Trading System Directive, consolidated Directive 2003/87/EC

The EU ETS covers industrial installations and aircraft operators. The EU ETS compliance cycle includes monitoring, reporting, and verification (MRV) of GHG emissions.

Before the start of a monitoring period, industrial installations must have their **monitoring plans** approved by the relevant national authority. Throughout the calendar year, they are required to **track their emissions** and **compile an annual report**. This report is first **verified by an accredited verifier** and then **submitted to the competent authority** by 31 March of the following year. By 30 September, operators must **surrender the corresponding number of allowances** to account for their verified emissions. The European Commission provides templates for monitoring plans, annual emission reports, verification reports, and improvement reports.

The European Commission's **ETS Reporting Tool** is a free, web-based platform that supports MRV obligations for both the ETS and ETS2. It is regularly updated as Member States propose improvements, helping to keep procedures aligned with evolving practice. The ERT allows ETS operators to:

- **submit monitoring plans**, which can then be reviewed and approved by the competent authorities;
- **submit annual emission reports** (including data on GHG emissions, fuel consumption, and production levels);
- **have their reports reviewed and verified by accredited verifiers**;
- **manage** their account, **obtain** compliance information, or **conduct** other performance-related operations.

Before reporting data, economic operators have to **register with the online ETS Reporting Tool** by following the next steps:⁵⁵⁴

1. Access the EU ETS reporting tool via the official portal: <https://ets-reporting.ec.europa.eu/>. Access is obtained through EU Login with two-factor authentication (e.g., email, password, and a registered second authentication method).

⁵⁵² The list of RRM^s is available at <https://remitportal.acer.europa.eu/portal/list-of-rrm>.

⁵⁵³ The MoP and TRUM are available at <https://www.acer.europa.eu/remit-documents/remit-reporting-guidance>.

⁵⁵⁴ The ERT user manual is available at https://climate.ec.europa.eu/eu-action/carbon-markets/eu-emissions-trading-system-eu-ets/monitoring-reporting-and-verification/ets-reporting-tool-ert_en/.

2. If a user is not already appointed to an organisation, they should contact the local Competent Authority or the service desk. First time users are also expected to register by completing the following two steps.
3. Non-registered users are displayed a “You are not registered in the system” message. To register, they need to agree with the “Terms and conditions” and the “Privacy statement” by clicking on the corresponding checkboxes.
4. Click Register

Once a person has access to their organisation on EU ETS Reporting, they can **submit a monitoring plan** to the EPA for approval and **submit their Annual Emission Report (AER)**. For the latter, the following steps must be followed:

1. Click on “Create A New Report”, available in the Annual Emission Report (AER) area.
2. If a monitoring plan has previously been approved, a pop-up window will appear, displaying the “Monitoring Plan” (MP) section. Users can then reset existing data in the AER and import it from the MP.
3. Users fill in the mandatory fields of the form. Different sections of the webform can be accessed, provided that they are enabled.
4. The AER is saved in status DRAFT.
5. Operators can share an AER saved in status DRAFT with their Competent Authority by ticking the “Share report with CA” checkbox. In this case, the Competent Authority will be able to access the details page of the report. Operators also have the option to edit an AER as well as add comments or attachments. To delete an AER, operators must click on the “Delete” button.
6. If a verifier has already been assigned to an organisation, the operator can initiate the verification process by clicking “Request verification.” A confirmation message will appear, requiring the user to click the button a second time. Once submitted, the AER status changes to “Under verification.”
7. From there, two outcomes are possible:
 - a. If revisions are needed, the status changes to “Under modification requested by verifier,” prompting the operator to update the AER.
 - b. If the report is accepted, the status becomes “Verified,” allowing the operator to submit it to the Competent Authority.
8. The Competent Authority may then either request further modifications or validate the report. In the latter case, the status will update to either “Estimated conservatively” or “Reported data finalised.”
9. To revise a report, the operator must click the “Revise” button, which resets the AER status to “DRAFT” and restarts the process. Note that the “Revise” option is only available when the status is either “Under modification requested by verifier” or “Under modification requested by CA.”

European Pollutant Release and Transfer Regulation 2006 (E-PRTR) (Regulation (EC) No 166/2006)

Under E-PRTR, facilities engaged in any of the 65 economic activities listed in Annex I and exceeding specific capacity and emission thresholds must **annually report the amounts**. Reporting is based on data from the previous calendar year. The E-PRTR currently covers 91 pollutants, grouped into seven categories: greenhouse gases, other gases, heavy metals, pesticides, chlorinated organic substances, other organic substances, and inorganic substances.

Facility operators **submit their data through national reporting systems**, which serve as the initial validation layer. The Member States establish, via national legislation, the deadlines for submitting the data. Additionally, if the competent authority does not already possess the necessary facility identification details, operators must **provide this information electronically** using the format specified in Article 7(2).

In addition to reporting the quantities of pollutant releases, operators must **indicate whether the data is derived from measurement, calculation, or estimation**. If measurement or calculation is used, they must also **specify the method applied**. When compiling their reports, operators are expected to use the most reliable information available. This may include direct monitoring, emission factors, mass balance approaches, indirect measurements, engineering assessments, or other scientifically sound techniques, in line with Article 9(1). Furthermore, operators are required to **keep supporting documentation for five years** following the reporting year. These records must detail both the reported figures and the methodologies used and must be accessible to national authorities upon request.

National authorities validate, compile, and forward the dataset to the European Commission, which publishes the results on the Industrial Emissions Portal (with the assistance of the European Environment Agency). This allows authorities, companies, and the public to track performance across facilities and over time.

Regulation on submission and publication of data in electricity markets (Commission Regulation (EU) No 543/2013)

According to Article 4, **primary data owners are required to submit data to TSOs** that is complete, meets the established quality standards, and is communicated in a manner that enables the TSO to process and transmit it to ENTSO-E in a timely fashion. Alternatively, **with prior agreement from the relevant TSO, primary data owners may submit the data directly to the central information transparency platform, provided they do so through a third party acting as a data provider on their behalf**.

The ENTSO-E Transparency Platform acts as the central hub for collection and publication of this information, providing free and continuous access to the European electricity market data in near real time. According to Article 5, ENTSO-E has the responsibility of developing a manual that specifies the details and format of the submission of data, standardised ways and formats of data communication and exchange, technical and operational criteria, and an appropriate classification of production types.

3.7.2.3 Overview of existing challenges

Evidence from previous studies, stakeholder consultations and surveys shows that economic operators in the energy sector face **recurring friction where legal obligations meet day-to-day data work**, especially when templates, identifiers and timelines differ across regimes.⁵⁵⁵ The picture that emerges is consistent across the EED, REMIT, the EU ETS, the E-PRTR, and the electricity transparency regulation. Capacity constraints, fragmented portals, and evolving technical specifications are the source of pain points for most operators.

Under the **Energy Efficiency Directive**, the Commission's Joint Research Centre surveyed Member State experts during the first Article 8 compliance cycle and recorded practical obstacles that translated into burdens for large enterprises.⁵⁵⁶ Economic

⁵⁵⁵ European Commission. (2023). *Annual Burden Survey 2022: The European Union's action to simplify legislation*. https://commission.europa.eu/system/files/2023-09/ABS_20230912_0.pdf.

⁵⁵⁶ Serrenho, T., Analysis on the practices to collect, store and assess information arising from Energy Audits in the EU-28 - EED Article 8 provisions on Energy Audits' information management,

operators reported fragmented reporting channels and formats, which increases effort and error risk. **The report shows a patchwork of web portals, functional mailboxes and templates, with different procedures across, and sometimes within, countries.** Enterprises are also often unsure whether they qualify for reporting, especially in complex group company setups or when they meet some, but not all, of the SME criteria. Cross-border and multi-jurisdiction operations multiply the above issues. The report flags problems with international companies whose parent is located in another country, creating uncertainty over which entity is responsible for filing and in which jurisdiction. Similarly, another study from 2025,⁵⁵⁷ supports these findings. It highlights significant variations in energy audit data collection practices across EU Member States. These inconsistencies often lead to information loss and undermine the reliability of energy data reporting.

Another challenge stems from the **significant financial burdens associated with compliance. Energy audits can be quite costly**, especially since they have to be conducted every four years. Moreover, **non-compliance can lead to consequential fines and penalties** – up to EUR 25 million for REMIT violations and EUR 50 million for breaches under the EU ETS. In 2024 alone, regulatory authorities across the EU imposed approximately EUR 113 million in fines for REMIT non-compliance, marking the highest annual total to date. This figure represents more than half of all penalties issued under REMIT since enforcement actions began in 2015.⁵⁵⁸ These costs and risks create a **disproportionate strain on smaller firms**, which often lack the resources to invest in robust compliance infrastructure.

ACER carried out a public consultation in 2024 on reporting obligations under REMIT.⁵⁵⁹ **Participants highlight that the reporting scale and technical detail have grown while the disclosure toolkit has diversified.** ACER's evaluation of responses to its 2024 public consultation on the revision of the Implementing Regulation's Annex shows high engagement from industry and focuses attention on data-field definitions, uniformity and proportionality, with 92 submissions feeding into the redesign of transaction and fundamental data reporting. In parallel, position papers from the electricity industry press for simple, proportionate schema choices and oppose proposals that would in practice require revealing trading strategies. Companies view these as operationally impractical and lacking a legal basis.⁵⁶⁰ Together, these findings portray a sector that is committed to compliance but highly sensitive to complexity, excessive reporting obligations, and overlapping communication channels.

Similarly, **E-PRTR operators share concerns about disclosing commercially sensitive and confidential data, duplication with existing reporting obligations, and**

monitoring and verification, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-08766-3, doi:10.2760/82756, JRC109382.

⁵⁵⁷ Johnsson, Simon, Mariana Andrei, and Maria Johansson. "Harmonizing energy audit reporting: Addressing data loss and policy challenges in the EU member states." *Energy* 319 (2025): 135040.

⁵⁵⁸ Agency for the Cooperation of Energy Regulators (ACER). Enforcement decisions. <https://www.acer.europa.eu/remit/coordination-on-cases/enforcement-decisions>.

⁵⁵⁹ European Union Agency for the Cooperation of Energy Regulators, *Evaluation of responses to the public consultation of ACER on the revision of the Annex of the REMIT Implementing Regulation*, 2024, https://www.acer.europa.eu/sites/default/files/documents/Official_documents/Public_consultations/PC_2024_R_01/PC_2024_R_01_Evaluation_of_Responses.pdf.

⁵⁶⁰ Eurelectric, ACER consultation on the revision of the REMIT implementing regulation - A Eurelectric response paper, 2024, <https://www.eurelectric.org/wp-content/uploads/2024/09/M-ACER-consultation-on-the-revision-of-the-REMIT-Implementing-Regulation.pdf>.

excessive administrative and financial burden. This could lower the competitiveness of European industries.⁵⁶¹ These concerns are echoed across other regimes, where the lack of harmonised templates and identifiers, combined with evolving technical specifications, creates uncertainty and increases the risk of error. The cumulative effect is a compliance landscape that demands constant adaptation, often without adequate support or clarity, with a disproportionate burden being placed on smaller businesses.

The most detailed user-side assessment under Regulation 543/2013 comes from an external review of the ENTSO-E Transparency Platform that combined statistical checks with an online user survey and expert interviews.⁵⁶² **The review pointed to issues regarding completeness, accuracy, timeliness and user friendliness.** Despite this being part of a Regulation, there were inaccurate and inconsistent definitions used between national sources and the central platform. The authors recommend stronger, public data-quality reporting by data providers, clearer machine-readable metadata and a more user-centric governance model.

In conclusion, **businesses face a common set of reporting and compliance hurdles across the analysed instruments.** Fragmented portals, divergent templates and identifiers, and mismatched timelines duplicate work effort and raise the risk of error, concerns that are significantly amplified for cross-border groups that must reconcile different national and EU rules. Technical burdens keep growing as schemas evolve, real time feeds expand, and guidance shifts, which strain budgets and affects smaller businesses the hardest. Unclear scoping for energy audits, verification bottlenecks, and recurring audit costs add pressure, while confidentiality concerns persist where datasets can expose commercial strategies without clear public value. The ENTSO-E transparency experience highlights persistent issues with data completeness, accuracy, timeliness and usability, as well as inconsistent definitions between national sources and the central platform – even under a single legal instrument. Multinational companies operating across Member States must navigate this complexity while staying informed about national-level deadlines and obligations, and monitoring updates to governance regulations and implementing acts. They are expected to align internal energy strategies with national targets and reporting frameworks, identify opportunities to optimise energy use and reduce costs, and maintain full compliance. These layered demands underscore the need for stronger confidentiality safeguards, proportionate requirements for smaller participants, and targeted capacity support for auditors, verifiers and operator compliance teams to help firms meet obligations without unnecessary cost.

3.7.3 Overview of reporting and compliance techniques

3.7.3.1 Description of techniques and tools

Standard techniques

Standard compliance and reporting techniques in the energy sector have traditionally relied on manual methods and procedural approaches, forming the foundation for regulatory adherence. These legacy methods include **maintaining spreadsheets and static databases to monitor emissions, energy usage, and operational metrics.** Reporting typically occurs **via email workflows or even paper submissions.**

⁵⁶¹ IOGP Europe, IOGP position on Industrial Emissions Portal (IEP) Regulation proposal, 2022, <https://iogpeurope.org/wp-content/uploads/2022/07/Industrial-Emissions-Portal-paper.pdf>; Cefic, Cefic views on the revision of the E-PRTR, 2022, <https://cefic.org/app/uploads/2024/05/Cefic-views-on-the-revision-of-the-E-PRTR.pdf>.

⁵⁶² VVA, Copenhagen Economics, Neon & Deloitte, A review of the ENTSO-E Transparency Platform Output 1 of the “Study on the quality of electricity market data” commissioned by the European Commission, 2017, <https://neon.energy/transparency-platform-report.pdf>.

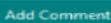
Regular internal audits and inspections, guided by standardised checklists derived from regulatory frameworks, offer a key internal control mechanism. These checklists are often crafted by institutions like the European Commission, national energy regulators, or industry associations.

An example is the EU Emissions Trading System (EU ETS).⁵⁶³ The European Commission supports compliance via the Monitoring and Reporting Regulation (MRR) and the Accreditation and Verification Regulation (AVR) by providing official templates and guidance documents. These include:

- Monitoring plan templates outlining how operators will track emissions,
- Annual emission report formats to standardise submissions,
- Verification checklists used by accredited third-party verifiers.

All of these are generally provided in Excel or PDF formats, allowing for digital editing or physical use. Access is largely restricted to registered entities through the ETS Reporting Tool (ERT).⁵⁶⁴

Figure 31. EU ETS Reporting page screenshot

Action	Interface
Download the report file (in xlsx format)	
Add attachments (except viewer roles)	
Download All attachments	
View attachments (depending on the visibility selected when uploading an attachment)	
Delete attachments (depending on your role and the visibility of attachment)	
Add comments to it (except viewer roles)	
View comments (depending on the visibility selected when adding your comment)	 Comments 
Edit and delete comments submitted by you	 
Consult the History of submission of all changes applied to the report. You can also download previous versions of the reports and the whole history of submission in PDF.	 History of submission   

Beyond the EU ETS, **other compliance systems operate with similar structured support**. For example, the REMIT requires participants in wholesale energy markets to first register via the Centralised European Register of Energy Market Participants (CEREMP),⁵⁶⁵ hosted by ACER, through their relevant national regulatory authority, in order to be able to report trades and orders using dedicated Registered Reporting Mechanisms (RRMs). Guidance on transaction reporting, including templates and TRUM (Transaction Reporting User Manual) is available through the webpage.⁵⁶⁶ The Agency also offers the stakeholders the opportunity to submit any REMIT-related questions that

⁵⁶³ European Commission. *Monitoring, reporting and verification*. https://climate.ec.europa.eu/eu-action/carbon-markets/eu-emissions-trading-system-eu-ets/monitoring-reporting-and-verification_en

⁵⁶⁴ Environmental Protection Agency (EPA). EU ETS reporting for stationary installations. <https://www.epa.ie/our-services/licensing/climate-change/eu-emissions-trading-system-/emissions-trading-system---stationary-installations/eu-ets-reporting>.

⁵⁶⁵ https://www.acer-remit.eu/ceremp/home?lang=en_AT&nraShortName=1.

⁵⁶⁶ Agency for the Cooperation of Energy Regulators (ACER). REMIT reporting guidance. <https://www.acer.europa.eu/remit-documents/remit-reporting-guidance>.

are not addressed in the existing online documentation via the online REMIT Query Form.⁵⁶⁷

Advanced techniques

Advanced techniques build on structured digital workflows, centralized data management, and partial automation of routine tasks by adding **regulatory intelligence** and **risk-oriented orchestration**. In practice, these integrated software platforms combine automated data capture/validation and report generation with capabilities such as horizon scanning and machine-readable obligation tracking, cross-framework mapping and taxonomy harmonization, predictive analytics for risk and workload forecasting, scenario modelling for audit/readiness, and risk-based prioritization of compliance actions across jurisdictions.

These systems are **not fully autonomous**: they streamline and standardize processes but still require **human oversight** for configuration, governance, interpretation, and decision-making. Implementation typically embeds compliance workflows within **multifunctional enterprise platforms** (e.g., ERP/QMS/EHS modules) alongside procurement, production, quality, finance, or HR, and may be delivered as **horizontal compliance suites** or **sector-specific solutions** integrated with domain operations and regulatory regimes. The result is **proactive, context-aware compliance management** that remains anchored in accountable human-in-the-loop controls.

Illustrative example 1: Navitasoft⁵⁶⁸

This is universal reporting tool for compliance with the Regulation on Wholesale Energy Market Integrity and Transparency (REMIT). It integrates with in-house or third-party trading systems to automatically collect and submit required data to ACER. The platform tracks regulatory updates and adapts accordingly. The system also reduces the manual effort through predefined processes, but it is not fully automated. Users remain actively involved in reviewing outputs, resolving data inconsistencies, and interpreting regulatory nuances to ensure accurate reporting. Reports are generated in formats compatible with ACER's REMIT Portal, reducing duplication and administrative burden.

Key features include:

- Automated data collection and validation from internal trading systems, eliminating the need for duplicate data entry;
- Integration with third-party vendor platforms and in-house trading applications;
- Regulatory tracking, with built-in updates to reflect amendments to REMIT and ACER reporting requirements.

Automated techniques

Automation is increasingly applied in compliance processes within the energy sector, particularly in areas such as **data collection, analysis, and reporting**. These systems are designed to **improve operational efficiency and consistency by reducing the need for manual input**. Once configured, automated tools can continuously process large volumes of operational data, identify anomalies, and generate compliance reports with limited human intervention. This reduction in manual input is achieved through features such as **pre-configured data, rule-based validation checks, and automated**

⁵⁶⁷ Agency for the Cooperation of Energy Regulators (ACER). REMIT documents. <https://www.acer.europa.eu/remit-documents>.

⁵⁶⁸ <https://navitasoft.com/en/products/remit-cs>.

formatting of reports according to regulatory standards. By eliminating repetitive tasks—such as manual data entry, cross-referencing of regulatory thresholds, and formatting of submission templates – automation allows compliance teams to focus on oversight and interpretation rather than execution.

Illustrative example 1: Enreg-E⁵⁶⁹

An automated regulatory compliance platform offering a single access point to all relevant energy laws, regulations, and publications. It streamlines compliance management by centralising information, reducing the risk of missing updates, and supporting long-term regulatory planning. This approach helps support both operational compliance and strategic planning, particularly in contexts where regulations span multiple jurisdictions or evolve rapidly.

Key features include:

- Centralised access to energy laws, directives, and regulatory publications;
- Change tracking tools to monitor updates across jurisdictions and highlight relevant regulatory amendments;
- Machine learning and NLP-based analytics to extract obligations and keywords from complex regulatory texts;
- Built-in workflows to map regulations to internal policies, processes, risks, and controls for full traceability.

3.7.3.2 Current use of reporting techniques

Available evidence on the use of different techniques for regulatory compliance and reporting in the energy sector remains limited. Preliminary findings suggest that these methods tend to be applied selectively, with different techniques being employed to conduct specific activities.

Targeted interviews with software providers have highlighted a **low adoption rate of automated solutions for legislative monitoring**. Most companies still rely on standard methods (i.e., Word or PDF documents) and manual work to understand and interpret regulatory requirements, follow legislative updates, identify compliance gaps, and take corrective action. Although several automated solutions exist that reduce reliance on these formats and eliminate the back-and-forth between disparate data sources and applications, their adoption in the energy sector remains limited – particularly when compared to more digitally mature industries such as finance.

For report preparation and submission to official portals, reliance on a mix of standard and advanced techniques appears as most likely. Submission via web-based forms, for example, allows for the automation of data collection and processing. However, it still demands manual input when sharing the data with the competent authorities. By contrast, machine-to-machine integration enables far greater automation, including during the submission process. **Full end-to-end automation of compliance and reporting processes, however, remains unlikely at present.**

3.7.4 Challenges and barriers

3.7.4.1 Overview of challenges per reporting technique

Standard tools

⁵⁶⁹ <https://enreg-e.com/>.

Standard compliance and reporting tools, such as **paper files and spreadsheets**, present recurring problems. **Time and errors** are a major issue: manual data entry and audits are slow, resource-intensive, and prone to mistakes. Manual audits required under the Energy Efficiency Directive absorb substantial resources for large companies and their auditors, while the need to repeatedly enter information by hand prolongs reporting processes. **Data quality** is another difficulty, as fragmented templates, inconsistent formats and repeated transcription reduce the reliability of records. Evidence from surveys of energy audit practices shows that fragmented templates and inconsistent formats contribute to data loss and unreliable records, making compliance more burdensome and less reliable.⁵⁷⁰ **Security and confidentiality** also represent significant concerns. Physical records can be accessed or copied without leaving an electronic trail, which makes it difficult to monitor who has accessed sensitive data. This is particularly problematic when confidential information about market positions or plant operations is exposed, potentially violating REMIT obligations or data-protection rules. Another difficulty lies in the limited capacity to share information across locations. Energy companies often operate assets across several Member States, and paper files cannot be quickly shared with cross-border teams or regulators. This leads to delays and inconsistent submissions.⁵⁷¹

Spreadsheets also fall short in compliance terms. They generally **lack encryption, immutable audit logs, and role-based access controls**, meaning that unauthorised access or alterations may go unnoticed. Weak safeguards in turn undermine data privacy, integrity, and cybersecurity – particularly when compliance reporting involves commercially sensitive information or operational data.⁵⁷² There are also clear **technical limits**. Spreadsheets are vulnerable to corruption, offer weak version control, and cannot reliably process very large datasets required under rules such as the Electricity Transparency Regulation. Their analytical capacity is also limited, making it difficult to detect anomalies or assess risks.

Advanced tools

Advanced compliance software brings new challenges. One of the most significant challenges is **data privacy and ownership concerns**. Advanced platforms often rely on cloud services and external vendors, raising the question of who ultimately controls sensitive trading or operational data.⁵⁷³ Once a company commits to a particular software ecosystem, **vendor dependence** can also become a problem: proprietary formats and custom integrations make switching costly and difficult, while technical standards set without broad consultation risk reinforcing this lock-in.⁵⁷⁴ At the same time, **cybersecurity**

⁵⁷⁰ TwinEU (2025). *Digitalization challenges and opportunities in the future energy system. Deliverable D2.1*. https://twineu.net/wp-content/uploads/2025/01/TwinEU_D2.1_V1.0.pdf

⁵⁷¹ UNECE (2024). *Governance policy of digitalization in the energy sector: Case study. Group of Experts on Energy Efficiency, Task Force on Digitalization in Energy*. https://unece.org/sites/default/files/2024-09/GEEE-11.2024.INF_3_governance.policy_case.study_.2024.pdf

⁵⁷² Ibid.

⁵⁷³ UNECE (2024). *Governance policy of digitalization in the energy sector: Case study. Group of Experts on Energy Efficiency, Task Force on Digitalization in Energy*. https://unece.org/sites/default/files/2024-09/GEEE-11.2024.INF_3_governance.policy_case.study_.2024.pdf

⁵⁷⁴ Shievam Kashyap, Wieland Schwinger, Stefan Grunberger, Christoph Schaffer, Georg Hartner, *Energy Data Governance Fragmentation in Europe: Challenges, Models, and Pathways to Integration, Department of Smart and Interconnected Living, University of Applied Sciences Upper Austria* (2022). https://eddie.energy/files/eddie/media/deliverables/EEM25_EDDIE.pdf

risks are growing, with energy utilities facing a rising number of attacks and high costs of data breaches.⁵⁷⁵

Finally, **technical reliability** remains a risk. Bugs, failed updates or crashes can interrupt reporting cycles. Given strict deadlines, such as the annual ETS emissions surrender, even short disruptions may result in non-compliance and financial penalties.

Automated tools

Technical complexity is one of the biggest challenges. Automated platforms require advanced data architectures and integration with existing production and compliance systems, which is often difficult given ageing infrastructure and fragmented formats. Interoperability problems persist when tools rely on proprietary protocols, making it hard to exchange data across platforms.⁵⁷⁶

The effectiveness of automated systems using AI in regulatory compliance also depends on the **quality and availability of the underlying data**.⁵⁷⁷ Inaccurate, incomplete, or biased datasets can significantly compromise the reliability of AI-generated analysis, leading to flawed compliance assessments. For renewable energy companies operating across multiple jurisdictions, data collection is further complicated by regional disparities in regulatory standards, reporting formats, and access to information. The evidence collected during targeted interviews with software providers underlined that automation depends on good-quality input data and reliable integration with existing processes. Poor data governance or weak system connectivity can quickly erode the expected efficiency gains of automated tools.⁵⁷⁸ These inconsistencies undermine the reliability and uniformity of AI-driven compliance evaluations. Furthermore, AI systems are prone to inheriting and amplifying historical biases embedded in training datasets, which, if left unchecked, can distort compliance outcomes and exacerbate regulatory risks.

A further issue is the **lack of guidance and standards**. With few benchmarks or common frameworks, companies are hesitant to trust automated outputs for compliance, particularly in cross-border contexts. Data security concerns add to this reluctance, since these systems depend on access to large datasets and often operate on cloud-based resources, raising the risk of breaches or unauthorised use.⁵⁷⁹ This lack of clarity discourages investment and slows adoption of automation for compliance purposes.

Finally, there is **regulatory uncertainty**. Technology evolves faster than legislation, and companies remain unsure how regulators will treat reports generated automatically.⁵⁸⁰

⁵⁷⁵ International Energy Agency (IEA), *Mapping Green and Digital Energy Jobs: Trends and Insights from Online Job Postings*, 2023. : <https://iea.blob.core.windows.net/assets/16986de9-b1f9-4243-8411-a2a12f00413c/MappingGreenandDigitalEnergyJobs.pdf>

⁵⁷⁶ TwinEU (2025). Digitalization challenges and opportunities in the future energy system. Deliverable D2.1. https://twineu.net/wp-content/uploads/2025/01/TwinEU_D2.1_V1.0.pdf

⁵⁷⁷ Future priorities for AI in an evolving digital-energy cyber-security regulatory framework. <https://www.smart-energy.com/industry-sectors/cybersecurity/future-priorities-for-ai-in-an-evolving-digital-energy-cyber-security-regulatory-framework/>

⁵⁷⁸ Based on interviews with a regulatory compliance software provider conducted for this study.

⁵⁷⁹ Future priorities for AI in an evolving digital-energy cyber-security regulatory framework. <https://www.smart-energy.com/industry-sectors/cybersecurity/future-priorities-for-ai-in-an-evolving-digital-energy-cyber-security-regulatory-framework/>.

⁵⁸⁰ TwinEU (2025). Digitalization challenges and opportunities in the future energy system. Deliverable D2.1. https://twineu.net/wp-content/uploads/2025/01/TwinEU_D2.1_V1.0.pdf.

3.7.4.2 Barriers to adoption

Although advanced and automated tools offer significant potential to streamline compliance, their uptake is slowed by several broader barriers.

A first barrier is **human capital and skills**. Deploying advanced compliance software often requires extensive training for employees to be able to use new systems, which can temporarily reduce productivity and efficiency. According to a report conducted by the International Energy Agency, energy utilities already face significant competition when recruiting among a limited pool of workers with digital skills and that companies will increasingly need to recruit or retrain staff in data management and analytics.⁵⁸¹

Integration and interoperability are recurring problems. Energy companies often rely on legacy systems such as SCADA or asset management platforms, and new compliance software does not always connect smoothly with them. Inconsistent formats and national differences further complicate cross-border reporting. In addition, there are also **concerns around customisation**, as off-the-shelf software may not fully match the reporting needs of specific firms or regulators, forcing companies to adapt their processes to the tool rather than the other way around. This can create compliance inefficiencies and disadvantages for smaller operators.⁵⁸²

Stakeholders highlighted that interoperability challenges are often compounded by **long implementation periods** and **difficulties in integrating automated platforms into existing company systems**. Without smooth integration, automation can actually add complexity rather than reduce it.⁵⁸³

Financial costs represent a further obstacle. The high upfront spending required for licences, infrastructure and secure networks, combined with long-term maintenance needs, remains prohibitive for smaller operators and discourages cautious investment by larger firms. The acquisition and maintenance of enterprise-level software packages can be prohibitively expensive, particularly for small and medium-sized energy providers. Digital technologies demand substantial upfront investment and long-term financial resources, and smaller providers are often uncertain about the return on investment.⁵⁸⁴ These high costs discourage adoption and risk widening the gap between large incumbents and small operators. Automation requires major investment in computing capacity, sensors and communications infrastructure, which many smaller providers cannot sustain.⁵⁸⁵ This is compounded by the risk of vendor lock-in, as automated systems often rely on proprietary algorithms and data models that are difficult or costly to replace.⁵⁸⁶ Interview inputs confirmed that cost is the primary barrier to adoption of AI-

⁵⁸¹ International Energy Agency (IEA), Mapping Green and Digital Energy Jobs: Trends and Insights from Online Job Postings, Paris: IEA, 2023. <https://iea.blob.core.windows.net/assets/16986de9-b1f9-4243-8411-a2a12f00413c/MappingGreenandDigitalEnergyJobs.pdf>.

⁵⁸² Shievam Kashyap, Wieland Schwinger, Stefan Grunberger, Christoph Schaffer, Georg Hartner, Energy Data Governance Fragmentation in Europe: Challenges, Models, and Pathways to Integration, Department of Smart and Interconnected Living, University of Applied Sciences Upper Austria (2022). https://eddie.energy/files/eddie/media/deliverables/EEM25_EDDIE.pdf

⁵⁸³ Based on interviews with a regulatory compliance software provider conducted for this study.

⁵⁸⁴ Ibid.

⁵⁸⁵ Ibid.

⁵⁸⁶ Shievam Kashyap, Wieland Schwinger, Stefan Grunberger, Christoph Schaffer, Georg Hartner, Energy Data Governance Fragmentation in Europe: Challenges, Models, and Pathways to

powered compliance tools. The perceived lack of a clear return on investment, combined with lengthy deployment and integration processes, makes it difficult for providers to convince clients of the financial advantages of automation.⁵⁸⁷

Finally, **regulatory uncertainty** continues to weigh on adoption. The absence of clear recognition of automated outputs means that companies cannot be sure how regulators will evaluate machine-generated reports, which reduces confidence and slows wider uptake.

3.7.4.3 Assessment of costs and benefits

Overview of costs

The costs associated with **Article 8 of the EED** arise from the requirement for large enterprises to conduct an **energy audit** every four years. These costs include **internal resources** dedicated to data collection and processing, analysis, and reporting, as well as monitoring key performance indicators in preparation for the next audit cycle. In addition, **auditor fees** further increase the overall compliance and reporting burden.

The costs associated with **REMIT compliance** and reporting can be grouped into several categories. First, firms must **adapt their IT infrastructure** to ensure interoperability with Inside Information Platforms (IIPs) and Registered Reporting Mechanisms (RRMs). These expenses may involve developing XML schemas, APIs, and other automated tools to collect and transmit data in the required format. Second, **labour costs can be substantial**, as companies need specialized teams to manage reporting activities and legal experts to monitor regulatory updates and handle tasks such as registration and updates of registration. Finally, ACER charges **specific fees** to cover its operational costs. The recent Commission Decision (EU) 2025/1771 introduced a revised framework combining fixed annual contributions with variable fees linked to market activity levels.⁵⁸⁸ Although these fees are paid by reporting entities (RRMs), they are typically passed on to their clients, resulting in direct compliance costs for market participants.

Compliance with the **EU ETS Directive** entails a broad range of costs. The first category relates to **monitoring, reporting, and verification (MRV)**. Companies must prepare a monitoring plan, which may require updates following process changes or regulatory revisions. Data collection itself generates significant expenses, including monitoring emissions, data processing, and quality assurance. Reporting adds further costs, as firms must complete standardized templates, conduct internal reviews, submit reports via official platforms, and address potential queries from accredited verifiers or competent authorities. Verification costs are also borne by companies, which must engage accredited verifiers before submitting reports. Estimates from a study on German firms indicate that median MRV-related costs between 2010 and 2012 were around EUR 7,000 per year.⁵⁸⁹

⁵⁹⁰ Additional compliance costs include **IT and system integration**, as well as **legal and advisory services**. Companies covered by the EU ETS must also open an account in the Union Registry, for which Member States charge **annual administrative fees**. For

Integration, Department of Smart and Interconnected Living, University of Applied Sciences Upper Austria (2022). https://eddie.energy/files/eddie/media/deliverables/EEM25_EDDIE.pdf

⁵⁸⁷ Based on interviews with a regulatory compliance software provider conducted for this study.

⁵⁸⁸ European Commission (2025). New EU rules on REMIT fees. Directorate-General for Energy, News announcement: https://energy.ec.europa.eu/news/new-eu-rules-remit-fees-2025-09-24_en

⁵⁸⁹ Heindl, Peter. "The impact of administrative transaction costs in the EU emissions trading system." *Climate Policy* 17.3 (2017): 314-329.

⁵⁹⁰ This amount corresponds to EUR 10,000 in 2024 prices.

example, in Ireland the fee is EUR 350 for an Operator Holding Account and EUR 1,000 for a Trading Account,⁵⁹¹ while Dutch authorities charge EUR 400.⁵⁹²

To comply with **E-PRTR obligations**, companies incur costs to **monitor emissions, compile data, verify its accuracy, fill out national reporting templates, and submit reports**. **Indirect costs** arise from staff training and periodic process updates in response to regulatory changes. A survey of reporting facilities in the Czech Republic found that average annual reporting costs amounted to EUR 365.2.⁵⁹³ These costs encompass staff wages for reporting activities, in-service training, software and technology (including depreciation and updates), measurement or calculation efforts, and outsourcing expenses such as contractor agreements.⁵⁹⁴

To comply with the **transparency regulation**, companies must ensure **IT integration** with the data providers, TSOs and the ENTSO-E. Additional initial setup costs include **automating data flows and implementing compliance checks**. **Ongoing operational expenses** cover staff wages for monitoring, quality assurance, and data submission, as well as system maintenance, training, and potential outsourcing – such as engaging platform intermediaries for reporting.

Interviews with software providers reveal that **labour is a major cost driver** in achieving and demonstrating compliance in the energy sector. The compliance process remains highly labour-intensive, requiring substantial resources to interpret the regulatory framework, track updates, and manage variations in requirements. Added to this are the personnel costs for collecting and processing data, preparing reports, and submitting them to authorities. These costs become **particularly significant when companies rely on conventional methods with limited automation**.

Automated solutions, on the other hand, offer **significant reductions in labour costs** as tasks typically performed by hired workers are now delegated to trained robots. However, these software solutions require **significant upfront investments**. To ensure seamless integration across the different systems that a company is using, the **IT infrastructure of the company may need to be adapted**, raising the costs of implementation. Similarly, **subscription and licensing costs** can also be quite consequential, especially when compared to more traditional methods.

Expected benefits

Interviews with software providers indicate that the **primary benefit of adopting automated tools stems from cost reductions**, particularly through decreased labour requirements. For example, automating legislative monitoring can significantly reduce the reliance on legal experts for time-consuming tasks such as gathering legislative acts, structuring embedded data, and tracking ongoing developments. A further advantage is that the time spent on low-value activities can be redirected toward higher-value activities, including analysis and strategic planning.

⁵⁹¹ <https://www.epa.ie/our-services/licensing/climate-change/eu-emissions-trading-system-/union-registry/union-registry-accounts/>.

⁵⁹² <https://www.emissionsauthority.nl/topics/emissions-trading-eu-ets-2021-2030/eu-ets-registry-2021-2030/accounts/trading-account>.

⁵⁹³ Vejchodská, E., Slavíková, L., & Malý, V. (2016). Evaluating the Regulatory Burden: Pollutant Release and Transfer Reporting Costs. *Prague Economic Papers*, 25(6), 671–685. <https://doi.org/10.18267/j.pep.583>.

⁵⁹⁴ Note that the Czech PRTR concerns eight times more firms than E-PRTR and may therefore imply higher reporting costs.

Another significant benefit of adopting automated tools lies in the **increased efficiency**. By automating manual, repetitive tasks, companies can process larger volumes of data at a much faster rate. This results in a more streamlined compliance and reporting process.

Lastly, automated solutions **enhance and augment the possibilities of real-time monitoring**, allowing companies to identify gaps in the compliance system and implement corrective actions. This reduces the risk of fines and reputational damages.

Although advanced and automated techniques may enhance output accuracy, the available data does not provide conclusive evidence to confirm this effect.

Qualitative cost-benefit analysis

Taking stock of the previous results, the **adoption of automated techniques in the energy sector would entail higher capital and operating costs**, as available software solutions remains particularly expensive. **Technical costs would also be higher**, as adopting advanced and automated techniques would require adaptations of the IT infrastructure to ensure seamless integration with other internal processes. **Labour costs**, on the other hand, are **expected to decrease** by a significant margin.

Advanced and automated techniques are also expected to generate consequential benefits across the following dimensions: enhanced efficiency, greater cost savings, and real-time monitoring.

Overall, the cost-effectiveness of replacing standard techniques with advances and automated techniques depends on the net costs and the size of the relative benefits. The analysis may vary across companies. Larger companies that are subject to a higher number of regulations and reporting obligations are arguably in a better position to spread the sunk cost of such investment and, therefore, more likely to find advanced and automated tools as cost effective.

Table 31. Qualitative assessment of the costs and benefits associated with different reporting techniques

Type of reporting technique	Costs				Relative Benefits			
	Capital & operating	Labour	Training	Technical	Accuracy	Efficiency	Cost savings	Real-time monitoring
Standard	Reasonable	Significant	Inconclusive	Reasonable				
Advanced / automated	Significant	Reasonable	Inconclusive	Significant	Inconclusive	Significant	Significant	Significant

Source: The data used to create this table were collected through interviews of software providers.

3.7.5 Policy recommendations

Interview insights and desk research show that the uptake of digital compliance tools in the energy sector remains limited due to low awareness, regulatory uncertainty, and lack of institutional endorsement.

Promoting visibility, trust, and endorsement of digital compliance tools

Increase the visibility and credibility of automated solutions in the energy sector through coordinated EU-level awareness and endorsement activities.

Operational actions:

- **Launch EU-led communication campaigns** showcasing the benefits of automation in reporting, monitoring, and auditing, with practical examples from energy utilities and market operators.
- **Organise webinars, workshops, and industry roundtables to showcase practical applications and software for compliance in the energy sector.**

- **Support demonstration projects and cross-border pilots** that test automated compliance systems in real operational environments, assessing impacts on cost, accuracy, and reporting timeliness.
- **Conduct comparative reviews and case studies** to assess effective digital tools and pilot projects that demonstrate measurable efficiency gains and error reduction.

Establishing regulatory guidance and frameworks for responsible adoption

Provide a clear and enabling policy environment for the safe and responsible integration of automated compliance tools across the energy sector.

Operational actions:

- Incorporate supportive provisions for digital and automated compliance systems into relevant **EU energy and digital legislation**.
- **Develop EU-level guidance or voluntary standards** defining good practices for automation in compliance and reporting, including data security, transparency, and human oversight requirements.
- **Create regulatory sandboxes** or controlled testing environments where energy companies can pilot automated solutions under supervisory support.

Facilitating interoperability and data harmonisation across compliance regimes

Reduce the administrative burden and error risk created by fragmented reporting portals, divergent templates, and inconsistent data standards by promoting interoperability and harmonisation across EU and national energy-reporting systems.

Operational actions:

- **Develop a common EU data model and reporting taxonomy** for energy compliance and monitoring obligations under different regulations.
- **Set interoperability requirements and open APIs** for EU and national portals to enable a “report once” approach and cross-regime data reuse.

4 Task 3 International aspects on the data economy

4.1 Methodological approach to Task 3

Task 3 aims to analyse measures that promote data access, to assess the barriers to cross-border data flows, and to propose actionable recommendations to address the barriers and exploit opportunities for international data cooperation. The task is divided in five highly interrelated subtasks:

Task 3.1 Analysis of measures to promote data access

Task 3.2 Analysis of measures to protect sensitive and highly sensitive data

Task 3.3 Assessment of potential opportunities for international data cooperation

Task 3.4 Interlinking data sharing ecosystems

Task 3.5 Non-EU compliant data processing

The task targets 8 third countries, Australia, Japan, South Korea, China, United Kingdom, Brazil, United States, and Canada.

The main methodological aspects of the work carried out under the respective subtask are briefly described below.

4.1.1 Mapping and analyses of measures to promote data access and use; to protect sensitive and highly sensitive categories of data; and of potential opportunities for international data cooperation (Task 3.1 to Task 3.3)

Due to the strong links and overlaps between the subtasks 3.1, 3.2 and 3.3 the work was performed in parallel.

Desk research

The search for relevant measures was performed separately for each targeted third country and with additional searches for international fora, following a predefined search strategy, including relevant search terms for different types of measures. Leveraging the overlaps between the scope of the different tasks, measures found using the search criteria defined for Task 3.1 were screened for their relevance also for Tasks 3.2 and 3.3. The same procedure was then followed for Tasks 3.2 and 3.3.

An initial round of desk research resulted in a total of 185 measures (including 18 literature references). Additional searches have followed and the total number of identified measures are over 200.

Selection of most interesting measures

Starting from the long list of identified measures an iterative process was used to narrow down and select those of particular interest (or potential interest), for the subsequent production of policy fiches. The Study team shared an initial short list of 53 measures with DG CNECT, which were discussed in a technical meeting, leading to revisions according to agreed priorities. A final shortlist with 20 measures and 20 back-ups was agreed after a few iterations.

Analysis and production of policy fiches

The production of policy fiches (Task 3.1 and 3.2) is based on additional desk research, following pre-defined templates for the type of information to be included. Two pilot fiches were produced, leading to adjustments of the templates. In the technical meetings with DG CNECT flexibility in the use of the templates was agreed, due to the big differences between the type, scope and jurisdictions of the measures covered.

The produced policy fiches are presented in Annex 4. Findings from the policy fiches are integrated into the country- and global analyses presented in section 4.2.

Mapping of definitions on sensitive/highly sensitive data

A mapping of existing definitions on sensitive/highly sensitive data was done, considered to be of significant added value for DG CNECT for the purpose of policy-development. This mapping of definitions and related laws and regulations is presented in Annex 5, and the concluding analysis is integrated in the country- and global analyses (section 4.2).

4.1.2 Country and global analysis of opportunities (Task 3.3) as well as non-EU-compliant data processing and data gathering practices (Task 3.5)

Based on the desk research performed for the targeted countries (Task 3.1-3.3), the policy fiches produced, and additional desk research, country analyses were performed (see section 4.2.1). The country analyses provide:

- A general introduction of the country analysed, selected initiatives and identified trends.

- An analysis of the degree of regulatory/legal convergence compared to the EU. This includes the identification of third country data handling practices that do not comply with EU law, i.e., practices that have no or limited convergence with the EU data regulations. It also includes any findings related to the mapping of sensitive and highly sensitive data processing practices.

- Conclusions and main lessons learnt.

The results from the country analyses, for example, regarding convergence (or divergence) with EU data regulations, are summarised in the global analysis (see section 4.2.2).

4.1.3 Use case analyses (Task 3.4 Interlinking data sharing ecosystems)

Task 3.4 involved the identification, investigation and reporting of six use cases representing opportunities for international data sharing and interlinkage with European data ecosystems including the Common European Data Spaces in relevant sectors.

4.1.3.1 Rationale for the selection of use cases

Our final approach for identifying and analysing these six use cases was based on a rigorous, two-phased methodology designed to select the most strategically valuable and impactful examples of international data sharing. The initial phase involved a broad scan and expert assessment of over ten potential data-sharing scenarios across a variety of sectors, including energy, finance, and public services. From this broader pool, a set of critical criteria was developed to evaluate each use case's potential to advance the European Union's strategic objectives. These criteria included the use case's alignment with key EU priorities (such as the Green Deal or industrial competitiveness), its potential to create significant economic or societal value through data, and its capacity to demonstrate the practical application of EU-championed federated architectures, particularly those from initiatives like Gaia-X and Catena-X.

The six cases detailed in this report were selected for a deep dive because they consistently scored highest across these criteria. They represent the most promising scenarios for translating the EU's vision for data sovereignty into a tangible, global reality. In addition to their strategic and high-value attributes, a key factor in their selection was the availability of robust information, sufficiently detailed for a preliminary analysis. This ensured that our discussion of each case could be grounded in specific details regarding

the data to be shared, the stakeholders involved, and the unique governance challenges that must be overcome. This selective process, guided by expert judgment, allows this report to focus on the most illustrative and strategically significant opportunities for the EU to shape the global data economy.

4.1.3.2 Use case research, interviews, report drafting and validation

Based on the draft descriptions of the selected use cases semi-structured interviews with minimum two stakeholders per use case were performed. The interviews served both to get validation of the preliminary use case rationale and definition, and to gather additional details and opinions from subject matter experts. In a first round the Study team engaged 20 persons in 15 interviews and/or written contributions. In a second step, the interviewees, as well as other contacted stakeholders who had demonstrated interest, were asked to comment and validate the updated use case descriptions, adding contributions from 4 additional persons (24 external experts engaged in total).

Overall, the stakeholder engagement activities have contributed significantly to improve the use case descriptions and analyses. However, the information collected are qualitative expert judgements from a small number of stakeholders per use case, and the feedback received can therefore not be considered to be representative for all stakeholders in the respective sectors.

The questionnaires used contained one common part with general questions for all the use cases and one part with use case specific questions. To increase the number of engaged stakeholders they were given the opportunity to answer either in writing or participate in an online meeting/interview. A list of the interviewees is presented in Annex 6.

Additional desk research was used to support a more in-depth analysis than what was performed for the policy fiches. Findings from the use cases are reported in section 4.3, covering for each use case a basic description, requirements and barriers along with potential opportunities and lessons to be learned.

It should be noted that due to the exploratory nature of the study, the selected use cases are quite widely defined. Full-fledged use case analyses are not feasible within the scope of this study. The goal is to present preliminary descriptions and conclusions for each use case, as well as recommendations on possible next steps. These next steps would surely involve further assessment and analysis, as well as clear definitions of scope and limitations.

4.2 Key findings from the country and global analysis

The present section includes a comprehensive analysis of the data *state of the art* in the eight selected countries through their legal, technological and strategic contexts in order to identify underlying trends and the degree of regulatory/legal convergence with the EU. To this end, the study of selected initiatives on data protection and sharing, the categorisation of sensitive and highly sensitive data in national legislations as well as additional desk research are central. The findings of each country analysis are subsequently integrated into a global analysis, designed as a synthesis of the identified global trends that define the present and future of data.

The conclusions and policy recommendations to the European Commission and the EU are later integrated in section 4.4.

4.2.1 Country analyses

4.2.1.1 Australia

Introduction

Since 2020 Australia has seen increasing efforts in keeping pace with developments in the field of data spaces and data sharing, as well as in the field of generative artificial intelligence. In 2021 the Australian Government published the first **Australian Data Strategy**, aimed at positioning “*Australia as a leading digital economy and society by 2030.*”⁵⁹⁵ Two years later, the **2023-2030 Australian Cyber Security Strategy** was released, focusing on “*improving cybersecurity, manage cyber risks and better support citizens and businesses to manage the cyber environment around them.*”⁵⁹⁶ Although not the only initiatives focused on building a secure and trusted environment for the use of data, the two strategies reflect the current state of the art in Australia regarding data.

On the one hand, Canberra is aware of the importance of launching initiatives and legislation that support the development of a secure and trusted data environment in Australia. I.e., to implement frameworks to boost data sharing through data spaces and sharing protocols as well as legislation on data protection. However, on the other hand, the rapid technological advancements on data, together with the relatively recent start of this governmental approach on data, is making it difficult to keep up with such developments. In this context, the two selected measures: the **2022 Data Availability and Transparency Act Scheme** and the **2024 Guidance on privacy and developing and training generative AI models** reflect the identified trend, as they represent the push for enforceable measures on data sharing and a voluntarily approach towards data protection in the absence of specific enforceable legislation respectively.

The Data Availability and Transparency Act Scheme (DATA Scheme) stands as a joint effort together with the searchable Australian Government Data Catalogue and Dataplace, to share Government data with accredited public users. The DATA Scheme is based on the 2022 Data Availability and Transparency Act and seeks to increase the availability and use of Australian Government data - both personal and non-personal - by government bodies and universities.⁵⁹⁷

The Guidance on privacy and developing and training generative AI models acts as voluntary guidelines to ensure compliance of generative AI development with Australian privacy law - the **1988 Privacy Act** and the **Australian Privacy Principles guidelines (APPs)**. The Guidance is framed in a larger effort by the Australian Government to properly regulate AI, as current measures such as the Guidance on privacy and the use of commercially available AI products, the AI Ethics Principles and the Voluntary AI Safety Standard are voluntary. Currently, the **2024 Proposals Paper for introducing mandatory guardrails for AI in high-risk settings** stands as the only enforceable AI-specific proposed initiative to date.

Regulatory and Legal Convergence - Partial convergence

The EU and Australia have both recognised the critical importance of promoting the share of data and innovation through data usage - generative AI learning for example - while ensuring a robust and adaptive data protection framework. In addition, both have core data protection legislations: EU’s 2016 GDPR and Australia’s 1988 Privacy Act, with similar categorisations for the protection of personal information under the concepts of

⁵⁹⁵The Australian Government (2021), *Australian Data Strategy: The Australian Government’s whole-of-economy vision for data*. Available online: <https://www.finance.gov.au/sites/default/files/2022-10/australian-data-strategy.pdf>

⁵⁹⁶ The Australian Government (2023), *2023–2030 Australian Cyber Security Strategy*. Available online: <https://www.homeaffairs.gov.au/cyber-security-subsite/files/2023-cyber-security-strategy.pdf>

⁵⁹⁷ For information on the specific accredited public users and on the data governance structure and data sharing restrictions see Annex 4.1.

sensitive data and personal information.⁵⁹⁸ Despite these laws acting as backbones for the entirety of data initiatives and legislation, both countries approach these two elements: promotion and protection, in distinct ways.

Australian data regulation tends towards pragmatism and a rather business-friendly approach, while the EU has developed a strong and protective framework, characterised by its comprehensive scope, its enforcement power and its protection of Europeans citizens' data rights within EU borders and beyond. In this sense, Australia's Privacy Act does not apply to small businesses - businesses with an annual turnover of \$3 million or less - employee data and political parties, amongst others.⁵⁹⁹ Additionally, Australian legislation does not include an equivalent to the "right to erasure" under GDPR, or a detailed differentiation between data processors and data controllers.⁶⁰⁰ For overseas transfers, Australian legislation only requires business to "take reasonable steps" to ensure that foreign recipients do not breach the APPs, however it does not include the need for consent from individuals for such transfers.⁶⁰¹ In fact, Australia does not currently hold the adequacy status granted under the data protection adequacy for non-EU countries framework.

Regarding the development of data spaces, Australia is actively promoting different initiatives including the Government-led **Australian Government Data Catalogue and Dataplace**.⁶⁰² The Catalogue and Dataplace currently include more than three thousand data assets from the Australian Government and its related agencies, responsible of assessing whether the asset is openly accessible or other conditions or restrictions apply and granting access to the 125 onboarded accredited public and private organisations.⁶⁰³ Additionally, this limitation of data spaces to assets exclusively held by the public sector is being complemented by the development of the **Australian Dataspaces**⁶⁰⁴, which seeks to create trusted and secure ecosystems for data exchanges between research, industry and government, with a special focus on underutilised data assets.

In the field of Artificial Intelligence, a clear and growing interest has been identified, with a proposed **National AI Capability Plan** to be delivered by the end of 2025 which is expected to be built on existing initiatives - National Reconstruction Fund, Research and

⁵⁹⁸ For more information on the treatment of sensitive and highly sensitive data in Australia see Annex 5.

⁵⁹⁹ Witzleb, Normann (2020). *Data privacy: stricter European rules will have repercussions in Australia as global divisions grow*. Monash University. Available online: <https://www.monash.edu/law/news-and-events/news/2020/data-privacy-stricter-european-rules-will-have-repercussions-in-australia-as-global-divisions-grow>

Potter, Angela et al. (2020). *Comparing privacy laws: GDPR v. Australian Privacy Act*. One Trust Data Guidance. Available online: https://www.dataguidance.com/sites/default/files/gdpr_v_australia.pdf

⁶⁰¹ Australian Government, Office of the Australian Information Commissioner (2018). *Australian entities and the European Union General Data Protection Regulation*. Available online: <https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/more-guidance/australian-entities-and-the-european-union-general-data-protection-regulation>

⁶⁰² For more information on the Australian Government Data Catalogue and Dataplace see Annex 4.1.

⁶⁰³ Dataplace. *Australian Government Data Catalogue*. Available online: <https://www.dataplace.gov.au/search/>

⁶⁰⁴ Australian Dataspaces is a limited company which combines the Australian National Data Service, the National eResearch Collaboration Tools and Resources and Research Data Services.

Development Tax Incentive, the National AI Centre and a network of government-supported AI Adopt Centres. However, these efforts have not yet been matched with enforceable legislation to guide generative AI development and usage. Unlike EU's **Artificial Intelligence Act**, which entered into force in August 2024, Australia has no specific legislation for regulating AI - apart from the ongoing Proposals Paper for introducing mandatory guardrails for AI in high-risk settings - and currently relies on voluntary measures and guidelines as analysed above.

Nevertheless, Australia has proven to be closely aligned with international efforts to regulate generative AI and build trustworthy data governance and data protection frameworks. Australia has signed together with Korea, Ireland, France and the UK a joint declaration to establish data governance that fosters innovative and privacy-protective AI.⁶⁰⁵ Additionally, the design of Guidance on privacy and developing and training generative AI models sought guidance in similar documents from the UK, France, Canada and -where possible - the OECD.⁶⁰⁶

Conclusions on lessons learnt and opportunities

Australia demonstrates partial convergence with the EU regarding data regulation, with significant potential to deepen both cooperation and alignment in key areas. The relative lack of enforceable legislation on data protection in contexts of AI, and the current reform process of the Privacy Act opens an important window of opportunity for legal alignment between Australia and the EU. This should be fostered by Brussels through diplomatic contacts, the ongoing negotiations for an EU-Australia Free Trade Agreement and the leverage of the data protection adequacy for non-EU countries framework. Furthermore, the relatively early stages in the development of data spaces and Artificial Intelligence present an opportunity for joint innovation initiatives and the adequacy of frameworks fostered by a parallel timeframe.

Despite the rather-wide room for cooperation and convergence, the analysis of the Australian data state of the art provides useful lessons learnt.

The EU does not yet have structured guidelines for all sectors that directly link the development and training of generative AI models to privacy requirements, which are in turn more developed than in Australia taking into account the GDPR and the AI Act. **As recommendation:** the EU could adopt a similar guidance framework to the Australian Guidance on privacy and developing and training generative AI models, which could act as a comprehensive guidance applicable to the entire public and private sectors. This will be highly beneficial for those entities and/or individuals dealing with personal data in light of growing regulation regarding data usage.

The EU currently has two key regulations to foster data sharing and to support the setup and development of the Common European Data Spaces: the 2022 European Data Governance Act and the 2024 Data Act. These are further reinforced by the 2024 Interoperable Europe Act, which enables public administrations to cooperate to ensure a seamless interoperability between IT systems and data assets. However, these regulations have a broad scope in terms of scope, types of data shared, size of the data spaces linked and the targeted users of the shared data. **As recommendation:** the Australian Data Availability and Transparency Act Scheme provides a useful model to

⁶⁰⁵ Australian Government, Office of the Australian Information Commissioner (February 2025), *News article – Joint statement on building trustworthy data governance frameworks to encourage development of innovative and privacy-protective AI*. Available online: <https://www.oaic.gov.au/news/media-centre/joint-statement-on-building-trustworthy-data-governance-frameworks-to-encourage-development-of-innovative-and-privacy-protective-ai>

⁶⁰⁶ For more information on the alignment with international frameworks see Annex 4.2.

develop a structured and secure framework to share government data between public entities based on accreditation. This accreditation system – which is key in Australia's data sharing model of government data assets between public entities – has not yet been explored by any specific EU initiative. Considering the amount of data generated by EU institutions, agencies and bodies, the adoption of a similar scheme could improve the share of data between EU, national and regional authorities. Furthermore, the establishment of an accredited users' framework and an oversight authority would provide greater security to these data transfers by guaranteeing that such accredited users consistently meet the necessary security criteria.

4.2.1.2 Brazil

Introduction

Brazil's data governance framework is anchored by the General Data Protection Law (LGPD)⁶⁰⁷, enacted in 2020, which establishes rights to data ownership, consent, and transparency, while mandating strict protections for sensitive data (e.g., healthcare, biometrics). To promote data sharing and innovation, Brazil launched the National Open Data Policy (INDA)⁶⁰⁸ in 2016, requiring public agencies to publish datasets via the Portal Brasileiro de Dados Abertos (dados.gov.br), which hosts over 10,000 datasets across sectors like transport and energy. For industry, the Brazilian Development Bank (BNDES) funds data-driven projects, including the Industry 4.0 Initiative, which incentivises manufacturers to adopt IoT and open data standards for supply chain optimisation.⁶⁰⁹ The ANPD also introduced the Data Governance and Innovation Programme in 2023, certifying companies that implement ethical AI and data-sharing practices.

Brazil is rapidly advancing its data economy through strategic expansions in data spaces and infrastructure. Collaborations like the Brazil IoT Association's (ABINC)⁶¹⁰ partnership with the International Data Spaces Association (IDSA) aim to build secure, interoperable frameworks for cross-sector data sharing, particularly in agriculture and smart cities. Concurrently, São Paulo is spearheading a tripling of data centre capacity by 2029, enabling scalable storage and access solutions. These efforts are bolstered by open-data projects and fintech innovations, such as consent-based credit models, to democratise data access while supporting SMEs and researchers.

On governance, Brazil is tightening data ownership and privacy rules, mirroring global standards like GDPR. Laws such as LGPD empower users to control personal data, while sector-specific policies address AI-driven ownership complexities and cross-border data flows. Challenges persist in balancing infrastructure growth with sustainable energy practices and adapting corporate strategies to evolving regulations.

Compressive mapping exercise of this study indicated Open Industry program⁶¹¹ as an initiative to further examine. This initiative – launched in April 2025 and managed by ABINC together with the National Confederation of Industry (CNI) – fosters collaboration by sharing anonymised industrial data (e.g., production metrics, R&D insights) among

⁶⁰⁷ The IAPP (October 2020), *Brazilian General Data Protection Law (LGPD, English translation)*. Available online: <https://iapp.org/resources/article/brazilian-data-protection-law-lgpd-english-translation/>

⁶⁰⁸ Government of Brazil, INPI website, *Access to information > Open Data*, available at: <https://www.gov.br/inpi/en/access-to-information/open-data>

⁶⁰⁹ OECD (2022), *Brazil's Digital Government Review*. Available online: https://www.oecd.org/en/publications/digital-government-review-of-brazil_9789264307636-en.html

⁶¹⁰ Brazil IoT Association's (ABINC) website, available at: <https://abinc.org.br/>

⁶¹¹ Open Industry Program website, available at: <https://www.openindustry.org.br/>

companies to drive sector-wide innovation. The program seeks to create the industry's first data space and support artificial intelligence initiatives for Brazilian industry.⁶¹² However, challenges persist, including fragmented sectoral regulations (e.g. telecommunications under Anatel, finance under BCB) and a digital divide limiting SME participation. Brazil's 2023 Digital Transformation Strategy⁶¹³ aims to address these gaps by expanding broadband access and promoting open APIs in critical industries like mining and energy. While Brazil aligns with OECD data governance principles, enforcement inconsistencies and limited public awareness of data rights hinder progress.

Regulatory and Legal Convergence – Partial convergence

Brazil's LGPD aligns partially with the EU GDPR but exhibits key compliance gaps, including broader legal bases for data processing (10 vs. GDPR's 6), lack of an EU adequacy decision requiring supplemental safeguards (e.g. SCCs) for cross-border transfers, and flexible cookie consent rules that may permit implied consent. Enforcement inconsistencies, lower penalties (capped at ~€8.6M vs. GDPR's 4% global revenue), and sector-specific laws overriding LGPD further challenge full EU alignment. Organisations processing EU data in Brazil must implement GDPR-specific measures (e.g., explicit consent, DPIA) alongside LGPD compliance to avoid non-compliance risks.

Brazil strengthens its data governance by aligning with key global frameworks such as the EU's GDPR (compatible with its LGPD law) to streamline cross-border data flows and trade, and the OECD's principles for responsible data sharing and AI governance. Additionally, its LGPD mirrors EU's GDPR by including specific categories of information under the definition of **sensitive personal data**.⁶¹⁴ Participation in the G20's Data Free Flow with Trust (DFFT)⁶¹⁵ initiative further supports secure international data exchanges while addressing sovereignty concerns, leveraging Brazil's 2024 G20 leadership in digital economy agendas. These alignments supported to enhance compliance for businesses, attract foreign investment, and position Brazil as a bridge between developed and developing nations in global data policy debates.

Additionally, sector-specific partnerships like the IDSA and the EU's GAIA-X infrastructure model supports to advance Brazil's IoT and smart city goals through interoperable, sovereign data ecosystems. Engagement with UN initiatives (e.g., the Global Digital Compact) and regional tools like APEC's Cross-Border Privacy Rules also promotes equitable data access for climate or health collaborations, particularly across Latin America and Asia-Pacific.

Conclusions on lessons learnt and opportunities

Brazil exhibits partial convergence with international data governance frameworks, evidenced by GDPR-inspired LGPD laws and engagement with G20 DFFT principles, though full alignment is hindered by regulatory fragmentation and uneven infrastructure. Emerging potential lies in partnerships like OECD guidelines, IDSA-driven data spaces, and GAIA-X cloud models, particularly in agriculture and smart cities. To deepen

⁶¹² Tiinside (April 2025), *Brazil launches Open Industry program and enters the global Data Economy race*. Available online: <https://tiinside.com.br/en/30/04/2025/Brazil-launches-open-industry-program-and-enters-the-global-data-economy-race/>

⁶¹³ Government of Brazil (2018), *Brazilian Digital Transformation Strategy - E-Digital*. Available online: <https://www.gov.br/mcti/pt-br/centrais-de-conteudo/comunicados-mcti/estrategia-digital-brasileira/digitalstrategy.pdf>

⁶¹⁴ For more information on the treatment of sensitive and highly sensitive data in Brazil see Annex 5.

⁶¹⁵ OECD website, *About > Data free flow with trust*, available at: <https://www.oecd.org/en/about/programmes/data-free-flow-with-trust.html>

convergence, Brazil must address rural digital gaps, harmonise sectoral regulations, and navigate geopolitical tensions. Strategic prioritisation of these areas could position Brazil as a bridge between Global North and South in shaping equitable data governance norms.

Europe could draw valuable lessons from Brazil's evolving data governance and open industry initiatives, particularly its sector-specific collaboration models and public-private data-sharing frameworks. Brazil's Open Industry Platform, which pools anonymised industrial data for sector-wide innovation, demonstrates how targeted data-sharing ecosystems can drive productivity while adhering to strict data protection laws like the LGPD (Brazil's GDPR equivalent). With data spaces created thanks to Open Industry Platform highlights the potential for Europe to adopt similar cross-industry partnerships, combining open APIs, anonymisation, and ethical AI certification (as seen in Brazil's ANPD programme) to balance innovation with privacy.

4.2.1.3 Canada

Introduction

Canada is a global leader in digital innovation and AI development, with a robust legal framework for data privacy and governance. Its data regulatory landscape is shaped by a combination of federal and provincial laws, international partnerships, and strategic initiatives aimed at fostering responsible data use and technological advancement. Key regulatory components include:

- Personal Information Protection and Electronic Documents Act (PIPEDA):⁶¹⁶ Governs private-sector data practices nationwide, except in provinces with “substantially similar” laws (e.g., Alberta, British Columbia and Quebec).
- Privacy Act:⁶¹⁷ Applies to federal government institutions handling personal data. It refers to specific categories of protected data under **personal information**, although only personal data held by federal government institutions is covered.⁶¹⁸
- Canada's Anti-Spam Legislation (CASL):⁶¹⁹ Regulates commercial electronic communications.
- Digital Charter Implementation Act (Bill C-27 from 2022):⁶²⁰ Proposed legislation to modernise PIPEDA and introduce Canada's first AI law (Artificial Intelligence and Data Act (AIDA)⁶²¹). Its future remains uncertain due to political dynamic.⁶²²

⁶¹⁶ Office of the Privacy Commissioner of Canada (May 2024), *PIPEDA requirements in brief*. Available online: https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda_brief/

⁶¹⁷ Office of the Privacy Commissioner of Canada (August 2019), *The Privacy Act in brief*. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-privacy-act/pa_brief/

⁶¹⁸ For more information on the treatment of sensitive and highly sensitive data in Canada see Annex 5.

⁶¹⁹ <https://ised-isde.canada.ca/site/canada-anti-spam-legislation/en>

⁶²⁰ Government of Canada (July 2024), *Canada's anti-spam legislation*, available at: <https://ised-isde.canada.ca/site/innovation-better-canada/en/canadas-digital-charter/bill-summary-digital-charter-implementation-act-2020>

⁶²¹ Government of Canada (January 2025), *The Artificial Intelligence and Data Act (AIDA) – Companion document*, available at: <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>

The identified data access and data sharing measures in Canada largely refer to high-level policy initiatives. Thus, it was agreed with DG CNECT to make short summaries of two of the identified initiatives instead of detailed policy fiches.

- Canadian Sovereign AI Compute Strategy:⁶²³ €1.27 billion investment to boost domestic AI infrastructure, support innovation, and ensure secure data storage within Canada.

The strategy to some extent mirrors EU's AI Continent Action Plan, e.g., emphasising the importance of domestic compute capacity and secure data environments through strategic investment in AI Infrastructure.

- Linkable Open Data Environment (LODE):⁶²⁴ Enhances access to harmonised microdata from various government levels, promoting transparency and civic engagement.

In terms of industrial data governance measures Canada's focus is broadening from personal data to robust governance of non-personal, industrial datasets. One of the key high-level initiatives identified is the Data Governance Standardization Roadmap,⁶²⁵ which establishes national standards covering data collection, organisation, grading, access, sharing, analytics and commercialisation across energy, agriculture, transport and manufacturing use cases. The roadmap put particular emphasis on interoperability, data quality metrics and indigenous data sovereignty. Another example is the Standard on Third-Party Data Access (CAN/DGSI 100-2)⁶²⁶, which provides a trusted framework for secure, auditable sharing of industrial data among public, private and non-profit sectors, ensuring clear roles, consent mechanisms and breach response procedures.

While no single Canadian "data space" programme has been identified several data-sharing initiatives at national and regional or sectoral level exist. Besides the LODE initiative mentioned above, one example is Borealis⁶²⁷ – The Canadian Dataverse Repository – a national, bilingual research data repository for discovery, secure storage and sharing of academic research datasets. It is operational and holds thousands of datasets, hosted in Canada. Another similar case is the Federated Research Data Repository⁶²⁸ of the Digital Research Alliance of Canada, providing a national federated platform for depositing, discovering and preserving research data, while supporting multi-institution workflows. On regional level, several initiatives of provincial open-data portals were identified, providing access to government and municipal datasets for public reuse,

⁶²² Osler (December 2024), *Privacy and data-related law reform to continue in 2025*, available at: <https://www.osler.com/en/insights/reports/2024-legal-outlook/privacy/>

⁶²³ Government of Canada (October 2025), *Canadian Sovereign AI Compute Strategy*, available at: <https://ised-isde.canada.ca/site/ised/en/canadian-sovereign-ai-compute-strategy>

⁶²⁴ Statistics Canada (July 2024), *The Linkable Open Data Environment*, available at: <https://www.statcan.gc.ca/en/lode>

⁶²⁵ Standards Council of Canada (May 2024), *Canadian Data Governance Standardization Roadmap*, available at: https://scc-ccn.ca/system/files/2024-05/scc_data_gov_roadmap_en.pdf

⁶²⁶ Digital Governance Council (July 2025), *Reaffirmed: Canada's National Standard on Third Party Data Access Remains a Trusted Framework*, available at: <https://dgc-cgn.org/reaffirmed-canadas-national-standard-on-third-party-data-access-remains-a-trusted-framework/>

⁶²⁷ Borealis website, <https://borealisdata.ca/>

⁶²⁸ Digital Research Alliance of Canada website, *Federated Research Data Repository*, available at: <https://www.alliancecan.ca/en/services/research-data-management/federated-research-data-repository>

for example: Données Québec, Ontario Data Catalogue, British Columbia Data Catalogue, and Alberta Open Data.⁶²⁹

Key trends identified include:^{630, 631, 632}

- Participation in the Global CBPR Forum to implement international privacy certifications for trusted cross-border data flow.
- Increasing convergence with international standards (e.g., GDPR).
- Stronger enforcement regimes and penalties (e.g., Quebec's Law 25 and proposed revisions of PIPEDA).
- Sector-specific reforms in health, cybersecurity, and children's online safety.
- Focus on open/public data and research/health data with strong governance and Canadian hosting (data residency) for sensitive research datasets.

Regulatory and Legal Convergence – Partial convergence

Canada has maintained adequacy status under the EU's GDPR since 2001, reaffirmed in 2024. This enables seamless data transfers from the EU to Canadian entities governed by PIPEDA.⁶³³

There are several **areas of alignment** when comparing Canada and the EU. Several core principles, such as, transparency, accountability, data minimisation, and consent are shared across GDPR and PIPEDA. Both frameworks also offer individual access and correction rights, although GDPR goes further with portability and erasure. In terms of international transfers Canada ensures safeguards for cross-border data flows, including transparency measures, contractual clauses and binding corporate rules.^{634, 635}

There are also some **limitations to the alignment** just described. For example, the provinces Quebec, Alberta, and BC are governed by “substantially similar” laws as PIPEDA, but these lack EU adequacy status. PIPEDA lacks strong enforcement

⁶²⁹ Various website links: Données Québec - <https://www.donneesquebec.ca/>; Ontario Data Catalogue - <https://data.ontario.ca/>; British Columbia Data Catalogue - <https://www2.gov.bc.ca/gov/content/data/finding-and-sharing>; and Alberta Open Data - <https://open.alberta.ca/opendata>.

⁶³⁰ Government of Canada (March 2025), *Government of Canada launches consultation to implement international data protection and privacy certifications to enable trusted cross-border data flows globally*, available at: <https://www.canada.ca/en/innovation-science-economic-development/news/2025/03/government-of-canada-launches-consultation-to-implement-international-data-protection-and-privacy-certifications-to-enable-trusted-cross-border-dat.html>

⁶³¹ Osler (December 2024), *Privacy and data-related law reform to continue in 2025*, available at: <https://www.osler.com/en/insights/reports/2024-legal-outlook/privacy/>

⁶³² Tokin Manes (February 2025), *What's on our Radar in 2025: Canada's Privacy and AI Landscape*, <https://www.torkin.com/insights/publication/what-s-on-our-radar-in-2025-canada-s-privacy-and-ai-landscape>

⁶³³ Dentons Data (January 2024), *Canada's PIPEDA remains “adequate” under the GDPR: what it means for business*, available at: <https://www.dentonsdata.com/canadas-pipeda-remains-adequate-under-the-gdpr-what-it-means-for-business/>

⁶³⁴ GDPR Local (March 2024), *Aligning Canadian Data Protection with EU Standards: A Comprehensive Guide to GDPR and Canada*, available at: <https://gdprlocal.com/aligning-canadian-data-protection-with-eu-standards-a-comprehensive-guide-to-gdpr-and-canada/>

⁶³⁵ Captain Compliance (May 2024), *PIPEDA Cross Border Transfer: What Are the Rules?*, available at: <https://captaincompliance.com/education/pipeda-cross-border-transfer/>

mechanisms compared to the GDPR. The Bill C-27 (proposed updated data regulation) aims to address this and provide stronger enforcement powers. Finally, regarding automated decision-making, Canada's current laws are less prescriptive than GDPR, while some provincial laws (e.g., Quebec's Law 25) are filling this gap.^{636, 637}

As regards **alignment with international frameworks** Canada is implementing privacy certifications to facilitate trusted cross-border data flows in the context of the Global CBPR System.⁶³⁸ Canada also supports privacy-by-design and child protection in digital environment, promoted by G7 and OECD.⁶³⁹ The proposed AIDA⁶⁴⁰ aligns with the EU's AI Act in regulating high-risk systems and promoting ethical AI.

Conclusions on lessons learnt and opportunities

Overall Canada's degree of convergence with EU data regulations is judged as partial. Canada demonstrates strong foundational alignment with EU and international standards, but gaps in enforcement, provincial consistency, and legislative modernisation (pending Bill C-27 approval) prevent full convergence.

The EU already treats truly anonymised data as outside the scope of the GDPR, but uncertainty persists over what constitutes robust anonymisation in practice. The EU relies on legal tests and guidance to judge whether data are anonymised. In parallel the Data Act and the DGA complement by governing non-personal data sharing and voluntary data-intermediation mechanisms. Canada's proposed federal legislation (Bill C-27, Part 1: Consumer Privacy Protection Act (CPPA)⁶⁴¹) provides a definition of de-identification, where **de-identify** means "*to modify personal information so that an individual cannot be directly identified from it, though a risk of the individual being identified remains*". It also contains a definition of **anonymise**, which means "*to irreversibly and permanently modify personal information, in accordance with generally accepted best practices, to ensure that no individual can be identified from the information, whether directly or indirectly, by any means*". The Bill C-27 proposal also makes it clear that anonymised information is not covered by the CPPA. This clear legal definition of anonymisation would offer a practical complement to the EU's legal test-based approach and could serve as inspiration for the EU in setting technical and administrative standards for anonymisation.

⁶³⁶ Dentons Data (January 2024), Canada's PIPEDA remains "adequate" under the GDPR: what it means for business, available at: <https://www.dentonsdata.com/canadas-pipeda-remains-adequate-under-the-gdpr-what-it-means-for-business/>

⁶³⁷ Tokin Manes (February 2025), What's on our Radar in 2025: Canada's Privacy and AI Landscape, <https://www.torkin.com/insights/publication/what-s-on-our-radar-in-2025-canada-s-privacy-and-ai-landscape>

⁶³⁸ Government of Canada (March 2025), Government of Canada launches consultation to implement international data protection and privacy certifications to enable trusted cross-border data flows globally, available at: <https://www.canada.ca/en/innovation-science-economic-development/news/2025/03/government-of-canada-launches-consultation-to-implement-international-data-protection-and-privacy-certifications-to-enable-trusted-cross-border-dat.html>

⁶³⁹ Office of the Privacy Commissioner of Canada (June 2025), 2025 G7 Data Protection and Privacy Authorities Roundtable Statement, available at: https://www.priv.gc.ca/en/opc-news/speeches-and-statements/2025/js-dc-g7_20250619/

⁶⁴⁰ Government of Canada (January 2025), *The Artificial Intelligence and Data Act (AIDA) – Companion document*, available at: <https://ised-isde.canada.ca/site/innovation-better-canada/en/artificial-intelligence-and-data-act-aida-companion-document>

⁶⁴¹ Parliament of Canada website, *Bill C-27*, available at: <https://www.parl.ca/DocumentViewer/en/44-1/bill/C-27/first-reading>

Analysing the future opportunities, the EU is already collaborating with Canada on digital transformation, through the Canada–EU Digital Partnership,⁶⁴² launched in November 2023, as an official framework for cooperation on connectivity, digital identity, data governance and data spaces, AI, semiconductors, quantum, and cyber security. A recent joint report (May 2025) from the Information and Communications Technology Council of Canada (ICTC) and the AI, Data and Robotics Association (Adra) of EU outlines six collaboration areas.⁶⁴³

1. *AI in business transformation* – e.g., AI testbeds, regulatory sandboxes, and best practice sharing.
2. *Promoting ethical use of AI* – e.g., harmonising regulation, and aligning on privacy protection.
3. *Building AI knowledge and skills* – e.g., AI talent exchanges, reskilling and upskilling, and alignment of curricula.
4. *Just transitions to an AI-enabled economy* – e.g., labour rights and sustainable AI development, and international cooperation on labour standards.
5. *AI innovation hubs* – e.g., leverage Canada's and the EU's ongoing initiatives for joint research on AI-quantum development, cybersecurity, and AI-driven climate solutions.
6. *Public-private partnerships* – e.g., developing co-designed AI policies, AI in infrastructure projects, and public-private AI collaboration models.

The commonalities in AI strategies in combination with the Canada-EU Digital Partnership and the ongoing ICTC-Adra collaboration offer fertile ground for co-developing solutions for e.g., ethical AI, quantum technologies, and climate solutions.

Canada's adoption of Global CBPR Forum certifications should be followed in the context of the overall development and uptake of this initiative, in order to learn how the EU could best take advantage of the mechanisms that are under development.

Canada's evolving data governance landscape reflects a pragmatic balance between innovation and regulation. While legislative modernisation is still in progress, its strategic initiatives and international engagement position it as a valuable partner and reference point for the EU's digital transformation journey.

4.2.1.4 China

Introduction

The People's Republic of China (PRC) has rapidly developed a comprehensive and multi-layered data governance framework, driven by national security, economic modernization, and digital sovereignty. In China, personal information protection and data security are governed through a complex legal framework spread across multiple laws and

⁶⁴² European Commission website (July 2025), *Digital partnerships*, available at: <https://digital-strategy.ec.europa.eu/en/policies/partnerships#:~:text=On%202024%20November%202023%2C%20the,agreed%20to%20work%20together%20on>

⁶⁴³ ICTC-CTIC (May 2025), *EU-Canada Collaboration on AI Could Lead to Transformational Economic Shifts*, available at: <https://ictc-ctic.ca/news-and-events/news-articles/eu-canada-collaboration-on-ai-could-lead-to-transformational-economic-shifts#:~:text=1,international%20cooperation%20on%20labour%20standards>; Full report at: <https://ictc-ctic.ca/media/1473/download>

regulations. The core components of this (still evolving) framework are three national laws:⁶⁴⁴

- The *Cybersecurity Law (CSL)* (2017),⁶⁴⁵ was the first national law to address cybersecurity and data privacy protection;
 - To safeguard network security, maintain national cyberspace sovereignty, and ensure the security of critical information infrastructure and data;
- The *Data Security Law (DSL)* (2021),⁶⁴⁶ addresses data security across a broad category of data (not just personal information);
 - To establish a comprehensive framework for data classification based on the concepts of **important data**, which cover both personal and non-personal information that once misused may endanger national security, and **core state data**, reserved to critical non-personal information for national security and interests.⁶⁴⁷
 - To promote secure data handling, and protect data relevant to national security, economic development, and the public interest;
- The *Personal Information Protection Law (PIPL)* (2021),⁶⁴⁸ the first comprehensive national law on personal information protection – does not replace, but instead enhances and clarifies earlier laws and regulations;
 - To protect individuals' personal information, regulate its processing, and strengthen individual privacy rights, while ensuring responsible use by organizations

They are accompanied by numerous measures and regulations on national, regional and industry level, e.g.: The Measures for the Security Assessment of Outbound Data Transfers (effective from September 2022); The Measures for the Standard Contract for the Outbound Transfer of Personal Information, (effective from June 2023); and The Network Data Security Management Regulation (effective from January 2025). From the international dimension perspective, the *Provisions on Promoting and Regulating the Cross-Border Flow of Data*,⁶⁴⁹ in force since March 2024, is of particular interest, introducing the concept of Free Trade Zones (FTZs) allowing them to develop their own "negative lists" for data export.

Another Chinese initiative was presented by the Study team in a policy fiche (see Annex 4.6), namely *The Action Plan for the Development of Trusted Data Spaces (2024-*

⁶⁴⁴ DLA PIPER (January 2025), *Data protection laws in China*, available at: <https://www.dlapiperdataprotection.com/?c=CN>

⁶⁴⁵ Full text (unofficial translation) of the Cybersecurity Law available at: https://www.dataguidance.com/sites/default/files/en_cybersecurity_law_of_the_peoples_republic_of_china_1.pdf

⁶⁴⁶ Full text of the Data Security Law available at: https://en.spp.gov.cn/2021-06/10/c_948426.htm

⁶⁴⁷ For more information on the treatment of sensitive and highly sensitive data in China see Annex 5.

⁶⁴⁸ Full text of the Personal Information Protection Law available at: https://en.spp.gov.cn/2021-12/29/c_948419.htm

⁶⁴⁹ Full text of the Cross-border data transfer regulation available at: <https://www.chinalawtranslate.com/en/Provisions-on-Promoting-and-Regulating-the-Cross--Border-Flow-of-Data/>

2028).⁶⁵⁰ The plan, aims to establish 100 trusted data spaces of five different types (focusing on Enterprises, Industries, Cities, Individuals and Cross-border), enabling compliant and efficient data sharing while safeguarding personal and national data security. Pilot work has recently been launched, for which the goal is to develop empirical models that can be replicated and promoted nationwide, enabling large-scale circulation and utilisation of data resources and accumulating operational experience for a national data infrastructure. The development is still in an early stage though and no detailed information has been found.

Key trends identified include:

- Strong emphasis on data localization and security assessments
- Introduction of positive and negative lists in FTZs to ease cross-border data flows
- Increasing compliance flexibility for businesses, especially foreign entities
- Strategic use of data regulation to assert digital sovereignty and control information flows

Regulatory and Legal Convergence – Limited/No convergence

PIPL, as well as EU General Data Protection Regulation (GDPR), adopts the OECD's eight basic privacy principles (collection limitation, data quality, purpose specification, use limitation, security safeguards, openness, individual participation and accountability) as a structural template for consent, data minimisation and impact assessments.

At the same time, China represents the most significant divergence from the EU model. PIPL differs significantly from the GDPR, primarily due to its emphasis on national security and state control over data, which distinguishes it from the more individual-centric and rights-based approach of the GDPR.

The GDPR allows data processing based on legitimate interests or contractual necessity, provided strict conditions are met. This offers businesses greater flexibility and supports consistent compliance across EU Member States. In contrast, China's PIPL places strong emphasis on explicit consent and central government approvals, which can hinder cross-border operations, especially due to requirements for state-approved infrastructure and oversight.^{651, 652}

The Chinese state has broad powers to restrict, monitor and intervene in data transfers, especially those deemed "important" or related to critical information infrastructure. The EU GDPR and the Data Act focus on empowering individuals and organizations, ensuring data portability and supporting the free flow of data within the EU while protecting fundamental rights.⁶⁵³

⁶⁵⁰ National Data Administration (NDA) (November 2024), Action Plan for the Development of Trusted Data Spaces (2024-2028), available at:
https://sjj.hubei.gov.cn/bmdt/sjjyw/gj/202411/t20241125_5426763.shtml

⁶⁵¹ DLA PIPER (January 2025), *Data protection laws in China*, available at:
<https://www.dlapiperdataprotection.com/?c=CN>

⁶⁵² IE International Policy Review, Elena Fernández-Novel Escobar, *How do the European Union's GDPR and China's PIPL regulate cross-border data flows?*, available at:
<https://ipr.blogs.ie.edu/2025/01/27/how-do-the-european-unions-gdpr-and-chinas-pipl-regulate-cross-border-data-flows/>

⁶⁵³ IE International Policy Review, Elena Fernández-Novel Escobar, *How do the European Union's GDPR and China's PIPL regulate cross-border data flows?*, available at:
<https://ipr.blogs.ie.edu/2025/01/27/how-do-the-european-unions-gdpr-and-chinas-pipl-regulate-cross-border-data-flows/>

In China non-personal data considered as “important data” and data linked to CIIOs and relevant to national security are tightly controlled in the DSL and CSL, with strict rules for localisation and export. In EU the Data Act and the Free Flow of Non-Personal Data Regulation put emphasis on free flow and fair access, with exceptions for security. They put no localization requirements, except for justified public security reasons.⁶⁵⁴

China’s Action Plan for Trusted Data Spaces aims to establish a national framework for secure, compliant and value-driven data sharing across multiple sectors. Reflecting the legal framework, state control is foreseen also in the developed data spaces, with central coordination, a top-down governance model, uniform national standards and infrastructure, and government-led auditing mechanisms. In contrast, the EU’s Common European Data Spaces initiative promotes sovereign, interoperable, and sector-specific data ecosystems to support innovation and reinforce digital sovereignty across the EU Member States. The EU initiative follows a decentralised, federated governance approach, allowing sectors and member states flexibility within a common EU legal and technical structure.

Conclusions on lessons learnt and opportunities

To conclude, the legal divergence is clear. The differences between EU’s individual rights-based and China’s state-control-based models make cross-border data sharing more complicated. China’s complex legal framework – based on three main pillars, the PIPL, the CSL, and the DSL – imposes strict controls and security assessments on cross-border data transfers, with ambiguous definitions of “important data” creating significant compliance risks for businesses.

While China has signalled a willingness to engage through a “global cross-border data flow cooperation initiative” and a specific communication mechanism with the EU for non-personal data,⁶⁵⁵ the underlying philosophy remains one of tight state control.⁶⁵⁶ Data localisation is identified as a strategic tool which China uses to assert control and protect national interests.

Regarding international data flows, China’s framework is highly restrictive, with cross-border transfers permitted only through pre-approved export controls. It will be important for the EU to monitor China’s progress, particularly in the development of its cross-border data space mechanisms.

In terms of opportunities, the Free Trade Zones may open up for EU business entry under eased compliance. Industry-specific lists offer a model for nuanced regulation. In addition, FTZs and pilot zones may serve as gateways for increased collaboration in innovation initiatives. For example, the upcoming work on cross-border data spaces under the Action Plan for trusted Data Spaces is likely to leverage the development in the FTZs.

EU can influence China’s regulatory evolution through trade and digital diplomacy. While the EU-China Cross-Border Data Flow Communication Mechanism (2024) offers a platform for resolving compliance issues the search for practical bilateral mechanisms

⁶⁵⁴ Tech GDPR (March 2025), *Data protection digest 16 Feb – 2 Mar 2025: Data Act to strengthen EU digital market, vigilance over US data transfers*, available at: <https://techgdp.com/blog/data-protection-digest-4032025-data-act-to-strengthen-eu-digital-market-vigilance-over-us-data-transfers/>

⁶⁵⁵ EU and China launch Cross-Border Data Flow Communication Mechanism - EU Trade, accessed July 10, 2025, https://policy.trade.ec.europa.eu/news/eu-and-china-launch-cross-border-data-flow-communication-mechanism-2024-08-28_en

⁶⁵⁶ Future of privacy forum (February 2025), *Geopolitical fragmentation, the AI race, and global data flows: the new reality*, accessed July 10, 2025, <https://fpf.org/blog/geopolitical-fragmentation-the-ai-race-and-global-data-flows-the-new-reality/>

should continue. In this context the development of model agreements and harmonised templates for data sharing could facilitate collaboration.

China's data regulation regime is strategic, evolving, and increasingly nuanced. For the EU, understanding its logic and engaging proactively offers a path to secure, legal, and mutually beneficial digital cooperation.

4.2.1.5 Japan

Introduction

In 2019, the **Data Free Flow with Trust (DFFT)** concept was launched as a driver “to promote the free flow of data while ensuring trust in privacy, security and intellectual property rights.”⁶⁵⁷ Since then, Japan has promoted a number of initiatives aimed at building a digital infrastructure “suitable for a digital nation in the 21st century”, with special attention to the reinforcement of data flows - both nationally and internationally - and the protection of personal data. In this context, the 2021 **National Data Strategy** was published seeking balance between economic growth, digital transformation towards the Society 5.0, safety and privacy protection.⁶⁵⁸ Furthermore, the Strategy places the government - identified as *platform of platforms* - at its core together with the promotion of public-private cooperation.

Data developments in Japan have systematically addressed three key needs: interoperable and secure data spaces capable of boosting data transfers between different sectors and across borders, ensure the security and reliability of data and reinforce the role of data as promoter of economic as well as technological development. In line with this, the three selected initiatives: the **Act on the Protection of Personal Information**, the **2020 Japan Data Exchange - JDEX** and the **2023 Ouranos Ecosystem** reflect Japan's efforts to fulfil the identified key needs, as they represent the blossoming of both publicly and privately-led data spaces as well as the constant legislative effort to ensure the security and reliability of data, critical for the DFFT concept.

The Act on the Protection of Personal Information (APPI), in force since 2003 and amended in 2015, 2020 and 2021, stands as the primary legislation to govern the collection and processing of personal data by both public and private sectors in Japan. In parallel, the APPI establishes the **Personal Information Commission (PPC)**, as the appointed body to provide guidance and advice on the Act, carry out reports and on-site inspections as well as issue recommendations and orders.⁶⁵⁹

Japan Data Exchange - JDEX is a private-sector-led data marketplace consisting of over 260 entities for data sharing, collaboration and data trading. It provides one-stop support for data registration, search, viewing, marketing, matching, sample acquisition, negotiation and contracts, in addition to being designed for the integration of different public clouds. Currently, it covers non-personal data related to economy, finance and business, real estate, consumer trends, healthcare, vehicle driving and teaching.⁶⁶⁰

The Ouranos Ecosystem represents Japan's forefront in promoting the data exchanges and the interoperability of data infrastructure. More precisely, the ecosystem is based on a

⁶⁵⁷ Digital Agency website, <https://www.digital.go.jp/en/policies/dfft>

⁶⁵⁸ Japan Government (No date), *Presentation-National Data Strategy*, available at https://www.digital.go.jp/assets/contents/node/basic_page/field_ref_resources/0f321c23-517f-439e-9076-5804f0a24b59/20210901_en_05.pdf

⁶⁵⁹ For information on the main regulatory provisions and enforcement mechanisms see Annex 4.7.

⁶⁶⁰ For information on data accessibility and the structure of data governance see Annex 4.8.

series of initiatives launched by the Japanese Government, focused on the design, research, development and promotion of an architecture that enables collaboration and interoperability between various information processing systems operated by different stakeholders. The initiative is the result of an intense public-private cooperation between industry, government and academia.⁶⁶¹

Regulatory and Legal Convergence - Full convergence

Japan's core purpose of developing an environment in which diverse, high-quality, and sufficient data can be used easily, safely, and reliably, together with the DFFT and the Society 5.0 concepts as leading principles, present a unique convergence with the EU in its data strategy.

Both have core data protection legislation: EU's 2016 GDPR and Japan's APPI - last amended in 2021 and currently under revision -, which include similar provisions on sensitive information, extraterritorial scope and obligations for operators and controllers/processors of personal data. Additionally, it includes similar categorisations of sensitive data under its definition of **sensitive personal information**.⁶⁶² Both legislations also detail specific rights for data subjects as well as supervisory authorities and the issuing of financial sanctions.⁶⁶³ As a result, the EU adopted in 2019 the **adequacy decision on Japan** under the data protection adequacy for non-EU countries framework.⁶⁶⁴ The decision - adopted the same year the EU-Japan EPA entered into force - has since been followed by a series of EU-Japan initiatives for digital cooperation and dialogue, crystallising into key milestones including the **2024 EU-Japan deal on data flows** or the annual **EU-Japan Digital Partnership Council** meetings.⁶⁶⁵ Despite the clear shared path to reap the benefits of data economy while ensuring a *human* perspective, Japan still maintains key particularities in its approach towards data.

Japanese legislation on data protection places greater emphasis on ethical behaviour and public perception rather than punishment, following a rather pragmatic and business-friendly approach. Under APPI, a data breach does not always result in immediate penalty, although the PPC is entitled to direct businesses to take measures or implement changes. Furthermore, the APPI does not include specific timeframes for data breach notifications and it is less exhaustive regarding handling personal information and processing data.⁶⁶⁶ The **2025 AI Act**⁶⁶⁷ further deepens this approach and takes a clear

⁶⁶¹ For information on the structure of data governance, data safeguards and geographic scope see Annex 4.9.

⁶⁶² For more information on the treatment of sensitive and highly sensitive data in Japan see Annex 5.

⁶⁶³ Kateifides, Alexis et al. (2020). *Comparing privacy laws: GDPR v. APPI*. OneTrust DataGuidance. Available online: https://www.dataguidance.com/sites/default/files/gdpr_v_appi_april_update.pdf

⁶⁶⁴ European Commission (2019). *Commission Implementing Decision (EU) 2019/419 on the adequate protection of personal data by Japan under the Act of the Protection of Personal Information*. Official Journal of the European Union. Available online: https://eur-lex.europa.eu/eli/dec_impl/2019/419/oj/eng

⁶⁶⁵ European Commission (2024). *EU-Japan deal on data flows enters into force*. European Union. Available online: https://policy.trade.ec.europa.eu/news/eu-japan-deal-data-flows-enters-force-2024-07-01_en

⁶⁶⁶ Lex Dinamica. *International Data Protection Laws (Japan APPI vs GDPR)*. Available online: <https://www.lexdinamica.com/post/international-data-protection-laws-japan-appi-vs-gdpr>

⁶⁶⁷ Officially enacted by the Japanese Parliament on May 28, 2025 as the Act on Promotion of Research and Development and Utilization of Artificial Intelligence-Related Technologies

pro-innovation stance as it primarily focuses on research, development and use of IA, while postponing the establishment of specific requirements or the imposition of a penalties framework.⁶⁶⁸

For its part, Japan's efforts to develop data spaces have increased rapidly over the past years, with a combination of public, private and joint initiatives. In addition to JDEX and the Ouranos Ecosystem, Japan launched the **Data Society Alliance (DSA)** in 2021 to bring together industry, academia and government to create a cross-sectoral data ecosystem by providing technologies and standards for data infrastructure development. Currently, sector-specific initiatives are being developed including the **National Land Transportation Platform** or the **Japan Mobility Data Space (JMDS)** under the **Strategic Innovation Creation Programme (SIP)**.⁶⁶⁹ These initiatives largely follow a data sovereignty principle, which places greater importance on data control by its providers rather than focusing on geographic requirements for data localisation.⁶⁷⁰

Regardless of these particularities, Japan is closely aligned with international efforts to build both trustworthy legal and exchange frameworks for data, taking the lead with the DFFT. The DFFT was introduced by Tokyo during the 2019 World Economic Forum to be later endorsed by the G7 in 2023 and the OECD, which has formed a dedicated **DFFT Experts Community**.⁶⁷¹

Conclusions on lessons learnt and opportunities

Japan demonstrates full convergence with the EU regarding data regulation, with significant potential to deepen cooperation and further alignment in key areas. In its efforts to strive for balance between economic development through data and the public interest, Japan has built a *consensus-driven* approach based on soft regulation. Furthermore, Japan has proven its commitment to international cooperation and to facilitate cross-border data flows as a fundamental part of the data economy. This has been particularly reflected by the promotion of the DFFT concept, the gradual alignment of its data protection framework with EU's GDPR and the different transnational initiatives of which Japan is part or have launched. In this context, a high degree of convergence with the EU in regulation, strategic priorities (AI, e.g.) and joint initiatives continues to be expected.

On top of the assessment presented above, the analysis of the Japan data state of the art provides useful lessons learnt.

Japanese *consensus-driven* approach reflected in the AI Act and in the APPI recognises the challenge that balance between privacy, commercial and public interests poses and will continue to pose. As such, the AI Act has prioritised AI innovation over requirements and penalties while the ongoing review of the APPI could lead to relaxing the breach reporting obligations to the PPC for organisations that have obtained prior certification from recognised third-party bodies as well as allowing organisations to use personal data without consent for the purpose of generative AI education.⁶⁷² As concerned voices rise in

⁶⁶⁸ AI Watch: Global regulatory tracker (2025). *White & Case*. Available online: <https://www.whitecase.com/insight-our-thinking/ai-watch-global-regulatory-tracker-japan>

⁶⁶⁹ Koshizuka, Noboru and Masaru Dobashi (2024). *The next generation data infrastructure in Japan*. International Data Spaces Association. Available online: <https://internationaldataspaces.org/the-next-generation-data-infrastructure-in-japan/>

⁶⁷⁰ For information on data localisation requirements see Annex 4.8 and Annex 4.9.

⁶⁷¹ OECD. *Data Free Flow with Trust*. Available online: <https://www.oecd.org/en/about/programmes/data-free-flow-with-trust.html>

⁶⁷² Nagae, Shunsuke (2024). *Amendments to the Act on the Protection of Personal Information of Japan Proposed for 2025*. Clifford Chance. Available online: <https://www.cliffordchance.com/insights/resources/blogs/talking->

Europe over the restrictive impact of the EU AI Act in R&D, the Japanese approach could contribute to further refine the balance between privacy, data rights and economic interests.⁶⁷³ **As recommendation:** the EU could implement a hybrid approach towards binding rules and penalties, with non-binding mechanisms where companies can opt into guidance and support programs, even before formal legal obligations apply. Additionally, the EU could consider limited exceptions under its privacy regulations to support AI research and training under controlled conditions, mirroring in a limited manner Japan's permissive copyright interpretations in this regard.

4.2.1.6 South Korea

Introduction

South Korea has established a robust data governance framework centred on the Personal Information Protection Act (PIPA),⁶⁷⁴ which mandates strict consent requirements, data minimisation, and enhanced protections for sensitive data such as health or financial information. Recent amendments to PIPA (2023) introduced flexibility for pseudonymised data use in AI and research, reduced consent barriers for contractual data sharing, and expanded data subject rights like portability.

Complementary laws such as the Three Data Acts (2021)⁶⁷⁵ promote innovation through sector-specific frameworks, including tax incentives for data-driven businesses and regulated fintech data sharing via open banking. Institutional bodies like the Personal Information Protection Commission (PIPC) enforce compliance, while initiatives like the MyData empower individuals to control and monetise their data across sectors. Public-private collaboration is further encouraged through platforms like the national Public Data Portal (offering 50,000+ datasets) and regulatory sandboxes (Data-Free Zones) for testing emerging technologies.

The country balances innovation with privacy by aligning cross-border data flows with GDPR standards under the EU-South Korea adequacy agreement (2021),⁶⁷⁶ while maintaining domestic safeguards like encryption mandates and breach notification rules. Challenges persist, including regulatory fragmentation across sectors, "consent fatigue" from stringent requirements, and cybersecurity vulnerabilities in critical infrastructure. South Korea's adaptive governance which seen in its National AI Ethics Framework and blockchain pilots for data traceability, positions it as a global leader in harmonising data utility and protection, though ongoing reforms aim to streamline compliance and address tensions between emerging technologies and privacy rights.

tech/en/articles/2024/06/amendments-to-the-act-on-the-protection-of-personal-information-set-for-2025.html

⁶⁷³ Csernaton, Raluca (2025). *The EU's AI Power Play: Between Deregulation and Innovation*. Carnegie Europe. Available online: <https://carnegieendowment.org/research/2025/05/the-eus-ai-power-play-between-deregulation-and-innovation?lang=en>

⁶⁷⁴ Korea Legislation Research Institute website, *Personal Information Protection Act*, available at: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=53044&lang=ENG

⁶⁷⁵ Konrad Adenauer Stiftung (2021), *Data Innovations and Challenges in South Korea*, available at: https://www.kas.de/documents/288143/19752764/220816_Laenderbericht_Korea_WEB.pdf/7a4ba995-a19b-f359-33bf-81ac14beb8e7?t=1661325969371#:~:text=laws%20to%20promote%20data%20innovation,collectively%20known%20as%20the%20%E2%80%9CThree

⁶⁷⁶ Euclid website (December 2021), *Commission Adopted Adequacy Decision for South Korea*, available at: [https://euclid.eu/news/commission-adopted-adequacy-decision-for-south-korea/#:~:text=On%2017%20December%202021%2C%20the,2%2F2021%2C%2099\).](https://euclid.eu/news/commission-adopted-adequacy-decision-for-south-korea/#:~:text=On%2017%20December%202021%2C%20the,2%2F2021%2C%2099).)

South Korea has emerged as a global leader in open API deployment, driven by collaborative public-private efforts and regulatory support. As of 2024, public institutions have released over 11,860 open APIs through platforms like data.go.kr, providing access to datasets spanning transportation, finance, and climate sectors. In finance, South Korea's open banking API platform which is launched in 2019 under the Financial Services Commission (FSC), has become a model for secure data sharing, enabling 18.6 million users to connect bank accounts with fintech apps via consent-based access. The FSC expanded this framework in 2024 to include insurance and pension data, further integrating financial ecosystems. These efforts are complemented by AI-focused initiatives, where open APIs facilitate training on public datasets and synthetic data experimentation.

During comprehensive mapping phase of this study, two initiatives were carefully analysed: South Korea's MyData initiative demonstrates how empowering individuals with granular control over their personal data (e.g., finance, healthcare) through consent-based platforms can drive innovation while maintaining trust, offering a model for user-centric data governance. Meanwhile, its Open API ecosystem, particularly in finance (open banking) and public services, highlights the value of standardised, secure data-sharing frameworks to foster sectoral collaboration, accelerate digital transformation, and balance regulatory oversight with private-sector agility.

Regulatory and Legal Convergence

The Personal Information Protection Act (PIPA, enacted in 2011 and revised in 2020), shares core principles with the GDPR, including robust protections for personal data, requirements for lawful processing, and accountability measures for organisations.⁶⁷⁷ It also aligns with the GDPR by the categories of information covered by its definition of **sensitive personal data**.⁶⁷⁸ The EU's recognition of PIPA as providing "equivalent" protections which formalised through a 2021 adequacy decision and facilitates seamless data transfers between South Korea and the EU, underscoring PIPA's alignment with global standards.

Key similarities between the PIPA and GDPR include broad definitions of personal data, obligations for data minimisation and purpose limitation, and strong individual rights (e.g. access, correction, and deletion). Both require explicit consent for processing sensitive data, mandate data breach notifications, and impose significant penalties for non-compliance. However, PIPA goes further in some areas: it imposes stricter consent requirements for third-party data sharing and uniquely regulates the use of resident registration numbers (RNNs), which are treated as high-risk identifiers requiring enhanced safeguards. South Korea also enforces specific technical and administrative measures for pseudonymised data, offering a more prescriptive approach compared to the GDPR's principles-based model.

Notable differences lie in enforcement structures and penalties. While the GDPR relies on decentralised enforcement across EU Member States, PIPA is overseen by a centralised authority such as the Personal Information Protection Commission (PIPC). GDPR penalties (up to 4% of global turnover) exceed PIPA's maximum fines (up to ₩500 million or ~€340,000), though South Korea's criminal sanctions for violations, including potential imprisonment, add a unique deterrent.

⁶⁷⁷ Iapp website (January 2018), *GDPR matchup: South Korea's Personal Information Protection Act*, available at: <https://iapp.org/news/a/gdpr-matchup-south-koreas-personal-information-protection-act>

⁶⁷⁸ For more information on the treatment of sensitive and highly sensitive data in South Korea see Annex 5.

Furthermore, the South Korea's participation in the APEC Cross-Border Privacy Rules (CBPR) and adherence to OECD guidelines, including support for the Data Free Flow with Trust (DFFT) initiative, reinforce its commitment to interoperable, accountable data flows. Sector-specific initiatives that were analysed in this study, MyData and Open Banking APIs, align with global trends in data portability, echoing the EU's Open Finance⁶⁷⁹ ambitions. Additionally, South Korea's adoption of the GA4GH Framework⁶⁸⁰ for ethical genomic data sharing and its AI Framework Act (2025), which aligns with the EU's risk-based AI governance, position it as a leader in ethical innovation.

Conclusions on lessons learnt and opportunities

South Korea exhibits partial convergence with global frameworks like the GDPR, evidenced by its GDPR-aligned Personal Information Protection Act (PIPA) and the 2021 EU adequacy decision, which facilitates seamless data transfers. However, sector-specific fragmentation (e.g., financial data laws), stricter consent rules for third-party sharing, and data localisation requirements for public-sector data limit full harmonisation. While initiatives like participation in the APEC CBPR and OECD Data Free Flow with Trust (DFFT) signal intent to align with global trends, challenges persist in balancing domestic sovereignty concerns with international interoperability. Ongoing reforms (e.g., 2025 AI Framework Act) and bilateral collaborations suggest gradual progress toward deeper convergence, contingent on reconciling localised enforcement rigor with flexible global standards.

Europe could draw inspiration from South Korea's integrated approach to balancing data innovation and protection. First, South Korea's sector-specific frameworks (e.g., Three Data Acts) demonstrate how tailored rules for finance, healthcare, and public data can complement broad regulations like the GDPR, enabling targeted innovation while maintaining safeguards. South Korea's MyData Roadmap initiative, which empowers individuals to control and monetise their data across industries, offers a model for Europe to enhance user agency in its own data portability initiatives. Second, South Korea's regulatory agility, which observed in Data-Free Zones (regulatory sandboxes) and sunset clauses for outdated rules, provides lessons for streamlining compliance and fostering experimentation in AI and IoT. Additionally, South Korea's centralised governance through the Korea Data Agency (KData) highlights the value of a unified authority to coordinate fragmented policies, an issue Europe faces with its multi-level enforcement structure. Finally, South Korea's adaptive AI governance (e.g., ethics frameworks and synthetic data pilots) could inform EU efforts to address AI risks without stifling innovation.

4.2.1.7 The United Kingdom

Introduction

The United Kingdom's data governance landscape combines privacy protections with initiatives to unlock data-driven innovation, and it anchored by the UK GDPR and Data Protection Act 2018. A centrepiece of this strategy is the NHS Federated Data Platform (FDP) (analysed Fiche, please see Annex 4.12), a 563 million Euros (£480 million) initiative launched in 2023 to securely connect patient data across hospitals, GP practices, and social care providers. Unlike centralised systems, the FDP uses a federated model, enabling analysis of de-identified data to avoid mass pooling, with strict access controls and encryption to address privacy concerns. The platform aims to improve care

⁶⁷⁹ European Commission website, *Framework for financial data access*, available at: https://finance.ec.europa.eu/digital-finance/framework-financial-data-access_en

⁶⁸⁰ GA4GH website, *Framework for responsible sharing of genomic and health-related data*, available at: <https://www.ga4gh.org/product/framework-for-responsible-sharing-of-genomic-and-health-related-data/>

coordination and medical research while allowing patients to opt out. This fiche reflects the UK's broader push to balance data utility and protection, guided by the National Data Strategy's emphasis on healthcare innovation and the Data Protection and Digital Information Bill (DPDIB), which seeks to streamline responsible data sharing.

Sectoral governance and safeguards are critical, as seen in the FDP's alignment with NHS Digital's Data Security and Protection Toolkit and oversight by the Information Commissioner's Office (ICO). Beyond healthcare, the UK promotes cross-sector data sharing via Smart Data Schemes (e.g., open banking) and international adequacy partnerships. However, the FDP has faced scrutiny over vendor selection (e.g., Palantir's role) and transparency, highlighting tensions between public trust and efficiency goals. The Centre for Data Ethics and Innovation (CDEI) advises on ethical AI use in such projects, while the National Cyber Strategy 2022 reinforces safeguards. The UK's approach—exemplified by the FDP—prioritises federated architectures and strict anonymisation to enable innovation without centralising sensitive data.

After mapping different initiatives within this study, it can be concluded that the UK has established a robust framework for data governance, balancing innovation with stringent protections. Central to this is the Data (Use and Access) Act 2025, which modernises data-sharing rules, mandates ethical oversight, and enables regulated data portability under amended UK GDPR provisions. The Act supports sector-specific "smart data schemes" (e.g., in finance and energy), managed by the Smart Data Council (please refer to analysed Fiche, see Annex 4.13), which oversees secure consumer data sharing with third parties while enforcing transparency and interoperability standards. These legislative and institutional measures aim to foster trust, drive economic growth, and ensure data is used responsibly across sectors.

Lastly, for sensitive data more broadly, the UK emphasizes technical and legal safeguards. The Data (Use and Access) Act requires Data Protection Impact Assessments (DPIAs) for high-risk datasets and enforces penalties for breaches. Initiatives like Trusted Research Environments (TREs)⁶⁸¹ enable secure analysis of sensitive data. These efforts, alongside exploration of data trusts and cross-sector collaboration, reflect the UK's dual priorities: unlocking data's societal and economic potential while maintaining rigorous protections for individuals and sensitive sectors like healthcare.

Regulatory and Legal Convergence – Partial convergence

The UK's NHS Federated Data Platform shows partial convergence with the EU's European Health Data Space (EHDS), sharing principles like GDPR-aligned security, interoperable standards (e.g., FHIR APIs), and ethical data reuse for research.⁶⁸² Furthermore, UK's GDPR includes similar categorisations of personal data under the concept of **sensitive personal data**.⁶⁸³ However, post-Brexit exclusion from EHDS governance limits direct integration, and the platform's domestic focus contrasts with the EU's cross-border health data ambitions. Similarly, the Smart Council Initiative mirrors EU smart city projects by leveraging GDPR-like privacy impact assessments and open data, but UK-specific reforms (e.g., Data Protection and Digital Information Bill) risk diverging from EU accountability standards, particularly in AI transparency and SME obligation.⁶⁸⁴

⁶⁸¹ Health data research UK website, *Trusted Research Environments*, available at: <https://www.hdruk.ac.uk/access-to-health-data/trusted-research-environments/>

⁶⁸² Reference 13

⁶⁸³ For more information on the treatment of sensitive and highly sensitive data see Annex 5.

⁶⁸⁴ Reference from the fiche

While both initiatives reflect shared foundational values with the EU (e.g., data sovereignty, ethical reuse), the UK's regulatory autonomy and lack of formal participation in EU mechanisms (e.g., EHDS, AI Act) result in partial convergence. Opportunities for deeper alignment exist such as adopting EHDS-like portability standards or harmonising AI governance. Strategic choices around the DPDIB's final provisions and bilateral health/AI partnerships will determine whether the UK sustains interoperability or drifts toward a distinct data governance model.⁶⁸⁵

The UK maintains partial alignment with global data frameworks, balancing post-Brexit regulatory autonomy with strategic partnerships. It upholds UK GDPR and retains an EU adequacy decision, enabling seamless data flows with the EU, though proposed reforms under the Data Protection and Digital Information Bill (DPDIB) risk divergence by relaxing consent and accountability rules. The UK is strongly involved in initiatives like the OECD's Data Free Flow with Trust (DFFT) and G7 Digital Trade Principles, promoting cross-border data interoperability, while its AI Safety Summit 2023 and leadership in ethical AI governance align with global trends. Bilateral agreements, such as the US-UK Data Bridge and CPTPP accession, prioritise frictionless data transfers with Indo-Pacific and transatlantic partners, while sector-specific frameworks (e.g., NHS data sharing, Open Banking) mirror international standards in healthcare and finance.

The UK faces risks to its EU adequacy status due to surveillance laws (e.g., Investigatory Powers Act) and potential regulatory drift from the DPDIB. Divergence from the EU's AI Act and fragmented sectoral rules (e.g., slow adoption of Open Finance) could complicate global interoperability.⁶⁸⁶ However, the UK's role in shaping synthetic data standards and hosting forums like the Global AI Safety Institute positions it to influence global norms. Sustaining adequacy, harmonising reforms, and deepening Indo-Pacific digital partnerships will be critical to reinforcing the UK's role as a bridge between EU rigor and global innovation-driven data governance.

Conclusions on lessons learnt and opportunities

The UK exhibits partial convergence with global data standards, balancing post-Brexit sovereignty with strategic alignment. Core GDPR principles persist through the EU adequacy decision (renewed in 2023), enabling seamless data flows with Europe, while initiatives like the Data Governance Conference Europe 2025 and participation in OECD/G7 frameworks signal commitment to cross-border interoperability. However, domestic reforms such as the Data (Use and Access) Act 2025 (DUAA) risk divergence by relaxing consent rules and SME obligations, potentially straining EU relations. Sector-specific efforts (e.g., the NHS Federated Data Platform) align technically with EU health data standards but lack formal integration with mechanisms like the European Health Data Space (EHDS), reflecting a domestically focused approach.

Europe could draw key lessons from the UK's approach to balancing data innovation and privacy, particularly through initiatives like the NHS Federated Data Platform (FDP). The FDP's federated architecture which enables decentralised, secure analysis of healthcare data without centralising sensitive information offers a model for Europe to scale cross-border data sharing (e.g., in the EU Health Data Space) while maintaining GDPR compliance. The UK's emphasis on strict anonymisation, granular opt-out mechanisms, and encryption standards highlights the importance of building public trust in sensitive data re-use, a critical challenge for EU projects. Additionally, the UK's regulatory agility,

⁶⁸⁵ Colm Peter McGrath (2024), *The state and values of AI governance in UK healthcare*. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK613210/>

⁶⁸⁶ Murat Soncul, Head of Data Protection and Privacy (2025), *Unlocking data sharing across government*. Available online: <https://dataingovernment.blog.gov.uk/2025/04/07/unlocking-data-sharing-across-government/>

seen in reforms like the Data Protection and Digital Information Bill (reducing compliance burdens for pseudonymised data), suggests pathways for Europe to streamline innovation without diluting rights. However, the FDP's controversies over vendor transparency (e.g., Palantir's role) underscore the need for Europe to prioritise ethical oversight and vendor neutrality in public-private partnerships. By adopting federated models, enhancing opt-in frameworks, and integrating ethical AI governance (as guided by bodies like the UK's CDEI), Europe could advance its data economy while addressing fragmentation and trust deficits.

4.2.1.8 The United States

Introduction

Data as a strategic asset. The statement included in the **2019 Federal Data Strategy** - designed as a 10-year vision for the acceleration of data usage by the federal Government - encompasses the importance of data as a core element for the US strategy as a global champion for innovation and as a global power.⁶⁸⁷ Since 2020, the US has accelerated its efforts to play a leading innovation role in the field of generative Artificial Intelligence, fostering a strong ecosystem of tech multinationals - OpenAI, Google, Meta, NVIDIA etc. -, startups, academic institutions and Government initiatives - about \$2.8 billion spent in federal AI and IT R&D for the 2021-2025 period.⁶⁸⁸ In parallel, the US has seen increased efforts to build secure and trusted environments for the use of data and the development of data spaces, although through a less structured path than in the EU. This trend has been reinforced by a strong federal structure, with a strong presence of state legislation, political changes at federal level - namely the beginning of Donald Trump's second term - and the strategic considerations of data.

The understanding of *data as a strategic asset* must be understood not only in R&D and economic development terms, but rather as a key element for US global predominance and strategic competition. In line with this, data has experienced an increasing process of securitisation regardless of the administration in office. This trend particularly affects data linked to sectors deemed critical and adopts a country-based framework to restrict data access and share, rather than a model based on safeguards. In this context, the three selected measures: the **2025 Generative Artificial Intelligence and Open Data: Guidelines and Best Practices**, the **2025 Preventing Access to US Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons Final Rule** and the **2021 Securing the Information and Communications Technology and Services Supply Chain Rule** reflect the identified trend. On the one hand, the US continues to foster data usage, and indirectly the establishment of interoperable data assets and data spaces, for the development of such cutting-edge technology as generative AI. On the other hand, the data securitisation process started by the **2019 Executive Order 13873** has not only been consolidated during both Republican and Democrat administrations but is currently in a phase of expansion and reinforcement.

The Gen. AI and Open Data: Guidelines and Best Practices is an initiative launched by the US Department of Commerce to establish technical guidance on the development of

⁶⁸⁷ United States Government website, *Federal Data Strategy - Overview*, available at: <https://strategy.data.gov/overview/>

⁶⁸⁸ Holohan, Maureen (2025). *Federal AI and IT Research and Development Spending Analysis*. Federal Budget. Available online: <https://federalbudgetiq.com/insights/federal-ai-and-it-research-and-development-spending-analysis/>

open data assets that are machine-understandable and AI-usable by providing a framework for open data publishers at all levels within Commerce and beyond.⁶⁸⁹

The Preventing Access Final Rule aims at prohibiting and restricting US entities from engaging in certain data transactions with specific foreign countries deemed as countries of concern and their nationals.⁶⁹⁰ For its part, the Securing ICTS Supply Chain Final Rule enables the Department of Commerce to effectively review and prohibit ICTS transactions linked to a *foreign adversary*.⁶⁹¹

Regulatory and Legal Convergence - Partial convergence

The EU and the US have both acknowledged the vital importance of data flows for economic development and technological innovation - generative AI as example -, while ensuring a trusted and secure data ecosystem. In this sense, the EU adopted in 2020 the **EU-US Data Privacy Framework** under the data protection adequacy for non-EU countries framework. Under this Framework, which entered into force following legal changes in the US linked to data access by intelligence agencies, US certified organisations participating in the DPF may receive personal data from the EU. Despite this, both countries are following different paths towards data regulation.

Unlike the EU GDPR, the US does not have a comprehensive federal data protection regulation with a unified definition on **sensitive data**, but rather a complex patchwork of federal sector-based rules and state regulations on data privacy, implemented first by California and followed by 19 more states as of 2025. These state regulations do include categorisations of personal information under terms such as **sensitive personal information** or **sensitive data**.⁶⁹² Regarding the sectors regulated federally, these include financial, telecommunications, credit reporting agencies, healthcare as well as driving records, children's online privacy, telemarketing, email marketing, biometrics, and communications privacy laws.⁶⁹³ Additionally, the **Standards for Security Categorization of Federal Information and Information Systems** takes a level-based approach for security categorisation of both personal and non-personal information linked to the potential adverse effect of its mismanagement and misuse.⁶⁹⁴

Similarly, EU AI Act has no equivalent at the federal level in the US, as Artificial Intelligence is at the centre of a legislative struggle between state regulations, initiatives at the federal level to precisely suppress AI regulation at state level and calls from the industry and the private sector to either de-regulate or implement a national strategy.⁶⁹⁵ Considering this legal void, the US has seen a large number of federal initiatives on AI emerge including the **2020 National Artificial Intelligence Initiative Act**, the **2020 AI in Government Act** or the **2023 EO on the "Safe, Secure, and Trustworthy Development**

⁶⁸⁹ For information on specific best practices and data development guidelines see Annex 4.14.

⁶⁹⁰ For information on targeted data and restricted countries see Annex 4.15.

⁶⁹¹ For information on targeted ICTS transactions and restricted countries see Annex 4.16.

⁶⁹² For more information on the treatment of sensitive data by specific US states see Annex 5.

⁶⁹³ DLA PIPER (2025). *Data Protection in the United States*. DLA PIPER. Available online: <https://www.dlapiperdataprotection.com/index.html?t=law&c=US>

⁶⁹⁴ For more information on the treatment of sensitive and highly sensitive data in the US see Annex 5.

⁶⁹⁵ Morgan, David and David Shepardson (2025). *US Senate strikes AI regulation ban from Trump megabill*. Reuters. Available online: <https://www.reuters.com/legal/government/us-senate-strikes-ai-regulation-ban-trump-megabill-2025-07-01/>

and Use of AI”.⁶⁹⁶ This effectively sets a path of divergence from the EU and its regulatory strategy towards data, and adds to the increasing differences between both overall data strategies.

Furthermore, the strategic understanding of data by the US Government has led to a securitisation process based not on applicable individual rights, safeguards and standards, but rather in the restriction or denial of access for countries deemed as *countries of concern* or *foreign adversaries*, i.e., strategic challengers and competitors in the global theatre including China, Russia, Iran, North Korea, Cuba and Venezuela. In line with this, legislation such as the Preventing Access Rule clearly places sensitive personal data under the same considerations of government-related data, classifying data as a national security asset. Additionally, the evolution of the Securing ICTS Supply Chain Rule since the **2019 EO 13873** reinforces the sectorial approach towards regulation, as critical technologies such as VCS for connected vehicles, cybersecurity software or drones (UAS) have been rapidly covered by the Rule or are in the process.⁶⁹⁷ Although there is a shared interest with the EU in regulating ICTS in critical sectors as Directive (EU) 2022/2555 - NIS2 and Regulation (EU) 2022/2554 - DORA reflect, both diverge as the EU is currently prioritising industry standards and foreign investment scrutiny.

Conclusions on lessons learnt and opportunities

The US demonstrates partial convergence with the EU regarding data regulation, with significant potential to both deepen cooperation and alignment and diverge in key areas. As a global actor and power, the US has built its own path towards data, shaped by deregulatory tendencies, especially in terms of comprehensive federal legislation, a strong data and AI R&D ecosystem as well as data securitisation. In this context, an increasing divergence can be identified between the models promoted by the EU and the US, which may pose a risk for cross-border data transfers, industrial cooperation and the promotion of joint standards in critical technologies including Artificial Intelligence. On the contrary, the lack of federal comprehensive regulation on data protection and AI, together with the EU-US DPF and relevant transnational data sharing initiatives such as CATENA X or IDSA, open a window of opportunity for legal and strategic alignment between the EU and the US.

Considering the assessment presented above, the analysis of the US data state of the art provides useful lessons learnt.

Despite the noted lack of federal legislation, US agencies have produced an important number of guidelines and best practices, including the Generative Artificial Intelligence and Open Data: Guidelines and Best Practices. **As recommendation:** the EU could adopt a comprehensive set of guidelines and best practices, based on the concept of the US initiative, for data publishers at the European level. This could act as a general framework to foster data accessibility, quality and common standards for ethical AI innovation and data sharing.

The provisions included in rules such as the Preventing Access and the Securing ICTS Supply Chain not only applies to entities in the US and in the targeted third countries, but also applies to entities in countries different from the targeted ones wishing to conduct any of the activities covered by the rules. This could potentially affect EU firms that are

⁶⁹⁶ Szczepański, Marcin et al. (2024). *United States approach to artificial intelligence*. European Parliamentary Research Service. Available online: [https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/757605/EPRS_ATA\(2024\)757605_EN.pdf](https://www.europarl.europa.eu/RegData/etudes/ATAG/2024/757605/EPRS_ATA(2024)757605_EN.pdf)

⁶⁹⁷ For detailed information on sectors covered by the Securing ICTS Supply Chain Rule see Annex 4.6.

currently involved in any type of covered transaction with firms from one of the targeted countries.⁶⁹⁸ **As recommendation:** Data has indeed become a strategic asset, and as such, the EU must wisely adapt to a growing trend of industrial and technological protectionism if it wishes to maintain its open market strategy. In this sense, the EU should review and reinforce its existing legislation on the protection of critical sectors, industrial standards and foreign investment scrutiny to both ensure its own strategic autonomy and strengthen its position as trading partner of two increasingly confrontational blocs: the US and China.

Finally, it is key to replicate and adapt the data ecosystem built in the US and based on strong synergies between tech *giants*, R&D SMEs, research institutions and strong funding efforts from the federal Government. As illustrative examples, the estimated 2020 public investment of the EU in AI was EUR 1 billion while the US invested EUR 5.1 billion. Similarly, EU private investments in AI and blockchain only accounted for 7% of the global total⁶⁹⁹.

4.2.2 Global analysis – Overview of results from country analyses

4.2.2.1 Introduction

Data has undoubtedly become a central element for international economy, innovation, technological development as well as competitiveness and, as its relevance to commercial interests grows exponentially, so do concerns about the protection of privacy and ethics.

The number and scope of data protection and privacy regulations continue to increase globally, while legislators are beginning to address data governance beyond privacy through initiatives aimed at facilitating data access, portability and reuse. Additionally, the emergence of generative AI and its rapid evolution have resulted in efforts to define and clarify the application of privacy regulations to it, i.e., the desired balance between economic interests and citizens' rights.⁷⁰⁰ This process is being shaped globally by a combination of enforcement, litigation, regulatory guidance and strategy. Precisely, strategy and national interests are equally rising as key factors for an issue with such global relevance and implications as data.

Data as a strategic asset. Technology has become a centrepiece of geopolitical struggle between global powers, and data along with it. As AI grows in relevance and impact, the data in which it relies for its development and training has effectively turned into a strategic asset. Cross-border data flows now face stricter oversight or restrictions under notions such as **digital sovereignty**, with the EU, the US and China implementing regulations to keep sensitive data within their borders or directly restricting access from particular countries.⁷⁰¹ Brazil is characterised by conditional cross-border data transfers under the General Personal Data Protection Act (LGPD), supplemented by sector-driven

⁶⁹⁸ For clarification, see example regarding the automotive sector and the Swedish-based firm Volvo Cars in Annex 4.6.

⁶⁹⁹ Special Committee on Artificial Intelligence in a Digital Age (2022). *REPORT on artificial intelligence in a digital age*. European Parliament. Available online: https://www.europarl.europa.eu/doceo/document/A-9-2022-0088_EN.html

⁷⁰⁰ Kewley, Jonathan and Devika Kornbacher (2025). *Data Privacy Legal Trends 2025*. Clifford Chance. Available online: https://www.cliffordchance.com/insights/thought_leadership/trends/2025/data-privacy-legal-trends.html#3

⁷⁰¹ Esposito, Mark (2025). *AI geopolitics and data centres in the age of technological rivalry*. World Economic Forum. Available online: <https://www.weforum.org/stories/2025/07/ai-geopolitics-data-centres-technological-rivalry/>

localisation in finance and health to address national security, public interest, and regulatory compliance. This reflects a balance between integrating global data protection standards and asserting sovereign control over sensitive data. South Korea, while securing an EU adequacy decision for cross-border data flows, maintains localised storage mandates for sensitive sectors like finance, illustrating its hybrid approach of balancing international interoperability with domestic security. Overall, countries are increasingly concerned that data access from certain third countries and data processing abroad could expose them to unfair economic competition, surveillance and foreign leverage. Technology and data are thus experiencing a gradual *securitisation* process as they intertwine with national security concerns and strategic assessments.

The categorisation of data is experiencing a similar trend. EU's GDPR concept of sensitive personal data has led the international adoption of similar formulas into national legislations, including *personal information*, *sensitive personal data* and *sensitive personal information*. Thus, the EU has become a standard-setter for the protection of personal information, a role which, however, has not achieved – even internally – regarding non-personal information as well as **highly sensitive data**, understood as classified or critical information for national security. Here, the possibility of structured definitions based on specific categories of information encounters a far more extended model: **disclose by default, protect when disclosure could cause harm**. This model, shaped by national particularities and strategic assessments, is articulated throughout a patchwork of sectoral-specific regulations and legal provisions. For highly sensitive data, national considerations also affect its treatment as reflected by the US **Standards for Security Categorization** or the Chinese **core state data** categorisation under its Data Security Law.

As a result, the once-borderless global data cloud is becoming increasingly fragmented and compartmentalised into geostrategic blocks, as different pieces of legislation and isolated governance frameworks multiply without a clear converging force. The table below provides an overview of the situation in the targeted countries compared to the EU, summarising the country analyses presented above.

Table 32. Targeted countries' degree of convergence and compliance with the EU

	Australia	Brazil	Canada	China	Japan	South Korea	United States	United Kingdom
Data protection*	Green	Green	Green	Green	Green	Green	Red	Green
Data localisation	Yellow	Yellow	Yellow	Red	Yellow	Yellow	Yellow	Yellow
Data licensing	Green	Green	Green	Red	Green	Green	Yellow	Green
Adequacy decision (EU)	Red	Red	Green	Red	Green	Green	Green	Green
International interoperability	Green	Yellow	Yellow	Yellow	Green	Green	Yellow	Green

To visualise the degree of convergence and legal compliance between the eight selected countries and the EU, the following colour coding system is applied: Green – Full convergence and compliance; Yellow – Partial convergence; Red – No/limited convergence and non-compliance.

* Data protection is understood as the existence of a comprehensive piece of legislation on protection of personal data or information (e.g. EU GDPR)

In this context, the establishment of transnational common baselines for such disruptive technologies as AI, especially regarding safety, transparency, infrastructure resilience and ethics continues to grow more urgent. So do the development of transnational data spaces and cross-border data flows mechanisms, as a key pillar to unlock the full potential of the data economy globally.

Based on the research carried out in this study and presented in this report, it is possible to distinguish **four principal models** for addressing data governance within the complex

landscape of initiatives, legislation, strategies and geopolitical developments: the market-driven, the state-driven, the rights-driven and the consensus-driven.⁷⁰²

4.2.2.2 The market-driven model - Deregulation as driver for innovation

The market-driven model champions market-led R&D and growth by harnessing private capital while the government acts as enabler by lowering operational barriers and, at pivotal times, boosting public investment.⁷⁰³ Deregulation is therefore seen as a key driver for innovation by lowering the legal *hurdles* associated with data protection, establishing itself as the prevailing narrative in critical areas including AI innovation. The US approach towards data currently embodies the market-driven model, characterised by the lack of comprehensive legislation, a fragmented sector-specific regime and strong synergies between tech giants, R&D SMEs, research institutions and - often overshadowed - public funding efforts. South Korea, despite its data protection framework (PIPA), has increasingly leaned into deregulation to spur AI innovation, exemplified by its 2024 “AI Deregulation Sandbox” and public-private partnerships under the “Digital New Deal”.

The model is not exclusive to the US and is gaining ground in strong regulatory contexts such in the EU, Japan or the UK under the premises of competitiveness, innovation or *smarter regulation*.⁷⁰⁴ In South Korea’s case, this shift reflects a strategic balancing act: loosening data-sharing constraints in sectors like autonomous vehicles and smart cities while maintaining core privacy safeguards to retain public trust.

4.2.2.3 The state-driven model - Data at the service of the state

The state-driven model is based on the subordination of data regulations, including those related to data transfers and data protection, to the interests and strategic assessments of the state. Since the introduction of data localisation requirements in 2017, China has progressively aligned its data strategy towards the state-driven model. Its legal framework imposes strict controls and security assessments on cross-border data transfers, with ambiguous definitions of “important data” creating significant compliance risks for businesses.⁷⁰⁵ In 2024 China launched the **Global Cross-Border Data Flow Cooperation Initiative** and, while it reflected the official willingness to deepen cooperation for cross-border data flows, the underlying line of action remains one of tight state control.

⁷⁰² Kümmerle, Harald and Franz Waldenberger (2024). Japan’s ‘consensual’ variety of digital capitalism and its global relevance. *Asia Pacific Business Review*, 30(3), pp. 596-612. Available online: <https://www.tandfonline.com/doi/full/10.1080/13602381.2024.2320553?scroll=top&needAccess=true>

⁷⁰³ Csernatoni, Raluca (2025). *The EU’s AI Power Play: Between Deregulation and Innovation*. Carnegie Europe. Available online: <https://carnegieendowment.org/research/2025/05/the-eus-ai-power-play-between-deregulation-and-innovation?lang=en>

⁷⁰⁴ Cousin, Rebecca (2024). *Smarter regulation? Data privacy impact of the government’s deregulation agenda*. Slaughter and May. Available online: <https://thelens.slaughterandmay.com/post/102j819/smarter-regulation-data-privacy-impact-of-the-governments-deregulation-agenda>

⁷⁰⁵ Kuner, Christopher and Gabriela Zanfir-Fortuna (2025). *Geopolitical fragmentation, the AI race, and global data flows: the new reality*. Future of Privacy Forum. Available online: <https://fpf.org/blog/geopolitical-fragmentation-the-ai-race-and-global-data-flows-the-new-reality/>

Although diluted into the market and the rights-driven models, the understanding of *data as a strategic asset* has expanded the influence of the state-driven model within global actors such as the EU and the US. Data localisation requirements and screening mechanisms have been joined by restrictive legislation in the US, blocking the access to critical data of third countries under categorisations such as *foreign adversary* or *country of concern*.

4.2.2.4 The rights-driven model - Data at the service of the individual

The rights-driven model places individuals' fundamental rights - including privacy, non-discrimination, integrity and autonomy - at the centre of data usage and innovation, effectively subordinating the economic domain to the protection of rights.⁷⁰⁶ Since the GDPR adoption in 2016 the EU has become the pioneer in data governance, with an unparalleled regulatory development which has effectively projected Brussels' normative power and set international benchmarks. Together with the EU, countries such as South Korea, Canada or the UK have followed similar paths, the former with a privacy-based and consent-based data protection law that predates the GDPR and the latter inheriting EU's regulatory developments after Brexit.

Brazil has increasingly adopted state-driven elements, mandating sector-specific data localisation for finance and healthcare while expanding its 2023 National Data Protection Authority (ANPD) powers to block transfers deemed risky to national interests. South Korea, enforces strict data export scrutiny in sensitive sectors like finance and healthcare, exemplified by 2025 enforcement actions requiring foreign companies to justify cross-border transfers. The UK, post-Brexit, has introduced sovereignty-focused amendments to its Data Protection and Digital Information Bill, enabling ministers to restrict data flows to "countries of concern" like China, blending market-driven flexibility with state-driven safeguards.

Since the first adequacy decision under EU's data protection adequacy for non-EU countries framework in 2003, Canada, Japan, South Korea, the US and the UK have joined the framework, leading to a sort of **adequacy alliance** of like-minded partners who share a fundamental commitment to high data protection standards.⁷⁰⁷ Despite this, key actors, such as the UK, Japan and even the EU are pivoting from a regulatory role towards a more innovation-focused path with significant potential consequences.⁷⁰⁸

The rapid technological developments around data carried out in the US and China have added significant competitiveness pressure in the right-driven model, with rising calls for deregulation. Yet, deregulation alone is nothing but an oversimplification of a challenge that requires extremely tailored actions combining governance, enablers, coordination and funding.⁷⁰⁹ In such an environment, the consensus-driven model offers an alternative.

⁷⁰⁶ Mitchell, Margaret (2023). *The Pillars of a Rights-Based Approach to AI Development*. TechPolicy.Press. Available online: <https://www.techpolicy.press/the-pillars-of-a-rightsbased-approach-to-ai-development/>

⁷⁰⁷ European Commission (2017). *What rules apply if my organisation transfers data outside the EU*. Available online: https://commission.europa.eu/law/law-topic/data-protection/rules-business-and-organisations/obligations/what-rules-apply-if-my-organisation-transfers-data-outside-eu_en

⁷⁰⁸ Csernatoni, Raluca (2025). *The EU's AI Power Play: Between Deregulation and Innovation*. Carnegie Europe. Available online: <https://carnegieendowment.org/research/2025/05/the-eus-ai-power-play-between-deregulation-and-innovation?lang=en>

⁷⁰⁹ Aloisi, Antonio (2025). *We're criticising GDPR for all the wrong reasons*. The Conversation. Available online: <https://theconversation.com/were-criticising-gdpr-for-all-the-wrong-reasons-256946>

4.2.2.5 The consensus-driven model - An alternative model in between

The consensus-driven model stands as a strive for balance between the protection of data and privacy of the rights-driven model and the flexibility of the market-driven model, carefully tailoring the synergies between privacy concerns and public as well as commercial interests. Japan, where the notion of *consensus* is deeply rooted in its legal and philosophical traditions, has effectively combined a strong data governance framework with soft regulations that prioritise voluntary compliance, trust and oversight over enforcement penalties.⁷¹⁰ Furthermore, the DFFT concept has led to new and less restrictive interpretations of *data sovereignty*, based on ensuring data is governed by regulations of where it is collected or stored, as well as respect for usage restrictions imposed by the data owner.⁷¹¹

At the global stage, the consensus-driven model presents a key alternative for the most pressing challenges on data. For countries like Australia and Brazil, with less developed data governance frameworks, the convergence of strong regulatory structures and flexibility could facilitate their **alignment to standards closer to the rights-based model**. The model also offers an **alternative** for countries that do not wish to choose a side and be dragged into the growing Sino-American competition. For the EU, the model presents a viable path to **accelerate competitiveness without renouncing** its principles and values, i.e., to the protection of rights. Moreover, it can act as a tool to maintain Brussels' role as normative power by leveraging its flexibility, especially in those countries with a notable pushback against what is perceived as *unilateral* EU regulations.⁷¹²

4.2.2.6 Conclusions

The simultaneous rise of the four data models and their inevitable coexistence have brought about latent tensions, increasingly visible and influential. First, there is the opposition between **innovation and protection**, reflected by the contrast between the market-driven deregulation and the rights-driven safeguards. Secondly, the division between **sovereignty and openness** is shaping the increasing distance between the state-driven protectionism and the consensus-driven strive for global data flows built on trust. Finally, the balance between **competitiveness and normative power** continues to define the long-lasting dilemma for key global actors such as the EU: how to remain innovative while ensuring the projection of its values of privacy and rights at internally and abroad.

The market-driven model seems to have proven effective at driving rapid innovation, but it risks eroding trust and the protection of individuals' rights. The state driven model prioritises sovereignty and national interests, not always understood from an individual-based approach, but can hinder cooperation and global data flows. The rights-based model, characterised until now by strong safeguards for individuals, seems to be under

⁷¹⁰ Kümmerle, Harald and Franz Waldenberger (2024). Japan's 'consensual' variety of digital capitalism and its global relevance. *Asia Pacific Business Review*, 30(3), pp. 596-612. Available online: <https://www.tandfonline.com/doi/full/10.1080/13602381.2024.2320553?scroll=top&needAccess=true>

⁷¹¹ Koshizuka, Noboru and Masaru Dobashi (2024). *The next generation data infrastructure in Japan*. International Data Spaces. Available online: <https://internationaldataspaces.org/the-next-generation-data-infrastructure-in-japan/>

⁷¹² Hanbury, Shanna (2024). *Brazil minister rejects EU deforestation law despite postponement plans*. Mongabay. Available online: <https://news.mongabay.com/short-article/2024/10/brazil-minister-rejects-eu-deforestation-law-despite-postponement-plans/>

increasing pressure in a fast-moving context due to its strict rules. Looking ahead, a suitable path for data is unlikely to come from one of these three models alone. Instead, the consensus-driven model emerges as an alternative capable of combining the rights-based approach of actors such as the EU, the flexibility of market-driven systems, and the strategic awareness of state-driven policies, while grounded in trust and consensus-based frameworks.

4.3 Use case reporting

This section provides the descriptions of the six selected use cases representing opportunities for international data sharing and interlinkage with European data ecosystems including the Common European Data Spaces in relevant sectors. The use case reporting includes an introductory part with basic description of e.g., the overall objectives, sector(s), technology focus, data types and stakeholders involved. This is followed by a description of the main identified requirements and challenges, and a completing subsection with conclusions on the value for EU and possible lessons learnt.

Table 33. Selected six use cases with primary sectors and value proposition for the EU

Use Case Title	Primary Sector(s)	International Partners	Primary Value Proposition for EU	Link to EU Data Space(s)	Key Challenge/Underdeveloped Aspect
1. Cross-Border Digital Battery Passport Data Exchange	Manufacturing, Mobility, Circular Economy	Canada, USA, Japan, South Korea	Industrial Leadership & Green Deal Enforcement	Manufacturing-X, Mobility, Green Deal	Lack of a global, interoperable technical and governance framework for compliance.
2. Federated Learning for Rare Disease Genomics	Healthcare, Research & Innovation	South Korea, USA	Public Health & Ethical AI Leadership	European Health Data Space (EHDS)	Nascent governance for multi-jurisdictional federated networks; lack of semantic interoperability.
3. Interoperable Safety Data Exchange for Autonomous Vehicles (AVs)	Mobility, Automotive, Public Safety	Japan, USA	Road Safety & Shaping Global Standards	Mobility Data Space	Trust deficit and competitive hurdles preventing data sharing; fragmented technical standards.
4. Deforestation-Free Agricultural Supply Chain Traceability	Agriculture, Environment, Retail	Brazil, China	Green Deal Enforcement & Sustainable Trade	Green Deal Data Space	Political friction and lack of scalable, inclusive traceability systems in producer countries.
5. Next-Generation PNR Exchange with PETs	Public Security, Aviation	Australia, UK	Fundamental Rights & Future-Proofing Security	Public Administration Data Space	Institutional inertia favouring bulk data transfer; technical complexity of PETs at scale.
6. Federated Digital Twins for Aerospace Supply Chains	Advanced Manufacturing, Aerospace	USA, Canada	Industrial Resilience & Strategic Autonomy	Manufacturing-X	Immense technical and governance complexity for sharing high-value IP across competitors.

Source: Study team, 2025

4.3.1 Use case 1: Cross-Border Digital Battery Passport Data Exchange

4.3.1.1 Overall description and use case rationale

This use case falls within the manufacturing, mobility and circular economy. The overall objective is to have interoperable data exchange for the EU Digital Battery Passport, connecting EU stakeholders with battery value chain actors in Canada, USA, Japan, and South Korea.

The EU Battery Regulation is a landmark piece of legislation under the European Green Deal. From February 2027, it mandates that every industrial and electric vehicle (EV) battery with a capacity over 2 kWh sold in the EU must have a “battery passport

accessible via a QR code”. This requirement applies regardless of the battery’s origin, effectively creating a compulsory global data-sharing system. To comply, international manufacturers and their entire supply chains must provide data covering the battery’s full lifecycle, from raw material sourcing and carbon footprint calculations to recycled content and end-of-life management. This makes transparent, verifiable cross-border data sharing a prerequisite for accessing the EU market.

While the EU has had a lead in many areas of battery digitalisation, especially regarding modelling and analytics, and the aforementioned EU regulation provides the legal impetus, the global technical and governance framework to make this data exchange a reality is still nascent. Key partners in the ecosystem such as Canada and the US are heavily investing in building domestic battery supply chains. Canada has allocated significant funding and policy support under its Critical Minerals Strategy,⁷¹³ while the US has implemented measures like the Inflation Reduction Act (IRA) to boost domestic production and secure supply chains for electric vehicles and batteries.⁷¹⁴ However, Canada and the US face a relative disadvantage in establishing a harmonised, mandatory traceability framework comparable to that of the EU. The US relies on a combination of voluntary standards and region-specific regulations (e.g. tied to IRA incentives), but lacks a unified, cross-border traceability system.⁷¹⁵ Canada, while aligning with international standards has likewise not implemented a mandatory traceability framework equivalent to the EU’s.⁷¹⁶ This **creates a vacuum**. Without a common standard, the risk is that a patchwork of fragmented, costly, and proprietary solutions will emerge, hindering interoperability and undermining the regulation’s goals.

The battery passport is a strong demonstration of **regulatory gravity** - where the EU law by force creates the demand for an entire data exchange ecosystem. **The critical strategic consideration** for the Commission is to move beyond mere enforcement and actively shape the architecture of this nascent ecosystem. The current vacuum in global standards presents a unique opportunity to make EU companies’ preferred architecture to become the global standard. Stakeholder feedback highlights the importance to build upon proactive national solutions such as the German Battery Pass project, launched in 2022,⁷¹⁷ and industry standards defined by Platform Industry 4.0 initiative. If the EU champions an open, federated standard for the passport’s data space, based on established principles from Gaia-X and IDSA, it can make this the path of least resistance for global compliance. By leveraging existing initiatives like Catena-X and Manufacturing-X, that already address automotive supply chain data exchange, the EU can provide a

⁷¹³ Clean Energy Canada (2021), *Turning Talk into Action: Building Canada’s Battery Supply Chain*. Available online: https://cleanenergycanada.org/wp-content/uploads/2021/05/Turning-Talk-into-Action_Building-Canadas-Battery-Supply-Chain.pdf

⁷¹⁴ The National Law Review (2023), *US-Canada Critical Mineral, EV Battery, and Semiconductor Cross-Border Supply Chain Issues*. Available online: <https://natlawreview.com/article/us-canada-critical-mineral-ev-battery-and-semiconductor-cross-border-supply-chain>

⁷¹⁵ Circularise (2025), *IRA vs EU Regulations: What global companies need to know*. Available online: <https://www.circularise.com/blogs/ira-vs-eu-regulations>

⁷¹⁶ Gowling WLG (2024), *Global traceability: Unpacking a rapidly evolving part of the Canadian battery ecosystem*. Available online: <https://gowlingwlg.com/en/insights-resources/articles/2024/global-traceability-canadian-battery-ecosystem>

⁷¹⁷ Federal Ministry for Economic Affairs and Climate Action (2023), *Battery Passport Content Guidance: Achieving compliance with the EU Battery Regulation and increasing sustainability and circularity*. Available online: https://www.acatech.de/wp-content/uploads/2023/04/1_202304_Battery-Passport-Content-Guidance-Version-1.0_Report_1_v2.pdf

ready-made, efficient blueprint. This would transform a regulatory requirement into a powerful vehicle for exporting EU technological and governance norms, ensuring that the global battery value chain becomes interoperable with and aligned to the principles of the Common European Data Spaces.

4.3.1.2 Requirements for enabling the use case

The data required for the passport is extensive and primarily non-personal, though it can touch upon personal data. Key data points include general battery information (manufacturer, model, composition, etc.), performance and durability metrics as well as detailed supply chain due diligence information, such as the origin of raw materials like lithium, cobalt and nickel. It also includes the battery's carbon footprint, recycled content percentages, and information to facilitate repair and recycling. When linked to a specific vehicle and owner, data on the battery's state-of-health could be considered personal data, requiring GDPR compliance.

Another requirement for enabling this use case is adaptation of common data standards. According to the interviewee usage of standardised frameworks such as Asset Administration Shell (ASS), OPC UA, Eclipse Data Space Components and specifically Resource Description Framework (RDF) for interoperability and semantics is pushed by industry in Germany and other several countries. Using a standardised framework which is exempt from language agnosticism will also ensure semantic interoperability so that data exchanged is understood uniformly across borders.

This data ecosystem is vast and global. As such, it is necessary to take into account the main stakeholders involved as well as its roles: this includes regulators (EU Institutions), battery cell and pack manufacturers (e.g., Verkor and ACCin the EU, LG Energy Solution and Samsung SDI in South Korea, Panasonic in Japan, and new ventures in the US and Canada), automotive OEMs (e.g., Volkswagen, Stellantis, Ford, GM), raw material suppliers (e.g., mining companies in Canada and Australia), recycling companies (e.g., Umicore in Belgium, Li-Cycle in Canada), data intermediaries (e.g., Battery Data Alliance from Linux Foundation Energy) and platform providers, and standardisation bodies like the Global Battery Alliance (GBA) or Catena-X, which Canada actively supports. With such stakeholders included in the use case, it is necessary to set up a clear governance model for data infrastructure and passport management, defining the roles and responsibilities for maintenance, updates and dispute resolution.

Harmonisation of regulatory requirements across the regions involved in the use case must focus on a core set of mandatory data attributes such as battery ID, expiration date and state of health, allowing flexibility for national or regional specificities beyond the agreed core standards. On the other hand, another stakeholder also remarked that harmonisation efforts can falter when not all stakeholders are engaged/involved. However, interoperability between systems can serve as an alternative, as it creates similar outcomes. During the interview phase, it was also mentioned that one of the requirements for enabling the use case is to ensure technical maturity of standards and solutions before the wide-scale rollout, which would push for market adaptation of open, standards-based solutions rather than proprietary and closed systems.

4.3.1.3 Barriers to the implementation of the use case

During the interview phase, it was concluded that a key barrier for this use case is legal and regulatory differences. Each region has its own laws regarding data privacy, product compliances and sustainability strategy that makes harmonisation among different countries very challenging. Another barrier in terms of legal and regulatory differences is intellectual property (IP) protection, where different approaches to IP rights may hinder data sharing and standardisation environments. Furthermore, companies may be reluctant to share proprietary or sensitive data, fearing loss of competitive advantage. There is also another barrier outlined in the interview phase, that also companies there is not yet

established trust and there is fear of IP leakage or misuse that prevent companies from participating in such initiatives.

As mentioned previously, there is still a lack of standard adaptation. While standards like AAS, OPC UA, RDF and Eclipse Data Space exist, their adaptation is inconsistent across regions and companies. Furthermore, some countries and companies may prefer proprietary systems, which hinders interoperability in data exchange. This leads to another barrier on technical interoperability challenges, such as integration issues which can lead to costly and time-consuming integration efforts. Furthermore, semantic differences were seen as another barrier.

Governance and institutional gaps were also identified as a barrier for this use case. According to the interviewee, there is absence of a clear, cross-border governance framework for managing and maintaining the battery passport infrastructure. There is no clear allocation of roles for long-term maintenance and dispute resolution. Furthermore, according to the interviewee, even though all relevant groups of stakeholders (industry, academia, associations) are involved, SMEs often lack the resources to lead or deeply engage in such initiatives. In particular, SMEs may not see immediate benefits and may resist adoption due to costs or complexity. The business model of such use case is uncertain and open, standard based infrastructures may have unclear or unattractive business models compared to proprietary solutions.

4.3.1.4 Conclusions. The value for the EU: Lessons learnt and opportunities

The Cross-Border Digital Battery Passport Data Exchange Use Case is a direct implementation of the EU Green Deal and the Circular Economy Action Plan, positioning the EU as a leader in sustainable industrial policy. Precisely, it is a foundational application for the upcoming **Manufacturing-X data space**, which aims to create a sovereign data ecosystem for Industry 4.0. It also has strong connections to the Mobility Data Space (through EVs) and **the Green Deal Data Space**.

In this sense, its primary value is cementing the EU's role as a *de facto* global standard-setter for sustainability. By mandating this level of transparency, the EU can foster the creation of more resilient, ethical and environmentally sound battery supply chains, reducing strategic dependencies on single sources and mitigating social and environmental risks associated with raw material extraction. It can equally promote a globally competitive market for battery recycling and second-life applications, creating new economic opportunities within the EU.

Furthermore, Europe has a significant opportunity to position itself as a global leader in setting standards for battery passports, digital twin technologies, and interoperable data spaces. By building upon its previous efforts in battery digitalisation, pioneering these frameworks and solutions, Europe can export its technologies and expertise, strengthen its sustainability credentials, and gain a first-mover advantage in global markets. This leadership also supports the EU's Green Deal and circular economy ambitions, while fostering innovation and digital transformation across the manufacturing and mobility sectors.

Key lessons from implementation highlight that true interoperability is technically complex and requires not just standards, but also widespread adoption and integration into legal and regulatory frameworks. Intellectual property protection remains a major barrier to open data sharing, necessitating robust legal and technical safeguards to build trust among stakeholders. Additionally, sustainable infrastructure often needs public-private collaboration, and special attention must be given to supporting small and SMEs to ensure broad participation.

In summary, Europe's experience demonstrates that harmonising essential data and processes — while allowing for regional flexibility — enables both innovation and compliance. Success depends on ongoing governance, strong stakeholder engagement, and balancing openness with the need to protect competitive advantages. These insights will be crucial as Europe seeks to lead in the global rollout of battery passports and similar digital initiatives.

4.3.2 Use case 2: Federated Learning for Rare Disease Genomics

4.3.2.1 Overall description and use case rationale

This use case falls within the healthcare and research & innovation sectors. The overall objective is to establish a federated learning network among leading research hospitals and institutions in the EU, South Korea, and the US, enabling the development of robust AI models for rare disease diagnosis using genomic and clinical data.

Research into rare diseases is critically hampered by the small and geographically dispersed nature of patient populations. To develop effective AI-powered diagnostic tools, researchers need access to large, diverse datasets, but pooling this highly sensitive health data across international borders is often legally and ethically challenging due to strict privacy regulations.⁷¹⁸ To move forward in rare diseases research, enabling legislation such as the GDPR in Europe and HIPAA in the United States, need to be leveraged.⁷¹⁹ Federated learning is a privacy-preserving machine learning technique which offers a revolutionary solution to this standoff. Instead of transferring raw patient data, federated learning allows algorithms to be trained locally within each institution's secure environment. Only model parameters (e.g., weights and gradients) are shared and aggregated centrally, ensuring that patient-level data remains protected. This allows for collaborative model training on a global scale without ever moving or exposing the underlying sensitive data.⁷²⁰

While the technology behind federated learning is maturing, its application in a highly regulated, multi-jurisdictional healthcare setting is still in its infancy⁷²¹. **Significant hurdles remain.** The adoption to available standardised data formats and clinical terminologies (semantic interoperability) between hospitals in different countries is limited, making it difficult to train models on harmonised data. The governance of the federated network - including rules for participation, model ownership, and liability - is complex and requires novel legal and operational frameworks. Finally, there is still legal uncertainty regarding the precise status of the shared model parameters under different data protection laws.

The use case aligns well with EU's strategic R&I priorities under Horizon Europe and the deepening partnership with South Korea. It is well in line with EU strategic priority to advance health data availability and exploitability through the European Health Data Space (EHDS)⁷²² and it also has links to the European Reference Networks (ERNs)⁷²³

⁷¹⁸ Ma, Fenlong and Jiaqi Wang (2023). *Federated learning for rare disease detection: a survey*. OAE Publishing Inc. Available online: <https://www.oaepublish.com/articles/rdodj.2023.16>

⁷¹⁹ Interview feedback.

⁷²⁰ Ibid.

⁷²¹ Ibid.

⁷²² The European Health Data Space (EHDS) website, available at: <https://www.european-health-data-space.com/>

initiative, which connect specialist hospitals across Europe to address rare and complex diseases requiring highly specialised care. Initiatives like Dataspace4Health, a Gaia-X project involving partners from Luxembourg and Japan's NTT DATA, are pioneering this approach but have yet to achieve a truly global, multi-jurisdictional scale.⁷²⁴

The use case represents a fundamental **shift in the concept of data collaboration, moving from "data sharing" to "insight sharing"**. Its successful implementation would provide a powerful counter-narrative to the prevailing assumption that breakthrough innovation requires the creation of massive, centralised data lakes. By demonstrating that value can be unlocked collaboratively without compromising sovereignty or privacy, the EU can establish a model that is both ethically superior and technologically advanced. If this EU-led pilot succeeds, it will prove that large-scale, international AI research can be conducted without creating privacy-invasive global databases. This would establish a **new best practice that could be replicated in other sectors** facing similar data-sharing barriers, such as financial fraud detection or critical infrastructure security. The EU would thus become a leader not just in health AI, but in the entire domain of privacy-preserving distributed computation.

The stakeholder interviews (see list of interviewees in Annex 6) confirmed the main potential opportunities identified for the use case. One stakeholder agrees both to a clinical research opportunity linked to increased access to data for AI modelling, and an opportunity to exploit this use case to pioneer a global methodology for collaborating on federated learning infrastructures, capability and governance that could be reused for many other, less rare, diseases. At the same time, interviewees also stated that the draft description is very broad, indicating the need for further use case definitions. They also pointed out several challenges and barriers for its implementation, presented below.

4.3.2.2 Requirements for enabling the use case

A successful network would involve a multi-stakeholder consortium, representative of the relevant ecosystem and able to agree on rules, controls and safeguards that promote transparency and establish feasible settings for implementation, even under conditions of less-than-optimal legal certainty. Key players include **regulatory oversight providers** (EU and national health authorities), leading **research institutions** and **university hospitals** (e.g., Fraunhofer in Germany, the Luxembourg Institute of Health, and their counterparts in Korea and the US), **patient advocacy groups** (ensuring patient-centricity) and other societal organisations, **pharmaceutical companies** specialising in rare diseases (e.g., Alexion/AstraZeneca⁷²⁵) and **technology providers** who would develop and manage the federated learning platforms.

Feedback from the engagement with stakeholders identified EHDEN⁷²⁶, an IMI (Innovative Medicines Initiative) project, as an important generic health data network supporting

⁷²³ European Commission website, *European Reference Networks*, available at: https://health.ec.europa.eu/rare-diseases-and-european-reference-networks/european-reference-networks_en

⁷²⁴ Gaia-X (2025). *Gaia-X and Partners to Showcase Cross-Border Data Collaboration at Hannover Messe 2025*. Available online: <https://gaia-x.eu/gaia-x-and-partners-to-showcase-cross-border-data-collaboration-at-hannover-messe-2025/>

⁷²⁵ Pharmaceutical Technology (2025). *JCR Pharma and Alexion partners on JUST-AAV gene therapy platform*. Available online: <https://www.pharmaceutical-technology.com/news/jcr-pharmaceuticals-alexion-just-aav-gene-therapy-platform/?cf-view>

⁷²⁶ The European Health Data & Evidence Network, <https://www.ehden.eu/>

federated analysis within Europe. DARWIN EU⁷²⁷, run by the European Medicines Agency, is applying a similar approach and can be seen like a pilot for the European Health Data Space functionality. This use case would benefit from methodological knowledge, tools, experiences and good practices from such existing health data ecosystems.

This use case is essentially defined by what is *not* shared. The sensitive personal data - including genomic sequences, electronic health records (EHRs) and medical imagery - remains securely within the institutional firewalls of the participating hospitals in the EU, South Korea and the US. The only data that crosses borders are the mathematical model updates (weights and gradients), which represent learned patterns but not individual patient information. This fundamentally changes the nature of cross-border data collaboration. However, feedback from stakeholders showed that even if the data shared will primarily be aggregated insights, the **AI model and machine learning processes that are being executed by one country on the data sets of another will need to be verified with a certification process** in order to be accepted by the stakeholders in the different jurisdictions.

In this context, interviews revealed that it is important for stakeholders across all involved jurisdictions to **understand the diversity of federated machine learning methodologies and tools** that are used, and to learn how to determine which methods are suitable and how best to govern them within different infrastructures or research contexts. Data access agreements may need to specify the types of machine learning permitted, the datasets to which they apply, and the safeguards required. These details are becoming increasingly significant. Another interviewee added that acceptance of the AI model is strongly linked to trust, which can be enhanced by clear rules and increased transparency, e.g., regarding what kind of models and analyses are being applied, with whom is data and/or results being shared and for which purpose. Another important issue linked to the trust is the current dynamic and geopolitical situation in different non-EU jurisdictions.

One of the interviewees stressed that the **most important requirements** are not linked to technology and infrastructure and connecting repositories into a federated machine learning environment, but to the **availability, quality, and interoperability of the multiple patient level data sources** that would need to be mapped and inter-connected for a reliable and trustworthy machine learning without the risk of bias. Examples of challenges linked to this aspect are the large numbers of healthcare organisations with diverse electronic health record systems and the variable quality of the data sets. Particularly the legacy data on rare disease patients is expected to be relatively disordered and complicated to interconnect. In this regard, standards that cover the different data modalities needed for rare disease research are considered key to support centres for rare disease to structure and map their data. They may need to leverage NLP (possibly AI powered) to extract structured concepts from narrative letters and reports. Data quality transparency is essential, requiring data quality assessment and data labelling. Metadata will be needed to guide the AI to correctly utilise data sources for different learning objectives, and to have bias transparency as well.

The same stakeholder points out that rare disease research combines clinical data, such as wellness and lifestyle data, and family history data with genetic data and sometimes other modalities such as medical images, photographs. **Interoperability issues and the lack of robust data standards for such data are longstanding concerns.** Standards for interoperability have yet to be widely embraced in electronic health record systems, and regular interoperable data exchange is still uncommon across healthcare systems. So

⁷²⁷ Data Analysis and Real World Interrogation Network, <https://www.darwin-eu.org/index.php>

far, the necessary incentives and motivations have not been sufficiently prioritised. The rare disease research set out in this use case, utilising federated machine learning, **does not create new semantic interoperability problems beyond those already present**, but instead serves as a useful and powerful driver to finally address these existing issues.

Regarding the **governance or institutional frameworks** needed to support the use case one interviewee stated that the model to apply should be a **one stop shop** for the researcher in all Member States and third countries to access the data. This calls for an **intermediary institution** that would be providing an oversight of the available data etc.

The legal aspects linked to sensitive data, privacy and transfers to non-EU jurisdictions are seen as important challenges by the interviewed stakeholders. In this respect one of them comment that **fresh approaches are required to define appropriate safeguards and standards for data protection** – whether under the GDPR or equivalent laws in other countries – that can be applied to federated machine learning involving small patient groups and highly sensitive information, such as genetic data or medical images. Privacy preserving federated machine learning is becoming more mature and better accepted. Thus, the methods developed in relevant European Commission funded projects may provide potential solutions that could be adopted or adapted.

4.3.2.3 Barriers to the implementation of the use case

According to the feedback received **the main barriers for implementation of the use case will be legal, centred on data protection**, as decision makers lack experience in applying existing legislation to these new types of processing. While frameworks such as the GDPR may not require amendment, clear guidance – for example through European Data Protection Board (EDPB) position papers – is essential to ensure consistency in decision making. Although some datasets will remain commercially owned and subject to restrictions, much rare disease data is likely to be accessible, since aggregated extracts are generally sufficient. In addition, rare diseases patients, and their families, are normally more supportive than the population in general to share their health data for research purposes.

The initial use case description, shared with the interviewees, identified it as a potential international flagship use case for the European Health Data Space (EHDS). The EHDS aims to facilitate the use of health data for research and innovation, and a federated learning network could be seen as a prime example of how this could be achieved in a secure and rights-respecting manner. However, while the feedback from the interviews recognises the EHDS as the most suitable framework, sharing communalities with the use case, implementing the use case under the EHDS legal framework is not feasible for reasons including:

- The EHDS secondary use provisions primarily focus on greater European availability of patient level data sets, usually anonymised or pseudonymised. According to one interviewee, the timeline currently foreseen for granting **third-country access to the EHDS is 10 years after the data space is implemented in Europe**. EHDS international data transfers main rules are still to be defined in the delegated and implementing acts. This makes the use case connection to EHDS interesting primarily in a long-term perspective.
- Current EHDS developments indicate resistance from some of the EU Member States to share highly sensitive data, e.g., genetic data from biobanks, even in a European data space implementing a federated data model. One interviewee means that EHDS might make the international data flows even more complex, because especially for the highly sensitive data, it gives more power to the Member States to decide on additional safeguards.

An article by the European Data Protection Supervisor highlights the pros and cons with federated learning from a privacy perspective.⁷²⁸ The international corporation, decentralisation, and data minimization, are identified as the positive impact of such a use case from a data protection perspective and on the other hand, the interpretability, the fairness and security issues are identified as negative risks. According to the same article **open issues under investigation** include communication efficiency, device heterogeneity, data heterogeneity and data privacy.

Stakeholder feedback suggests that implementation of the use would need to happen in a **controlled environment – a regulatory sandbox** – leveraging existing legal instruments such as data sharing agreements and the EHDS principles, while exploring, identifying challenges and proposing solutions that could be transposed to the EHDS data access mechanisms when they will become enforceable.

From a stakeholder onboarding perspective, especially for potential data holders such as hospitals or pharmaceutical firms, it is important to **recognise and incentivise the effort required to make data available**. A clear legal framework is needed to support data sharing, including safeguards for intellectual property and protection of commercially sensitive information, and accompanied by fair compensation. Such contributions should not be taken for granted and must be factored into projects to ensure long-term sustainability.

4.3.2.4 Conclusions. The value for the EU: Lessons learnt and opportunities

The feedback from stakeholders is overall positive, confirming the use case's great potential societal value, by accelerating the diagnosis and development of treatments for rare diseases – a significant area of unmet medical need. On the other hand, based on the feedback received, the Study team discarded the initial proposal to promote this use case as an international flagship use case directly under the EDHS. The current timeline for allowing international data exchange in the data space is a prominent barrier for direct integration under the EHDS. However, collaboration with the EHDS established ecosystem would still be of great strategic importance for such an exploratory use case to succeed, and for ensuring long-term interoperability. Implementing the use case pilots within a regulatory sandbox is considered the most advantageous approach, as it enables the achievement of quicker results while maintaining links to the EHDS framework and ensure long-term validity.

The initiative aligns well with EU's strategic R&I priorities under Horizon Europe and in 2025, South Korea officially joined Pillar II of Horizon Europe, which focuses on global challenges and industrial competitiveness, with health and antimicrobial resistance identified as key areas for cooperation.⁷²⁹ This creates a direct political and funding pathway for such a joint project.

For the EU, it is an opportunity to position itself as a global leader in the development and governance of ethical, privacy-preserving AI for health. It would create a new, powerful paradigm for international medical research that successfully reconciles the drive for innovation with the non-negotiable principles of data sovereignty and patient privacy. The federated model could serve as a template for other sensitive domains—such as cancer

⁷²⁸ European Data Protection Supervisor, European Union, (No date). Available online: https://www.edps.europa.eu/press-publications/publications/techsonar/federated-learning_en

⁷²⁹ European Commission (2025). *International cooperation with Korea in research and innovation*. Available online: https://research-and-innovation.ec.europa.eu/strategy/strategy-research-and-innovation/europe-world/international-cooperation/association-horizon-europe/korea_en

genomics, mental health, and paediatric care – where data sharing is constrained but collaboration is essential.

Despite its promise, the use case faces several implementation challenges.

- *Technical challenges and Legal ambiguity:* Although the idea is to share anonymised data or insights across borders, and not raw patient data, interviewed stakeholders still express uncertainty about the security of federated learning models that are being executed by one jurisdiction on sensitive data sets of another jurisdiction. This also leads to concerns about the necessary legal safeguards and regulatory framework needed for such machine learning processes.
 - Investigations in these areas are under way and a mapping of ongoing relevant projects and research on the state of the art of federated learning would be important activities at the start of the implementation of any use case initiative.
- *Data quality and interoperability:* Rare disease datasets are often incomplete, inconsistently labelled, and stored in legacy systems. Curating these datasets to agreed interoperable standards requires significant effort and investment.
- *Incentives for participation:* Hospitals and other data holders need clear incentives – financial, reputational, or regulatory – to participate. Without sustainable funding and recognition mechanisms, long-term engagement may be difficult to achieve.
- *Trust and transparency:* Stakeholders may hesitate to join federated networks due to concerns about model misuse, lack of visibility into decision-making processes, and geopolitical sensitivities around data collaboration.

Interview feedback confirms that the EU's role in this use case would be multifaceted, encompassing that of regulator, funder, convenor, as well as standard-setter. To address the various challenges and barriers identified a combination of top-down and bottom-up actions are needed. The European Commission and Member States can take a top-down role by shaping policies, promoting harmonisation and standardisation, providing financial incentives, and investing in infrastructure and capacity building. At the same time, they must support bottom-up initiatives through funding schemes and programmes that foster multi-stakeholder collaboration and drive organisational change.

4.3.3 Use case 3: Interoperable Safety Data Exchange for Autonomous Vehicles (AVs)

4.3.3.1 Overall description and use case rationale

Use case 3 is linked to the mobility, automotive and public safety sectors. The overall objective is to enable a federated data space for sharing critical AV safety data (e.g., disengagement events, near-miss scenarios, collision data) between OEMs, suppliers and regulators in the EU, Japan and the US.

The development of safe and reliable autonomous vehicles is a global endeavour, but the learning process is dangerously fragmented. Currently, each AV developer primarily learns from its own on-road testing and failures. This siloed approach means that the entire industry does not benefit when one company encounters a critical edge case or failure mode, forcing others to potentially repeat the same dangerous learning

experiences on public roads.⁷³⁰ A "safety commons," where anonymised data from safety-critical events is shared, would allow the entire ecosystem to learn collectively from every incident. This would dramatically accelerate safety improvements, reduce risks to the public, and build the societal trust necessary for widespread AV adoption.

The exchange of safety data is hampered by a **classic collective action problem**, rooted in a deep-seated tension between the public good of shared safety knowledge and the private value of proprietary data. Automakers are hesitant to share failure data due to concerns about liability and loss of competitive advantage.⁷³¹ There is no mandatory international framework compelling this type of sharing, and technical standards for data formats as well as sharing protocols are fragmented and largely voluntary.⁷³² While countries like Japan have established a legal framework for Level 4 AVs on public roads, the approach to data sharing is not yet harmonised at an international level, preventing the creation of a global safety database.^{733, 734}

The key to unlocking this use case lies not in a technological breakthrough, but in a governance action. A direct, peer-to-peer sharing model between *fierce* competitors is unlikely to succeed. In this context, the EU is uniquely positioned to pioneer a "**trusted third party**" or "**data trustee**" model and to promote regulatory solutions that push such data sharing to take place also on international level. This would involve establishing a data intermediary, that is, a neutral, independent body - perhaps linked to an existing agency such as the European Union Aviation Safety Agency (EASA)⁷³⁵ or a new entity under the Mobility Data Space. Companies would report incident data to this trusted entity, not directly to their competitors or regulators. The trustee would be responsible for anonymising, aggregating, and analysing the data, providing generalised insights and reconstructed test scenarios back to the industry and authorities without exposing the sensitive raw data of any single company. This governance structure, central to the IDSA and Gaia-X philosophy, would build the trust necessary to overcome the competitive and liability hurdles. One stakeholder interview emphasises that the EU needs to redouble its efforts to catch up with China and the US. However, if participation in this system were linked to the EU vehicle type-approval process, making it a requirement for market access, it could establish the EU as the global centre for AV safety data analytics and governance.

4.3.3.2 Requirements for enabling the use case

The focus of this use case is on sharing just enough information to allow for the reconstruction and analysis of safety-critical scenarios without revealing proprietary algorithms or competitive secrets. Data packages would be triggered by specific events (disengagements, near-misses, collisions, etc.) and could include anonymised or pseudonymised sensor data (LIDAR point clouds, camera snippets, etc.) vehicle telemetry (speed, acceleration, steering angle, etc.) and environmental context data (weather, road

⁷³⁰ Goedicke, David et al. (2024). *Mutual Benefit: The Case for Sharing Autonomous Vehicle Data with the Public*. Cornell University. Available online: <https://arxiv.org/html/2409.01342v1>

⁷³¹ Ibid.

⁷³² Connected Automated Driving (2025). *Standards*. Available online: <https://www.connectedautomateddriving.eu/standards/>

⁷³³ Gaia-X. *Data Spaces*. Available online: <https://gaia-x-hub.de/en/campus-en/data-spaces/>

⁷³⁴ GlobalFleet, (Nov 2024), From *lanes to laws: Japan's fast-tracking of a driverless future*, Available online: <https://www.globalfleet.com/en/technology-and-innovation/asia-pacific/analysis/lanes-laws-japans-fast-tracking-driverless-future>

⁷³⁵ The European Union Aviation Safety Agency website, available at: <https://www.easa.europa.eu/en>

conditions, etc.). The principle of proportional disclosure would apply: more data is shared for more severe events such as collisions.⁷³⁶

Building on this point, one interviewed expert underlined the importance of federation across multiple databases and scenario datasets. This includes **technical enablers** such as:

- Standardised connectors and interfaces,
- Semantic data models,
- Identity and access management systems,
- Traceability and auditability tools.

As a B2B (Business-to-Business) and B2G (Business-to-Government) use case, it aims to involve **automotive OEMs** (e.g., Volkswagen, Toyota, GM, Ford), **Tier 1 suppliers** (e.g., Bosch, Continental), **oversight bodies** (EU and national transport authorities and regulators), **standardisation bodies** (e.g., ISO, SAE - which are working on AV standards but without a harmonised global framework⁷³⁷) and **data space operators** such as the Mobility Data Space (MDS) in Germany⁷³⁸. Additionally, as one stakeholder notes, beyond manufacturers and regulators, national bodies such as the Ministry of Justice, forensic institutes, police, and other agencies involved in accident response should also be considered as stakeholders.

The main conclusion of the stakeholder interview is the **critical importance of governance and trust framework**, envisioning a neutral European-level entity. One interviewee envisioned it under the CCAM Partnership⁷³⁹ or under the Data Spaces Support Centre⁷⁴⁰ to ensure neutrality and acting as an intermediary between data providers (OEMs, Tier-1s, AV tech companies) and data users (regulators, researchers, policymakers), securing confidentiality, trust, and fairness in throughout the entire process of data sharing, management and usage. To guarantee the protection of confidentiality and commercial interests of all participants, the intermediary entity should rely on EU laws clarifying data usage rules, non-disclosure agreements and limiting the market access for actors that lack offering verifiable information on safety management and the handling of edge cases.

To ensure trust and **traceability**, the following aspects should be incorporated:

- Contractual and legal mechanisms for data access and usage rights,
- Data licensing models,
- Anonymisation and IP protection tools,
- Compliance with the Data Act and AI Act provisions.

While it can be a potential incentive for importers outside the EU, it is worth to note that the attitude and willingness to cooperate also vary among the manufacturers within the

⁷³⁶ Goedicke, David et al. (2024). *Mutual Benefit: The Case for Sharing Autonomous Vehicle Data with the Public*. Cornell University. Available online: <https://arxiv.org/html/2409.01342v1>

⁷³⁷ Dunn, Jessica (2025). *The Future of US Battery Manufacturing is At Stake*. Union of Concerned Scientists. Available online: <https://blog.ucs.org/jessica-dunn/the-future-of-us-battery-manufacturing-is-at-stake/>

⁷³⁸ Fanoukoe, Samuel. *Data Spaces: Unlocking the Power of Connected Information*. Dataroots. Available online: <https://dataroots.io/blog/data-space>

⁷³⁹ CCAM website, *CCAM Partnership*, available at: <https://www.ccam.eu/what-is-ccam/ccam-partnership/> CCAM website, *CCAM Partnership*, available at: <https://www.ccam.eu/what-is-ccam/ccam-partnership/> <https://www.ccam.eu/>

⁷⁴⁰ Data Spaces Support Centre website, available at: <https://dssc.eu/>

EU. Offering some kind of influence on the creation of the future data sharing standard might be an efficient incentive for manufacturers. It could be an effective way to convince companies they get some value out of sharing more data. One interviewee also acknowledged the importance of **economic sustainability** of such data space as its infrastructure and operation invoke costs. A viable business model is needed to support data sharing and generation, along with appropriate incentives to encourage data contribution and reuse. As another interviewee points out, the existing initiatives and regulations focus on potential flows, but ditching the negative approach and focusing on accelerating and enabling innovation might enable a more cooperative attitude from the manufacturers.

4.3.3.3 Barriers to the implementation of the use case

The stakeholder feedback highlights that the absence of a global governance or certification mechanism is the key barrier to the implementation. In detail, the core barriers are governance-related, legal and geopolitical, leaving technical aspects as secondary barriers. The competitive dynamics between global OEMs, different liability regimes and fragmented technical standards lacking harmonisation in terms of definition, categorisation and methods on what qualifies as a “safety-critical event”, what data should be shared and how it should be anonymised or abstracted.

The base of harmonisation lies in the engagement of the government. As one stakeholder underlined, the governance strategy of China represents another alternative, establishing a highly centralised, state-controlled AV data regulation, which is managed by public authorities that show limited consideration for transparency. As another interviewee adds, the EU on the other hand is characterized by slow progress and fragmentation, as such data control is regulated and executed on a national level. The relations between each national governmental agencies and manufacturers are very variable, but even on a Member State level there are not many examples of proactive initiatives.

On the other hand, the US provides an example of a company-driven data management, but these practices tend to be non-transparent because companies are inclined to operate in silos. The US proves that minimal regulations enable strong corporate control, giving rise to disengagement and reluctance to share data, as AV developers operate in a highly competitive and litigious environment. These tendencies could be balanced by a national mandate or trusted intermediary for aggregating or anonymising AV safety data across companies so the AV developers would have a guarantee to preserve their liability and competitive advantage.

Shifting the focus to the global scale, the example of Japan highlights that even a robust legal framework for Level 4 AV deployment would be insufficient to be the norm-setter unless it has the incentivisation mechanisms and operational framework to coordinate with international counterparts on cross-border AV safety data exchange. Without policy alignment across jurisdictions, companies will continue to operate in silos. The lack of incentives or obligations to share data, combined with the absence of legal protection and trust, remains the biggest roadblock to implementation. To guarantee proper representation of European standards, the respondent expert brings up the example of EASA and the ongoing negotiations with its US counterpart and suggests the establishment of an EU agency specifically dedicated to AV data management, as at present it can be only negotiated on a country level, potentially leading to imbalanced dynamics at the conference table, especially with partners who are global leaders in the automotive sector.

4.3.3.4 Conclusions. The value for the EU: Lessons learnt and opportunities

The Interoperable Safety Data Exchange for AVs use case supports EU’s “Vision Zero” goal of eliminating road fatalities as well as its overarching strategy for connected and

automated mobility. It represents a crucial evolution for the Mobility Data Space, moving beyond consumer-facing services to become a critical safety infrastructure for the automotive industry.⁷⁴¹ This aligns with the vision for Gaia-X to support sovereign, federated data ecosystems for key European industries.⁷⁴²

The primary value of this use case is the enhancement of road safety for all citizens. It also accelerates the safe deployment of AVs, a sector of intense global industrial competition. Besides, as one stakeholder concludes, such collaboration and the subsequent technological advantage would also accelerate reducing emissions, improving traffic flow, and contributing to more sustainable mobility systems which can effectively support the current societal and sustainability goals of the EU.

As some interviewee notes while taking the lead in creating a trusted, non-punitive environment for this type of data sharing would offer advantages and potential influence, it would take significant efforts, as EU has already fallen behind China and the US. However, the aforementioned trust-based approach would help to reposition the EU as a global convener and rule-maker in a transformative technology domain, offering a third way between China's highly centralised, state-controlled AV data regulation and Silicon Valley's company-driven and non-transparent ecosystem. A well-functioning governance and trust framework, supported by EU law, to safeguard commercial interests and confidentiality is considered key for success. This would be strongly linked to the establishment of a neutral European-level intermediary to collect, anonymise, and analyse safety incidents on behalf of all industry players. Industry participation would be crucial for the development of contractual and legal mechanisms for data usage, access rights, licensing, and IP protection. As potential allies, one interviewee names South Korea and Japan, noting that both countries already maintain substantial government agencies dedicated to AV data management, demonstrating not only a robust toolkit but also a strong commitment. Furthermore, the EU could investigate the possibility to connect participation in the AV safety data exchange to the EU vehicle type-approval process, making data contribution a market-access requirement.

4.3.4 Use case 4: Deforestation-Free Agricultural Supply Chain Traceability

4.3.4.1 Overall description and use case rationale

The Deforestation-Free Agricultural Supply Chain Traceability Use Case is linked to the agriculture, retail, environment and finance sectors. The overall use case aim is to enable a cross-border data space for the traceability of key agricultural commodities to ensure compliance with the EU Deforestation Regulation (EUDR). This would connect EU importers with producers and processors in key sourcing countries such as Brazil and major processing hubs such as China.

The EUDR, which entered into force in June 2023,⁷⁴³ is an ambitious piece of environmental legislation. It requires any company placing covered commodities (cattle, cocoa, coffee, palm oil, rubber, soy and wood) on the EU market to prove that their

⁷⁴¹ Gaia-X (2022). *How can dataspaces contribute to the development of mobility activities in Europe and better serve travellers*. Available online: https://gaia-x.eu/wp-content/uploads/2022/06/Mobility-Data-Space-Report_FINAL.pdf

⁷⁴² Gaia-X. Data Spaces. Available online: <https://gaia-x-hub.de/en/campus-en/data-spaces/>

⁷⁴³ Originally the EUDR was supposed to take full effect in December 2025, but in September 2025 the European Commission has proposed a delay to December 2026.

products are not linked to land deforested after December 31, 2020.⁷⁴⁴ Besides following relevant legislation of the producer country, compliance necessitates the submission of a due diligence statement containing precise, plot-level geolocation coordinates from the farm of origin.⁷⁴⁵ This creates a non-negotiable, legally mandated requirement for granular data flow from thousands of farms in different countries, through complex global supply chains involving traders and processors, to the desks of EU importers and customs authorities.

In such a context, the Deforestation-Free Agricultural Supply Chain Traceability Use Case faces **significant political and practical headwinds**. There is a strong political pushback from governments of major producer countries, notably Brazil, which have criticised the EUDR as a unilateral, punitive and discriminatory trade barrier that infringes on their sovereignty.⁷⁴⁶ This has led to diplomatic friction and calls for the regulation's postponement. Even within Brazil, there are deep divisions between agribusiness lobby and environmental authorities.⁷⁴⁷ On a technical level, creating affordable, scalable and reliable traceability systems represents a challenge even for high-tech level organisations, which is even more serious when it comes to the accommodation of millions of small, often low-tech farmers.⁷⁴⁸ The added complexity of navigating China's restrictive data export laws for commodities processed there further complicates the picture.⁷⁴⁹

The EUDR is effectively creating a new global market for "compliance data". The EU's **strategic opportunity** lies in shaping this market to be fair, inclusive and collaborative, rather than purely punitive. Instead of acting solely as a regulator imposing rules from afar, the EU can also act as an *enabler*. This involves moving beyond a confrontational stance and building partnerships by bringing both pro-sustainability and non-pro-sustainability actors to the table within producer countries, such as Brazil's IBAMA and progressive farmer associations. By using instruments like development aid or Horizon Europe funding, the EU could support the co-development of open-source, low-cost traceability platforms tailored to the needs of smallholders. Interview feedback highlighted the importance of building upon already existing and trusted initiatives and platforms in Brazil and EU. The most efficient way for expansion will be through peer-to-peer recommendation. This proactive engagement would counter accusations of "green protectionism," demonstrate goodwill, and build a more resilient and cooperative ecosystem for EUDR compliance. In sum, it would transform a point of diplomatic friction

⁷⁴⁴ CMS Law-Now (2025). *Guidance: Geolocation Data Sharing from China under EUDR*. Available online: <https://cms-lawnow.com/en/ealerts/2025/07/guidance-geolocation-data-sharing-from-china-under-eudr>

⁷⁴⁵ Ibid.

⁷⁴⁶ Hanbury, Shanna (2024). *Brazil minister rejects EU deforestation law despite postponement plans*. Mongabay. Available online: <https://news.mongabay.com/short-article/2024/10/brazil-minister-rejects-eu-deforestation-law-despite-postponement-plans/>

⁷⁴⁷ Mendes, Karla (2025). *EUDR divides Brazil's environmental and agribusiness authorities*. Mongabay. Available online: <https://news.mongabay.com/2025/01/eudr-divides-brazils-environmental-and-agribusiness-authorities/>

⁷⁴⁸ Igini, Martina (2024). *Brazil Urges EU to Postpone and Reassess 'Unilateral' Anti-Deforestation Law Over Fears It Will Affect Trade Relations*. Earth · ORG. Available online: <https://earth.org/brazil-urges-eu-to-postpone-and-reassess-unilateral-anti-deforestation-law-over-fears-it-will-affect-trade-relations/>

⁷⁴⁹ CMS Law-Now (2025). *Guidance: Geolocation Data Sharing from China under EUDR*. Available online: <https://cms-lawnow.com/en/ealerts/2025/07/guidance-geolocation-data-sharing-from-china-under-eudr>

into a genuine partnership for sustainable trade, strengthening the hand of those in producer countries who advocate for greater environmental protection.

4.3.4.2 Requirements for enabling the use case

The primary processed data is non-personal, including geolocation coordinates of farm plots, transaction records, transport documents, and certificates of compliance. However, the situation is nuanced. In supply chains involving many smallholder farmers, this plot-level data, when linked to an individual or family, can be classified as personal information. This triggers the data protection regimes of both the source country (e.g., Brazil's LGPD) and transit/processing countries (e.g., China's PIPL), adding significant legal complexity to the data sharing process.⁷⁵⁰ However, one stakeholder pointed out that there are already existing technological solutions that allow data to be anonymized to a certain extent when uploaded into the system. Additionally, regulations like the GDPR provide further protection regarding personal information, such as land ownership details. The stakeholder group is highly complex and global. It includes **regulators** (the European Commission), **EU importers and retailers** (e.g., Danone, which has its own soy traceability project⁷⁵¹), major **commodity traders** (e.g., COFCO International⁷⁵²), **producers in Brazil** (ranging from large, sophisticated agribusinesses to small family farms), **Brazilian government agencies** (which are internally divided, with the Ministry of Agriculture opposing the EUDR and the environmental agency IBAMA supporting it) and **technology providers** offering traceability solutions using blockchain, IoT and satellite imagery. The stakeholder interviews emphasized the importance of a fair and collaborative approach by involving non-EU countries early in the process of developing a partnership supporting the design of a platform for data sharing, especially as interoperability is crucial later on, in the operational phase. Along with interoperability, adequate governance of the data platforms is also essential to ensure credibility, trust, and willingness to participate. The overarching engagement could also support local innovation, create business opportunities, and help implement new regulations effectively. Several stakeholders also acknowledged the efforts by Brazil to strengthen traceability via regulations and technology projects and they have shown progress related to timber and soy which can serve as an enabler of cooperation.

The stakeholder inputs underlined that the effective sharing of supply chain information, along with associated economic, social, and environmental performance metrics depends on the integration of diverse data sources, such as shipment records, inspections, tax data, export licenses, and geographic information. It is also important to recognise that the use case covers a cross-sectorial data value chain with multiple intersections. However, scalable traceability does not necessarily require the development of new hardware and/or digital technologies. Instead, the focus should be on coordinated approach to promote standardised data collection, suggested data governance methods and framework intended to assure data quality along the whole data value chain, and improving access to and use of existing data under “fair, reasonable, and non-discriminatory” (FRAND) terms as prescribed by the recently entered into force Data Act regulation. Much of the necessary information is already being collected but remains siloed across various government agencies or proprietary platforms, with limited

⁷⁵⁰ Ibid.

⁷⁵¹ Danone Regenerative Agriculture (2023). *The Soy Traceability Project*. Available online: <https://regenerative-agriculture.danone.com/projects/the-soy-traceability-project/>

⁷⁵² COFCO International (2020). *COFCO International aims for full traceability of all directly sourced soy across Brazil by 2023*. Available online: <https://www.cofcointernational.com/newsroom/cofco-international-aims-for-full-traceability-of-all-directly-sourced-soy-across-brazil-by-2023/>

coordination or data sharing between them. Unlocking these existing datasets would eliminate the need to create entirely new systems, avoiding duplication of effort and accelerating progress toward effective traceability in agricultural supply chains.

In addition to the Data Act, the recently adopted **AI Act** introduces further regulatory dimensions relevant to traceability systems, particularly those leveraging artificial intelligence for data analysis, risk assessment, or decision-making in supply chains. If AI-based tools are used to process geolocation data, assess deforestation risks, or automate compliance checks, they may fall under the scope of the AI Act as **high-risk systems**. This implies requirements for transparency, robustness, human oversight, and conformity assessments. Stakeholders should therefore consider not only data governance and interoperability, but also **algorithmic accountability** and **compliance with AI-specific obligations**, especially when deploying AI in cross-border traceability platforms. Early alignment with the AI Act can help ensure legal certainty and foster trust among international partners.

As stakeholders noted, embedding available traceability systems within acknowledged and robust regulatory framework could support the overcome of the barriers. In the EU, regulations such as the EU Deforestation Regulation (EUDR), the Corporate Sustainability Reporting Directive (CSRD), and the European Sustainability Reporting Standards (ESRS) require companies to disclose environmental and social impacts across their supply chains, reinforcing the need for reliable and granular traceability mechanisms. The upcoming Ecodesign for Sustainable Products Regulation (ESPR) will introduce Digital Product Passports, enabling companies to communicate verified sustainability data directly to consumers, turning compliance into a strategic branding tool. At global level, frameworks such as the OECD-FAO Guidance for Responsible Agricultural Supply Chains, the UN Guiding Principles on Business and Human Rights, and the CFS Principles for Responsible Investment in Agriculture and Food Systems provide comprehensive guidance for responsible sourcing and due diligence. Also recognized technical standards/certification (ISO22000, ISO140064) support this approach for third party verification needs in the sustainability domain. By aligning traceability systems with these regulatory and normative instruments, companies can not only ensure compliance but also enhance transparency, build consumer trust, and differentiate their products in increasingly sustainability-conscious markets. This approach transforms traceability from a cost centre into a value-generating asset that supports both regulatory alignment and brand positioning.

It is also important to recognize that data gaps are likely to arise throughout both the collection and technical implementation phases. Addressing this will require a robust and standardised methodology for identifying and managing uncertainties or incomplete segments within supply chain data.

One of the key challenges ahead lies in determining exactly which types of data and metadata should be shared, along with establishing consistent formats and structures for doing so, together with the right metadata to understand the data and give value to the data shared even outside the context of origin. Valuable lessons could be drawn from how the life cycle assessment community has approached similar issues, particularly through databases like HESTIA⁷⁵³ or ecoinvent⁷⁵⁴ – and comparable systems which may offer useful models for structuring and standardising complex environmental data.

⁷⁵³ HESTIA website, available at: <https://www.hestia.earth/> HESTIA website, available at: <https://www.hestia.earth/>

⁷⁵⁴ ecoinvent website, available at: <https://ecoinvent.org/> ecoinvent website, available at: <https://ecoinvent.org/>

Feedback from stakeholders further highlight that a potential leadership position is established by taking the lead in making data more available in order to build trust and catalyse data sharing among the EU and its partners such as Brazil which already has a rather open data landscape. In addition to serving as a catalyst, stakeholders also refer to it a cost-effective resource. Stakeholder interviews also note that an EU agency can potentially be an effective governance institute or at least a connecting body where knowledge can be translated into action. This area however falls into the Member States competence, and a high level of data sharing among the authorities of each Member State is necessary to actually be able to trace all of this data, and where companies are importing or exporting. It is worth mentioning that companies export into multiple countries at once, so there probably need to be a high level of data sharing between enforcement and confidence authorities in Member States. Such nuanced cooperation would require a moderator or an intermediary organization commissioned to safeguard the implementation of standards across the EU and manage any conflicts between Member States.

Finally, a critical requirement for ensuring that accessible, scalable, and reliable traceability systems can accommodate millions of small farmers, who often have low technology, is the effective usage of GenAI to enable more robust and reliable actions that are transparent to the users (especially those who are not experts) but at the same time explainable, so to get for example **automatic data validation** by detecting and correcting errors in farmer-related data, **data format normalisation** if different data sets come from diverse sources, **translation of data** entered using different dialect, **accurate predictions** such as risk ones (e.g., unexpected supply chain delays or more accurate crop yields forecasting). In addition, GenAI can pave the way to (semi)automate **key workflows** within such traceability systems, by pre-filling compliance forms with EUDR based on both structured and non-structured traceability data, executing compliance check in documentations against standards, and to generate automated reports synthesizing trends, performance metrics and so on.

4.3.4.3 Barriers to the implementation of the use case

In sectors such as wine, traceability is primarily used for regulatory compliance and brand protection. While centralized consortia foster uniformity and credentialing, they often **limit flexibility and innovation** in incorporating digital tools or in promoting new business opportunities given by new regulations instead of considering them only as new costs.

Desk research highlights that agricultural commodities often pass through multiple layers of intermediaries before reaching export or processing stages, with minimal or inconsistent documentation and digital tracking at each step. One of the interviewed expert underlines that current regulations often focus only on the production level, not the entire supply chain, leaving important gaps. Consequently, significant portions of global commodity flows remain untraceable to the farm level, particularly in regions at high risk of deforestation. The barriers to establishing deforestation-free supply chains are rooted in deep structural and systemic challenges that cannot be resolved solely through technological solutions or corporate commitments.⁷⁵⁵ As one interviewee noted, technologies like blockchain can help, but they work best when participants already know and trust each other. Another expert also pointed out that relying on blockchain alone is not an infallible system as it does not verify the authenticity of data. Therefore, in the context of relying on blockchain, further costs should be considered for applying additional data verification layers, for example the cross-checking of GPS data. **Trusted ecosystems, such as data spaces**, overcomes this technological limitation trying to

⁷⁵⁵ Renier, Cécile et al. (2023). *Transparency, traceability and deforestation in the Ivorian cocoa supply chain*. Environmental Research Letters, Volume 18, Number 2. Available online: <https://iopscience.iop.org/article/10.1088/1748-9326/acad8e>

enforce trust among all stakeholders, data space operators, data product providers and consumers, regardless the specific technologies adopted.

Implementing effective traceability systems faces major obstacles related to technical capacity, financial feasibility, organizational readiness, and regulatory clarity.⁷⁵⁶ A lack of clear and harmonized regulations impedes data interoperability between digital systems, which critically limits traceability, independent verification, and scalability. Weak data governance further undermines credibility, erodes trust, and reduces stakeholder participation in digital traceability platforms.⁷⁵⁷ This barrier is also pointed out by one of the stakeholders in relation to the current Brazilian institutional framework. Besides the harmonization of regulations, the stakeholder input also stressed the importance of harmonizing the regulation of different policy fields as it is unclear how these tools will comply with both food regulations and digital regulations at the same time, not only for exporters and importers, but also for food producers within the EU. This overlap between the digital and food sectors is still evolving, and we are waiting on more clarity from regulators.

Stakeholders input also confirms that in addition to these challenges, **data sharing policies** in countries such as Brazil and China can act as inhibiting factors, particularly when domestic regulations restrict the exchange of commercially sensitive or personally identifiable information. These barriers may be partially overcome if there is sufficient buy-in from data holders, such as companies, who are willing to authorize the release of relevant information. However, further obstacles arise from geopolitical dynamics and divergent policy perspectives. For instance, some consumer countries do not perceive their import-related consumption as a driver of deforestation, in contrast to the EU's approach. In these contexts, trade-related sustainability requirements are often viewed with scepticism or as non-tariff barriers. Moreover, global geopolitical tensions have further eroded trust and cooperation, reducing the willingness of some governments and stakeholders to engage in data sharing and joint monitoring initiatives.

Financial constraints hinder the development of secure and reliable systems, as there is often insufficient investment in the technological, infrastructural, and human resources required. This in turn contributes to technical limitations, such as the absence of affordable, interoperable, and user-friendly tools for data collection and reporting. Financial incentives are also necessary to avoid exclusion and resistance, particularly among smallholders.

Organizational challenges include inconsistent definitions, lack of standardised data governance frameworks, limited access to independent verification mechanisms, and inadequate training and support for farmers and cooperatives. Collectively, these issues reinforce distrust and disengagement, especially among smallholders, creating significant barriers to the establishment of inclusive and functional traceability systems.⁷⁵⁸

⁷⁵⁶ Cuellar, David- Johnsohn, Zechariah (2022). *Barriers to implementation of blockchain technology in agricultural supply chain*. Available online:

https://www.researchgate.net/publication/366092563_Barriers_to_implementation_of_blockchain_technology_in_agricultural_supply_chain

⁷⁵⁷ Hinker, Cordula - Günter, Peter (2019). *A conceptual framework for deforestation-free supply chains applied to soy certification*. Emerald publishing. Available online:

<https://www.emerald.com/sampj/article-abstract/11/7/1159/337403/Traceability-mattersA-conceptual-framework-for?redirectedFrom=fulltext>

⁷⁵⁸ Fripp, Emily et al. (2023). *Traceability and transparency in supply chains for agricultural and forest commodities. A review of success factors and enabling conditions to improve resource use and reduce forest loss*. Report. World Resources Institute. Available online:

Concerning GenAI adoption, lack of trust and fear of losing control of actions on own data is probably a strong barrier. Indeed, from the perspective of several stakeholder (regulators, EU importers, retailers etc) adopting GenAI tools and techniques integrated within their traceability system, especially if built on data spaces present psychological and practical barriers. The psychological barrier appears as farmers could show reticence towards AI systems recommending actions for their certifications, particularly when these AI systems lack GenAI explainability mechanisms such as recommendations. Additionally, from the farmers' perspective, the operational costs of GenAI tools adoptions, maintenance costs or fees could also represent a practical barrier.

4.3.4.4 Conclusions. The value for the EU: Lessons learnt and opportunities

The Deforestation-Free Agricultural Supply Chain Traceability Use Case is a cornerstone of the Green Deal Data Space. It is the practical, data-driven enforcement mechanism for the EU's Farm to Fork and Biodiversity strategies. It demonstrates the EU's willingness to use its market power and data policy to achieve global environmental objectives, making data a key tool for enforcing sustainability standards.

The value for the EU is the creation of a credible, enforceable mechanism to deliver on its Green Deal commitments, significantly reducing the EU's global deforestation footprint. On a global scale, it promotes much-needed transparency and sustainability in opaque agricultural supply chains. By creating a clear market incentive for verifiably deforestation-free products, it has the potential to drive positive land-use changes in producer countries.

The interviewees highlighted that the rollout of the EU Deforestation Regulation opens up new possibilities for improving forest governance and reinforcing the application of environmental laws in producer countries like Brazil. Realising this potential will require the EU to work closely with key stakeholders across the supply chain, including governments and private sector actors, to build on existing efforts, particularly the ones who aimed at developing and scaling national traceability systems for agricultural goods. Such engagement should include financial support targeting technical development, preventing fragmentation by mandating interoperability and setting clear standards as well as the facilitation of knowledge and best-practice sharing.

The Forest Code of Brazil already lays out a comprehensive legal framework for sustainable land use on private property, including provisions that enable landowners to restore legal compliance. Despite this, implementation and enforcement have proven difficult. The timing of the EUDR could help strengthen these domestic efforts by aligning EU market access with improved compliance. Additionally, carefully coordinated financial and technological support from the EU and major commodity traders are essential to incentivise farmers to align with national regulations and meet EUDR requirements.

In this context the EU's sustainability framework provides legal and strategic incentives for companies to adopt traceability systems that go beyond compliance, enabling them to demonstrate environmental and social responsibility and enhance brand value through verified sustainability claims.

In parallel, there is also strategic value in engaging with China, given its role as a major consumer of agricultural commodities. Insights gained from the EU's experience with implementing the EUDR, specifically around traceability, monitoring, and evaluation could serve as useful input into China's own environmental governance. This cooperation could also pave the way for future bilateral agreements in which the EU can continue to position itself as a leader in global sustainability standards.

4.3.5 Use case 5: Next-Generation Passenger Name Record (PNR) Exchange with Privacy-Enhancing Technologies (PETs)

4.3.5.1 Overall description and use case rationale

The Next-Generation PNR Exchange with PETs use case falls within the sectors of public security, aviation and border management. The overall objective of the use case is to establish a pilot project with Australia and the UK to re-architect the Passenger Name Record data exchange system. The goal is to move from the current model of bulk data transfers to a more targeted system that utilises Privacy-Enhancing Technologies such as federated analytics, secure multi-party computation or zero-knowledge proofs.

Existing PNR systems are built on a paradigm of mass data collection. Airlines are obligated to transfer large dossiers of personal data for all passengers on international flights to government authorities for the stated purpose of combating terrorism and serious crime.⁷⁵⁹ This model of indiscriminate, bulk transfer and retention has been heavily criticised by civil society and data protection authorities as a form of mass surveillance that is disproportionate and infringes on fundamental rights to privacy and data protection.⁷⁶⁰ The Court of Justice of the European Union (CJEU) has repeatedly scrutinised these schemes, imposing strict limitations on the EU's own PNR Directive and invalidating previous agreements (e.g., the EU-Canada PNR agreement). The court demands that such systems be limited to what is "strictly necessary".⁷⁶¹ A PET-based approach would therefore offer a path to meet security objectives in a manner that is more compliant with these fundamental rights principles.

The primary barrier is institutional and political inertia. The security and law enforcement communities are deeply invested in the "collect it all" paradigm of bulk data transfer, which they see as essential for effective analysis.⁷⁶² An additional important hurdle is the technical implementation of PETs at the scale and speed required for PNR processing. This is a complex endeavour which has not yet been attempted in a live operational environment. Furthermore, existing international PNR agreements, such as the one between the EU and Australia, are legally and technically structured around the old model of bulk transfer and would need to be fundamentally renegotiated⁷⁶³. The poor quality and

⁷⁵⁹ European Parliament and the Council (2016). *Directive (EU) 2016/681 on the use of passenger name record (PNR) data for the prevention, detection, investigation and prosecution of terrorist offences and serious crime*. Official Journal of the European Union. Available online: <https://eur-lex.europa.eu/eli/dir/2016/681/oj/eng>

⁷⁶⁰ EU Migration Law Blog (2016). *I travel, therefore I am a Suspect*. Available online: <https://eumigrationlawblog.eu/i-travel-therefore-i-am-a-suspect-an-overview-of-the-eu-pnr-directive/>

⁷⁶¹ Statewatch (2022). *EU: Travel surveillance: member states seek to circumvent court judgment on PNR*. Available online: <https://www.statewatch.org/news/2022/september/eu-travel-surveillance-member-states-seek-to-circumvent-court-judgment-on-pnr/>

⁷⁶² EU Migration Law Blog (2016). *I travel, therefore I am a Suspect*. Available online: <https://eumigrationlawblog.eu/i-travel-therefore-i-am-a-suspect-an-overview-of-the-eu-pnr-directive/>

⁷⁶³ European Union (2012). *Agreement between the European Union and Australia on the processing and transfer of Passenger Name Record (PNR) data by air carriers to the Australian Customs and Border Protection Service*. Available online: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2012:186:0004:0016:EN:PDF>

lack of standardisation in the PNR data provided by airlines also presents a significant challenge to any new system.⁷⁶⁴

This use case is a direct and necessary response to the "Schrems effect" and the consistent message from the EU's own judiciary. By proactively developing and piloting a privacy-preserving alternative to bulk data transfer, the EU can get ahead of the next inevitable legal challenge. It is a strategic move to "future-proof" international security cooperation. Instead of relying on legal patches and political assurances that have repeatedly failed to satisfy the courts, this approach embeds fundamental rights protection directly into the technical architecture of the system. Piloting this with trusted, like-minded partners such as Australia and the UK, which have strong rule-of-law traditions and existing PNR agreements, is the logical first step.⁷⁶⁵ A successful outcome would create a new global standard that is "privacy by design and by default," fundamentally altering the terms of the global debate on security and data sharing.

4.3.5.2 Requirements for enabling the use case

To ensure the successful implementation of next generation Passenger Name Record exchange with privacy enhancing technologies between the EU, Australia, the UK, and Canada, several key requirements must be met. First, there must be harmonised legal and regulatory frameworks across all participating countries. This includes clear, aligned privacy standards, interoperable data protection rules, and formal international agreements to enable secure and lawful cross-border data sharing. Legal certainty and mutual recognition of standards are essential to avoid ambiguity and ensure compliance by all stakeholders.

From a technical perspective, robust, modern infrastructure is needed to support secure and efficient data exchange. This includes upgrading legacy aviation systems to support advanced encryption, strong authentication, and privacy-enhancing technologies such as secure multi-party computation⁷⁶⁶ and/or zero-knowledge proofs⁷⁶⁷. Efficient, high-speed network connections and resilient gateways for data transmission are crucial, as is the adoption of common data standards (including PNRGOV⁷⁶⁸) to facilitate semantic interoperability between different systems and operators. Regular updates and certification of encryption methods are necessary to keep pace with technological advances and emerging threats.

Operationally, standardised procedures and comprehensive training for all personnel involved in PNR data handling are required. This involves developing and implementing

⁷⁶⁴ Aposkiti, Christina and Freideriki Maki (2025). *Navigating the challenges of passenger name record data and the way forward*.

⁷⁶⁵ Office of the Australian Information Commissioner (2025). *Cross-border disclosures of personal information: Passenger Name Records*. Australian Government. Available online: <https://www.oaic.gov.au/privacy/privacy-assessments-and-decisions/privacy-assessments/cross-border-disclosures-of-personal-information-passenger-name-records>

⁷⁶⁶ TNO website, *Secure multi-party computation*, available at: <https://www.tno.nl/en/technology-science/technologies/secure-multi-party-computation/>

⁷⁶⁷ Zero-knowledge proofs (ZKPs) are cryptographic protocols that allow one party (the prover) to prove the truth of a statement to another party (the verifier) without revealing any information beyond the validity of the statement itself. For more details see: https://en.wikipedia.org/wiki/Zero-knowledge_proof

⁷⁶⁸ The International Air Transport Association (IATA), (2021), *Passenger and Airport Data Interchange Standards*. Available online: https://www.iata.org/contentassets/18a5fdb2dc144d619a8c10dc1472ae80/pnrgov-edifact-implementation-guide_21_1.pdf

uniform operational manuals, certification programmes, and ongoing education to ensure proper handling of sensitive information and compliance with privacy protocols. Clear guidelines for data collection, processing, and sharing must be established, and all actors – including airlines, authorities, and technical vendors – should be actively engaged in collaborative working groups to address challenges and share best practices.

Finally, strong governance, stakeholder engagement, and incentives are vital. Governments and regulators should facilitate transparent communication, inclusive participation from all relevant industry players (including smaller operators and non-IATA members), and provide funding or support for necessary technological upgrades. Building trust among stakeholders, ensuring transparency in data handling, and creating a centralised or harmonised data gateway for PNR exchange will help overcome resistance and foster long-term cooperation. The key stakeholders are EU and national law enforcement and border agencies (specifically the Passenger Information Units, or PIUs), the EU Counter-Terrorism Coordinator, national data protection authorities, the European Union Agency for Fundamental Rights (FRA), and the relevant authorities in partner countries (Australian Department of Home Affairs, UK Home Office). Airlines would still be the source of the data, but the mechanism of government access would change. Experts in PETs from academia and the private sector would be crucial technical partners

4.3.5.3 Barriers to the implementation of the use case

Stakeholder feedback confirmed that many citizens and airlines are uncomfortable with personal data or datasets being shared globally, especially beyond countries they travel to. At the same time, one stakeholder expressed doubts about the feasibility of the use case, pointing out that it is yet to be proven if the use of PET solutions – with less privacy invasive data exchange – would admit the same level of security analysis. The interviews also highlighted significant legal and regulatory barriers to implementing next generation PNR exchange with privacy enhancing technologies between the EU, Australia, the UK, and Canada. Each jurisdiction has its own legal frameworks and privacy regulations, making harmonisation complex and slow. Even within the EU, member states interpret and implement directives differently, resulting in inconsistent data handling and compliance standards. Moreover, the absence of unified international agreements and ongoing legal disputes further complicates the process, creating uncertainty for stakeholders about their obligations and rights in cross-border data exchange.

Technical and infrastructural challenges also pose major obstacles. The aviation sector relies heavily on legacy systems and outdated data formats, which are often incompatible with new privacy-enhancing technologies. Not all airlines and countries employ robust encryption or secure transmission methods; some still use unsecured channels such as plain text email for sensitive passenger data. Additionally, large technology vendors exhibit inertia, making it difficult for the industry to quickly adopt and implement new standards, especially when upgrades require significant investment and time.

Operational barriers stem from the lack of standardised training and certification for personnel handling PNR data. Airlines and authorities follow different operational manuals and procedures, leading to mistakes and inefficiencies in data processing. The sheer size of the aviation industry means that retraining and certifying staff on new privacy and security standards would be a massive and costly undertaking. These inconsistencies are further exacerbated by fragmented stakeholder engagement, as not all relevant industry players are involved in policy and technical discussions.

Political, commercial, and cultural factors also hinder collaboration. There is a persistent lack of trust between stakeholders regarding data sharing, with concerns about losing control of sensitive information or exposing competitive business data. Overcoming these barriers will require greater transparency, incentives such as funding for technical

upgrades, and centralised or harmonised data gateways – though achieving these solutions will likely be a long-term process.

4.3.5.4 Conclusions. The value for the EU: Lessons learnt and opportunities

This use case directly addresses the core EU challenge of balancing security with the protection of fundamental rights, a foundational value of the Union.⁷⁶⁹ It represents a significant potential innovation for developing Common European Data Space for Public Administration. By pioneering a more privacy-respecting model for international security cooperation, the EU can set a new global standard that aligns with its legal and ethical commitments.

The primary value is the establishment of a legally and ethically sustainable model for PNR and other sensitive law enforcement data exchanges. This would resolve the long-standing and damaging conflict between executive security imperatives and judicial oversight. For the EU, it is an opportunity to lead globally in the development of rights-respecting security technologies. A successful pilot could provide a blueprint for renegotiating agreements with all international partners, including the highly sensitive agreement with the US, moving beyond a cycle of legal challenges and makeshift solutions.

The next generation PNR exchange use case between Europe, the UK, Canada, and Australia presents a significant opportunity to enhance security and border management across these regions. By enabling real-time, harmonised sharing of passenger data with privacy-enhancing technologies, authorities could more effectively identify security threats and manage migration, while ensuring compliance with high data protection standards. This would lead to improved risk assessment and more efficient, less intrusive border control processes for travellers.

Additionally, the implementation of a standardised and secure PNR exchange system can streamline operations for airlines and border agencies. By reducing redundant checks and harmonising data flows, both passengers and authorities would benefit from smoother travel experiences and reduced administrative burdens. This operational efficiency could also lead to cost savings and improved service quality for all stakeholders involved in international travel.

The use case also provides a platform for Europe to reinforce its leadership in data privacy and international cooperation. Successfully integrating privacy-enhancing technologies into cross-border PNR exchange would set a global benchmark, demonstrating that robust security and strong privacy can coexist. This approach could influence global standards and encourage other regions to adopt similar practices, fostering greater international trust and collaboration.

From this initiative, Europe can learn valuable lessons about the importance of legal and technical harmonisation, stakeholder engagement, and system flexibility. Early involvement of all relevant parties, transparent communication, and adaptable technical solutions are critical to overcoming implementation challenges. Furthermore, balancing security with privacy and maintaining public trust through clear safeguards and communication will be essential for the long-term success and acceptance of such cross-border data sharing projects.

⁷⁶⁹ European Commission (2024). *Priorities 2024-2029*. Available online: https://commission.europa.eu/priorities-2024-2029_en

4.3.6 Use case 6: Federated Digital Twins for Resilient Aerospace Supply Chains

4.3.6.1 Overall description and use case rationale

The Federated Digital Twins for Resilient Aerospace Supply Chain Use Case is framed under the advanced manufacturing as well as aerospace & defence sectors. The use case aims to create a federated network of digital twins for critical aerospace components (e.g., jet engines, landing gear, avionics, etc.) across the complex, multi-tiered supply chains of EU, Canadian and US manufacturers.

Modern aircraft are among the most complex systems ever built, with their development and production relying on deeply integrated, international supply chains. A digital twin—a virtual representation of a physical asset, updated with real-time data—is becoming essential for design, manufacturing, and maintenance. The full value of a digital twin is only realized when it is a "living" model, continuously updated with performance and maintenance data throughout the component's lifecycle. This requires data to flow securely between numerous stakeholders: the aircraft OEM (e.g., Airbus), the engine manufacturer (e.g., Safran), and multiple tiers of parts suppliers. A centralized model is unworkable due to IP and security concerns. A federated approach, however, allows each stakeholder to maintain sovereignty over their own digital twin while enabling controlled, secure, and contractually defined data/information exchange between them. Moreover, depending on the sensitivity of the information, this federated model would serve to exchange data insights and model parameters between members of the network, rather than raw data. This enables use cases like predictive maintenance, performance optimization, and supply chain resilience analysis on a system-wide level.

The technical and governance challenges are immense. This use case pushes the boundaries of B2B data sharing. It requires achieving semantic interoperability between highly complex and proprietary digital twin platforms from different vendors. More importantly, the governance frameworks for sharing extremely high-value intellectual property and commercially sensitive operational data across corporate and national boundaries are not yet established. Companies are fiercely protective of this data. While initiatives like the German COOPERANTS project are making progress, they are largely national or intra-EU in scope and have not yet tackled the complexities of a truly transatlantic federated system.⁷⁷⁰

This use case represents the ultimate test for the concept of industrial data sovereignty. If the deeply intertwined yet fiercely competitive supply chains of Airbus and Boeing can find a way to securely share digital twin data in a federated, sovereign manner, the model can be replicated across any complex manufacturing sector, from semiconductors to medical devices. This is about more than just efficiency; it is about building the infrastructure for collective industrial resilience. In an era of increasing supply chain vulnerability and geopolitical competition, creating a trusted data/insight-sharing framework for the West's most critical industries is a strategic imperative. A successful pilot here would provide an irrefutable proof-of-concept for the EU's federated data space model, establishing a powerful, open, and collaborative alternative to the vertically integrated, closed ecosystems that might otherwise dominate the future of Industry 4.0.

⁷⁷⁰ Gaia-X (2025). *Gaia-X and Partners to Showcase Cross-Border Data Collaboration at Hannover Messe 2025*. Available online: <https://gaia-x.eu/gaia-x-and-partners-to-showcase-cross-border-data-collaboration-at-hannover-messe-2025/>

4.3.6.2 Requirements for enabling the use case

This use case involves the exchange of extremely sensitive, high-value, non-personal industrial data/information. This includes insights from proprietary **CAD models**, **engineering simulation data**, **real-time sensor telemetry** from aircraft in operation, **maintenance, repair, and overhaul (MRO) logs**, **detailed material composition data**, and **software configuration files**. The ability to share this data/information in a controlled and secure manner is the core challenge.

The ecosystem is composed of the champions of the global aerospace industry. This includes major **aerospace OEMs** (e.g., Airbus, Boeing, Dassault, Bombardier), **engine manufacturers** (e.g., Safran, Rolls-Royce, Pratt & Whitney), hundreds of **Tier 1-N suppliers** on both sides of the Atlantic, **MRO providers**, and national and international **aviation safety authorities** (e.g., EASA, FAA). Crucially, it also involves the **technology enablers** who provide the platforms and standards for this collaboration, such as Dassault Systèmes, Siemens, and the proponents of federated data architectures like **IDSA** and **Gaia-X**.

4.3.6.3 Barriers to the implementation of the use case

The primary obstacle faced by the use case is likely of political nature, particularly concerning the US lack of alignment with the EU on key policies and regulations, which is added to the potential imposition of barriers to access to the US market by Washington and its preference to support national companies, i.e., European competitors. Geopolitical considerations are also of great relevance, as the increasing competition between global actors is hindering the capability of sharing information – increasingly securitised and protected.

The potential presence of industrial champions in the use case such as Airbus or Boeing adds a new layer of factors in the implementation process. Geopolitical considerations that go beyond the private and commercial interests of each company. In this sense, the identified reluctance to share data due to commercial concerns meets with the reluctance to share data on the basis of geopolitical concerns – especially when dealing with strategic or security and defence projects. In turn, this limits the possibility to develop common standards, essential for the establishment of cross-border frameworks for data transfers.

In the technical domain, interoperability challenges pose a major barrier in terms of ontology conversion, the implementation of API services and the configuration management of digital twins, which are crucial for guaranteeing the timely availability of precise information. These challenges are further worsened due to a lack of development and adoption of common standards, which limit the establishment of transnational networks.

4.3.6.4 Conclusions. The value for the EU: Lessons learnt and opportunities

This is a potential flagship use case for Aerospace-X,⁷⁷¹ a development project under Manufacturing-X, the dedicated data space for Industry 4.0 that is a key pillar of the German and broader European industrial strategy.⁷⁷² The use case directly supports the EU's goals for industrial competitiveness, innovation, and strategic autonomy in a sector of critical economic and security importance. It builds directly on the principles of data sovereignty and interoperability championed by Gaia-X and IDSA, which are foundational

⁷⁷¹ Aerospace-X website, available at: <https://www.aerospace-x.net/en.html>

⁷⁷² Factory-X website, *Manufacturing-X*. Available online: <https://factory-x.org/manufacturing-x/>

to the EU's data space vision.⁷⁷³ Existing national or European focused projects and initiatives provide a starting point for this international expansion, like the German project COOPERANTS,⁷⁷⁴ which targets the aerospace sector within a Gaia-X framework, or DECADE-X,⁷⁷⁵ a non-profit association established in June 2025 with the aim to drive the development of a sovereign, interoperable, and secure digital ecosystem for the aerospace and defence sectors.

The potential benefits are substantial. On a practical level, it would drastically improve aviation safety and operational efficiency through system-wide predictive maintenance. Strategically, it creates more resilient and transparent supply chains, providing end-to-end visibility that is crucial for mitigating disruptions caused by geopolitical events or other crises. For the EU, it is a prime opportunity to establish its preferred data space architecture (based on IDSA and Gaia-X) as the global standard for high-stakes, collaborative industrial data sharing. This would create a significant competitive advantage for European technology providers and embed EU governance principles into the fabric of 21st-century manufacturing.

4.4 Task 3 policy recommendations

This section presents conclusions and policy recommendations related to Task 3 findings, linking back to the analyses performed in section 4.2 and 4.3. The first subsection focuses on recommendations linked to the global analysis while the second one is connected to the key findings from the analysis of sensitive and highly sensitive data categorisation.

Finally, the third subsection provide summarised conclusions and recommendations linked to the use case analyses. As mentioned in the methodological approach (section 4.1.3) the use cases are of exploratory character and quite widely defined, why full-fledged use case analyses were not feasible within the scope of this study. The conclusions and recommendations for each use case should therefore be interpreted with care.

4.4.1 Policy recommendations based on the Global analysis

The analysis of the global data landscape reveals a clear strategic path for the EU. Brussels is in a unique position to move beyond articulating a “third way” in data governance, distinct from the market and state-driven models, and to actively build the technical, legal and political infrastructure that underpins it. To seize this opportunity, a proactive and targeted strategy focused on five key pillars should be pursued:

Leverage regulatory gravity to export EU data space architecture

The EUDR and the Battery Regulation demonstrate that EU legislation can create global demand for new data-sharing ecosystems. The EU should not therefore be a passive regulator but an active architect. Building on the global influence, the EU should:

- Expand the "Adequacy Alliance" by accelerating negotiations with Brazil (partial GDPR alignment) and Australia, offering sector-specific adequacy agreements

⁷⁷³ Peters, Maj-Britt (2021). *Volkswagen builds supply chain efficiency through industry-wide data exchange*. Volkswagen Group. Available online: <https://www.volkswagen-group.com/en/press-releases/volkswagen-builds-supply-chain-efficiency-through-industry-wide-data-exchange-16755>

⁷⁷⁴ Gaia-X (2025). *Gaia-X and Partners to Showcase Cross-Border Data Collaboration at Hannover Messe 2025*. Available online: <https://gaia-x.eu/gaia-x-and-partners-to-showcase-cross-border-data-collaboration-at-hannover-messe-2025/>

⁷⁷⁵ DECADE-X website, available at: <https://www.decade-x.org/>

(e.g., for Brazil's finance or South Korea's smart cities) to bypass localisation barriers.

- Actively promote Gaia-X and IDSA-compliant architectures as the default solution for compliance with extraterritorial regulations, funding pilot projects (e.g., via Digital Europe or Horizon Europe) to build open-source platforms for Battery Passports and EUDR traceability.
- Secure critical data infrastructure by scaling EU-controlled cloud/AI platforms (e.g., Gaia-X), reducing reliance on US and Chinese tech while maintaining interoperability with trusted partners like Japan and South Korea.

Pioneer new governance models to solve trust deficits

Rather than technology, lack of trust and viable governance models stand as the primary barrier for cross-border data sharing, especially in sensitive and security-related areas. The EU can lead by designing and implementing novel governance structures.

- Japan-inspired hybrid frameworks to bridge divides between strict GDPR nations and innovation-focused states. Pilot a "data trustees" model (e.g., an AV Safety Data Trust under the Mobility Data Space) to overcome B2B/B2G trust gaps.
- Lead global AI governance by partnering with South Korea (via its AI Sandbox) and Japan to harmonise the AI Act's ethics standards with agile innovation policies, creating a template for transnational AI safety.
- Adapt GDPR for innovation through targeted regulatory sandboxes in AI/Health tech, mirroring South Korea's "Digital New Deal" partnerships, while upholding core privacy safeguards.

Transform friction into partnership through proactive engagement

Regulations like the GDPR and the EUDR can create diplomatic friction. Under the consensus-driven model, the EU can mitigate this and build stronger alliances by pairing regulatory requirements with tangible support and partnership offers. To counter fragmentation and build alliances, the EU should:

- Launch a "Traceability for Sustainable Trade" initiative with partners like Brazil, offering technical assistance and co-funding for inclusive traceability systems (e.g., smallholder-focused platforms), positioning the EU as a partner in sustainable development.
- Advance multilateral data diplomacy via G7/OECD forums to promote GDPR-inspired norms, isolating China's state-driven model while engaging the US on algorithmic transparency.
- Develop transnational data spaces in sectors like healthcare, leveraging South Korea's hybrid approach (localised storage and cross-border R&D) to demonstrate secure, rights-respecting data sharing.

Prioritise Privacy-Enhancing Technologies (PETs) to future-proof security cooperation

The consistent legal challenges to bulk data transfer agreements are a strategic vulnerability for EU security cooperation. The EU must move from a defensive, reactive posture to a proactive one by leading the transition to a new technological paradigm. To address security vulnerabilities and reinforce trust:

- Initiate PET-based pilots with the UK and Australia for next-gen PNR systems, combining DG HOME and EEAS efforts to embed fundamental rights into security architectures.

- Balance rights and competitiveness in trade deals by allowing limited deregulation in non-sensitive sectors (e.g., logistics) with the UK or Brazil, while enforcing GDPR standards in strategic areas like finance.

Scale and structure R&D Funding for AI and Data Spaces

The rapid pace of technological innovation, particularly in AI and data space architectures, demands a stronger and more coordinated investment strategy. While regulatory frameworks provide the backbone for trust and interoperability, these initiatives must be matched by sustained R&D efforts if the EU is to remain competitive and avoid overreliance on third-country technologies:

- Boost R&D funding as a strategic complement to regulation through increased and long-term EU investment in AI and data – Horizon Europe, Digital Europe and future framework programmes - essential to support innovation cycles and accelerate market-ready solutions.
- Prioritise strategic research domains that carry both economic and geopolitical weight, including generative AI, privacy-enhancing technologies (PETs), federated learning and cross-border data space infrastructure. These areas present both opportunities for leadership and critical vulnerabilities if left unaddressed.
- Bridge the gap between research and deployment with targeted support for pilot projects and testbeds (AI sandboxes, cross-border data space demonstrators, etc.) will ensure that EU-funded R&D translates into tangible outcomes. This will foster ecosystems that combine academic excellence, SME innovation and industrial scale-up.

Internationalise R&D partnerships by fostering collaboration with like-minded partners (Japan, South Korea, Canada, the UK, etc.) to strengthen EU's position as a hub for trustworthy AI and data spaces, while also diversifying access to global expertise and markets.

With the overall global analysis and raised opportunities from six use cases, the study provided strategic implementation for the European Commission to:

Short Term	Finalise sectoral adequacy agreements with Brazil (agriculture traceability) and expand GDPR-aligned data spaces with South Korea (smart cities).
Medium Term	Co-design AI ethics frameworks with other positioning partners (e.g. Indo-Pacific) to balance US-China dominance.
Long Term	Establish the EU as the hub for secure, rights-based innovation via Gaia-X and PET-driven security standards.

By pursuing these recommendations, the EU can translate its vision for a human-centric, sovereign and prosperous digital future into a concrete and impactful international agenda. Thus, the identified use cases provide the ideal laboratories for this work, while the selected initiatives provide a much-needed international context, allowing the EU to demonstrate the superiority of its model and shape the next frontier of the global data economy.

4.4.2 Policy recommendations based on the analysis of Sensitive and Highly Sensitive Data Categorisation

The protection and categorisation of data at the global stage is structured around a patchwork of national and regional legislations, each shaped in accordance with national contexts and priorities. Here, the EU has taken the lead in the categorisation of sensitive personal data through its GDPR, not only harmonising legislation at the European level but also setting an international precedent and a legal model. On the contrary, the protection and categorisation of non-sensitive personal data and **highly sensitive data** followed a different path focused on sector-specific regulations and, in some cases, undefined. This reality represents both a challenge for cross-border data flows, as well as an opportunity for the EU to continue leading international regulatory development through the following lines of action:

Reinforce the legal concept of sensitive personal data

The milestone in the protection of personal data entailed by the GDPR has had a *spillover* effect in national legislations across the globe, which have adopted similar categorisations and protections. These achievements should be further promoted by the EU with specific actions:

- Support the alignment of protection measures towards personal data with the EU of those countries that currently include in their legislations a categorisation of sensitive personal data.
- Use the EU Adequacy Decision as incentive to foster alignment, in the categorisation of sensitive personal data as a first stage, and in the protection of the established categories of personal data in a later stage.
- Promote a clear and auditable EU legal framework for de-identification and certified anonymisation, building on GDPR articles 42 and 43, to enable the lawful reuse of data assets for research and industrial innovation, while safeguarding citizens' fundamental rights and minimising re-identification risks. The resulting framework should incorporate the following elements: **EU-level anonymisation certification scheme** overseen by accredited official bodies (designated by the European Commission), based on specific, regularly updated technical and organisational criteria; **transparency and re-identification monitoring** to require the publication of anonymisation metadata for certified datasets as well as mandatory notification to the competent authority and affected individual if breaches occur; **guidance** provided by the competent authority including technical guidelines, test vectors and sandboxes to validate methods; **liability clarification** to limit the responsibility of data controllers and processors for downstream re-identification if compliance with the anonymisation criteria is proved.

Promote the legal codification of non-sensitive personal data internally and abroad

In contrast to the categorisation and protection of personal information, non-personal information potentially deemed as sensitive data is not covered by a comprehensive and unified piece of legislation as personal information is under the GDPR. Instead, it relies in a series of sector-specific and purpose-specific laws as in most of the eight selected countries. This absence of comprehensive and equivalent legislation poses a clear barrier to cross-border data flows, as differing legal safeguards and categorisations – i.e., what is deemed sensitive – in each jurisdiction often hinder compliance and meeting protection standards. This should be addressed by the EU through a series of measures:

- Establish and adopt a formal cross-cutting definition on sensitive non-personal data which shall include but not be restricted to the following categories of data: **trade secrets and business confidential information** – currently covered under Directive (EU) 2016/943; **intellectual property** – currently covered under Directive

(96/9/EC) and Directive (2001/29/EC); **geospatial and environmental data** – currently covered under Directive (2007/2/EC); **infrastructure and sectoral data** – currently covered under sectoral EU frameworks such as Directive (EU) 2019/944 or Directive (EU) 2022/2555; **scientific and research datasets**.

- Explore the possibility to develop a General Non-Personal Data Protection Regulation which includes formal definitions and categorisations of non-personal sensitive and highly sensitive data (see section below), specific safeguards and requirements for data controllers and processors as well as conditions for the exchange of information with third states and international organisations (see section below).
- Coordinate the development of a General Non-Personal Data Protection Regulation or at least the establishment of formal definitions and categorisations of non-personal sensitive and highly sensitive data with likeminded global partners. These efforts would allow for aligned legislative initiatives from the drafting stage, which would facilitate the subsequent creation of mutual recognitions of legal equivalence or other instruments to facilitate trusted and secure international data flows.

Reinforce the protection of highly sensitive data in the EU

The categorisation and codification of non-personal information usually interpreted as highly sensitive experiences similar challenges to the ones described for non-sensitive personal data. Sector, purpose and jurisdiction-specific laws continue to prevail over an EU-level comprehensive and unified piece legislation applicable to both EUIs and Member States. Additionally, the inclusion of classified and critical information, state secrets, etc. in this type of data adds further obstacles due to specific national legislations, criteria and priorities. The high degree of bilateral interdependence between Member States and between these and the EU requires however the development of a truly common protection framework. While the Data Governance Act does leave the door open for highly sensitive categories of data, a general definition and scope have not yet been established for highly sensitive data. In this sense, to accomplish such framework the EU should take action to:

- Establish and adopt a formal definition on special categories of non-personal data identified as highly sensitive, which shall include but not be restricted to: **public health and critical environmental datasets; critical infrastructure information; logistics and supply chain information with an impact to the security and defence of the Union; airspace and maritime control data; cybersecurity data; emergency and civil protection data.**⁷⁷⁶
- Taking into consideration the level-based model of the Protection of European Union classified information (EUCI) and of different frameworks from the eight selected countries such as the US Standards for Security Categorization of Federal Information and Information Systems, incorporate the EUCI's main features into the protection of non-personal sensitive and highly sensitive data.
- Define the specific requirements and safeguards to be applied to data controllers and data processors when handling non-personal sensitive or highly sensitive

⁷⁷⁶ This policy recommendation is based on Recital 24 and Article 5 of Regulation (EU) 2022/868 (Data Governance Act). The Act explicitly states the potential need to attach stricter conditions for certain types of non-personal data that may be identified as highly sensitive through future Union legislative acts.

data, which shall cover: **personal security** - measures to ensure access only to required individuals; **physical security** - physical and technical protective measures to prevent unauthorised access; **administrative security** - measures to reinforce the protection of data throughout its entire lifecycle; **communication and information systems security** - guarantees that such systems will protect the information they handle and will function as they need to, when they need to, under the control of legitimate users and processors; **industrial security** - measures to ensure the personal, physical, administrative and technical protection of data by contractors and subcontractors; **exchange of information with third states and international organisations** - measures including *security of information agreements* or cross-border data transfers screening mechanisms similar to the current EU Foreign Direct Investment screening mechanism.

4.4.3 Policy recommendations based on the Use Cases

The variable nature and scope of the above-presented six use cases as well as the difference between their sectors of application has resulted in the recommendations derived from their analysis not being exclusively regulation-related. Although the development of policy frameworks and tools is key, the promotion of technical standardisation, multi-stakeholder governance and technical interoperability is equally indispensable. In order to pursue the objectives of these use cases, the EU will therefore need to closely cooperate with industrial actors and associations, international standardisation organisations, sectoral stakeholders and academia.

Use case 1: Cross-Border Digital Battery Passport Data Exchange

To maximize the impact of the **Cross-Border Digital Battery Passport use case** and ensure alignment with Manufacturing-X and Green Deal objectives, the European Commission should give priority to **accelerating the development of a unified technical and regulatory framework as a whole ecosystem by moving beyond enforcement**. This should include the advancement of interoperable standards for battery passports and digital twins, guaranteeing their effective integration across the Mobility and Green Deal Data Spaces. To overcome technical and legal barriers, the EU must institutionalise robust safeguards for intellectual property (IP) and data sovereignty, such as secure federated data-sharing protocols and standardised IP agreements. These frameworks can be realized by building upon proactive national solutions such as the German Battery Passport Project. These measures will foster trust among stakeholders while protecting innovation. Additionally, the Commission should invest in scaling digital infrastructure, particularly for SMEs, through targeted funding and capacity-building programmes. This will ensure equitable public-private participation in the data ecosystem and prevent fragmentation, solidifying Europe's role as a global standard-setter.

Further, the EC should establish **dynamic governance mechanisms** to sustain long-term success. A multi-stakeholder platform, involving industry, academia, and civil society, could iteratively refine standards and address emerging challenges like supply chain traceability or circular business models. Public-private partnerships (PPPs) should be incentivised to co-develop sustainable infrastructure, while regulatory sandboxes could test innovative approaches to data sharing without compromising competitiveness. To reinforce Europe's first-mover advantage, the EU should proactively export its frameworks via trade agreements and multilateral collaborations, embedding its sustainability and interoperability principles globally. Finally, harmonising core requirements with regional flexibility such as allowing Member States to tailor implementation to local industries will balance compliance with agility, ensuring the initiative remains both scalable and adaptable in a rapidly evolving technological landscape.

Use case 2: Federated Learning for Rare Disease Genomics

The **use case on Federated Learning for Rare Disease Genomics** shows strong potential societal value by accelerating diagnosis and treatment of rare diseases, an area of significant unmet medical need. The initiative is well aligned with EU strategic R&I priorities under Horizon Europe, and South Korea's participation in Pillar II provides a direct political and funding pathway for collaboration. This use case, or initiative, creates an **opportunity for the EU to position itself as a global leader in ethical, privacy-preserving AI for health**, reconciling innovation with data sovereignty and patient privacy. The federated model developed here could be extended to other sensitive domains such as cancer genomics, mental health, and paediatrics.

Stakeholders broadly support its potential, but the current timeline for international data exchange under the EHDS makes it unsuitable as a flagship initiative directly under EHDS. Instead, **piloting within a regulatory sandbox** is recommended, as this would allow faster progress while ensuring future alignment with the EHDS ecosystem. As portrayed in the use case description (see section 4.3.2) its implementation comes with challenges linked to several interrelated areas, including technical challenges and legal ambiguity, data quality and interoperability, incentives for participation, and trust and transparency. The European Commission and Member States should set strategic direction by developing policies, harmonising standards, offering incentives, and investing in infrastructure. They must also encourage bottom-up action by supporting collaborative projects and funding initiatives that drive change across stakeholders.

Use case 3: Interoperable Safety Data Exchange for Autonomous Vehicles (AVs)

The **Interoperable Safety Data Exchange for AVs use case** supports EU's "Vision Zero" goal of eliminating road fatalities as well as its overarching strategy for connected and automated mobility and aligns with the ambition of Gaia-X to boost sovereign, federated data ecosystems for key European industries. Taking the lead in creating a trusted, non-punitive environment would result in advantages and potential influence, enabling the EU to reposition itself as a global convener, representing a third way between China's highly centralised, state-controlled AV data regulation and Silicon Valley's company-driven ecosystem which often lack transparency. Realizing this ambition however requires significant efforts, as EU has already fallen behind China and the US.

Stakeholders agreed that the emphasis should be put on governance action and trust framework, as technological breakthrough is not enough to create engagement in such a competitive industry. The EU can step in creating a pioneer "data trustee" model by clarifying data usage rules, non-disclosure agreements and limiting the market access for actors that lack offering verifiable information as well as by establishing a data intermediary. This neutral body could be linked to an existing agency such as EASA or a new entity under the Mobility Data Space. Thus, manufacturers can share just enough information to allow for the reconstruction and analysis of safety-critical scenarios without revealing proprietary algorithms or competitive secrets. On a global perspective, the consulted experts suggested South Korea and Japan as potential allies, as both countries already maintain substantial government agencies dedicated to AV data management, demonstrating not only a robust toolkit but also a strong commitment.

Use case 4: Deforestation-Free Agricultural Supply Chain Traceability

The **use case on Deforestation-Free Agricultural Supply Chain Traceability** is a cornerstone of the Green Deal Data Space. It is the practical, data-driven enforcement mechanism for the EU's Farm to Fork and Biodiversity strategies. Stakeholders underlined that close cooperation between the EU and the versatile group of key stakeholders is essential to realize the potential of the regulation. Such cooperation should build upon the currently developing traceability systems for agricultural goods and consider the already laid out legal framework such as the one adopted by Brazil. The timing of the EUDR could help strengthen these domestic efforts by aligning EU market access with improved compliance. Additionally, coordinated financial and technological support from the EU and

major commodity traders could incentivise farmers to align with national regulations and meet EUDR requirements. It is worth mentioning that companies export into multiple countries at once, indicating the need for a high level of data sharing between enforcement and confidence authorities in Member States which can probably be assured by an EU agency as an intermediary.

In parallel, there is also strategic value in engaging with China, given its role as a major consumer of agricultural commodities. Insights gained from the EU's experience with implementing the EUDR, specifically around traceability, monitoring, and evaluation could serve as useful input into China's own environmental governance. This cooperation could also pave the way for future bilateral agreements in which the EU can continue to position itself as a leader in global sustainability standards.

Use case 5: Next-Generation Passenger Name Record (PNR) Exchange with Privacy-Enhancing Technologies (PETs)

To advance the rights-respecting **Passenger Name Record data exchange** model as illustrated in the analysis of use case 5, the Commission should prioritise **developing a harmonised legal and technical framework** that embeds privacy-enhancing technologies as a cornerstone of cross-border security cooperation. This framework must reconcile security imperatives with fundamental rights by establishing clear safeguards, such as data anonymisation protocols, strict access controls, and independent judicial oversight mechanisms. By formalising these standards in revised agreements with international partners like the US, Canada, and Australia, the EU can position itself as a global leader in ethical security innovation. Simultaneously, the EC should invest in piloting PET-driven solutions such as secure multiparty computation or federated learning in order to demonstrate scalable, real-time data sharing that meets GDPR standards. This would not only resolve longstanding legal conflicts but also create a replicable blueprint for future partnerships, ensuring Europe sets the benchmark for interoperable, rights-centric security systems.

To ensure long-term success, the EC should adopt a **collaborative governance approach** that engages stakeholders early and sustains public trust. A multi-disciplinary task force or working groups, comprising law enforcement, data protection authorities, airlines, and civil society, should co-design implementation protocols, ensuring transparency and addressing operational concerns like redundant checks or passenger burdens. Regulatory sandboxes could test system adaptability across jurisdictions. To build and maintain public trust, targeted public awareness campaigns should clearly communicate the specific safeguards in place such as data encryption, anonymisation, and strict access controls to proactively address and alleviate privacy concerns, ensuring that citizens understand how their personal information is protected throughout the process. Additionally, the EU should leverage this initiative to promote global adoption of its standards through multilateral forums, reinforcing its leadership in ethical security. Finally, embedding continuous review mechanisms, such as impact assessments and stakeholder feedback loops, will maintain the balance between evolving security needs and unwavering commitment to fundamental rights, securing both public acceptance and international credibility.

Use case 6: Federated Digital Twins for Resilient Aerospace Supply Chains

By analysing the use case **Federated Digital Twins for Resilient Aerospace Supply Chains**, the Study team concluded that to accelerate the adoption of Manufacturing-X in critical sectors like aviation, the EC should prioritise **scaling and harmonizing existing national initiatives** such as Germany's COOPERANTS project **within a unified EU-wide framework**. By embedding Gaia-X and IDSA principles into binding interoperability standards, the EU can ensure seamless cross-border data sharing while safeguarding sovereignty. Targeted funding should support the integration of predictive maintenance systems and supply chain visibility tools, leveraging public-private partnerships to develop

shared infrastructure. This would maximize operational efficiency and crisis resilience while creating a blueprint for pan-European collaboration. Additionally, the EC should fast-track regulatory sandboxes to test data-sharing models in high-stakes industries like aerospace, addressing legal barriers and incentivising SME participation through grants and technical assistance.

To solidify the EU's leadership in global industrial data ecosystems, the EC must boost its architecture as the international benchmark. This requires proactive diplomacy to align key partners (e.g., NATO allies and major aviation markets) with EU standards for secure, ethical data exchange. A dedicated task force – co-led by industry leaders and member states – should drive certification schemes for compliant technologies, giving European companies a first-mover advantage in markets adopting Manufacturing-X principles. Concurrently, the EU should invest in training programmes to upskill manufacturing workforces in data-space technologies, ensuring readiness for digital transformation. By embedding its governance model into global supply chains, Europe can secure strategic autonomy, strengthen competitiveness, and position its norms as the foundation for 21st-century industrial innovation.

5 Conclusions and recommendations

This section presents an overview of the study's principal conclusions and recommendations. Based on the evidence and analyses discussed in the previous chapters, it summarizes the main findings and insights derived from each core task. The conclusions and recommendations are organized according to the study's three primary tasks, aligning the guidance with the specific issues and opportunities identified in each area.

5.1 Task 1: Data needs for innovation and the role of synthetic data

5.1.1 Conclusions

Data needs for GenAI companies in Europe

Europe's GenAI ecosystem is growing, but it is still outpaced by the US and China in terms of investment, foundational model development, and market share. Scientific literature shows that Europe's strengths lie in research output and innovation, with a high number of academic publications and a vibrant startup scene. However, most European GenAI companies focus on application layers (e.g., text, image, and speech tools) rather than building foundational models, which require significant compute and financial resources. Stakeholder perceptions confirm that adoption is concentrated among large enterprises in sectors like finance, healthcare, and creative industries, while SMEs face barriers due to limited digital maturity and resources. The market is also characterized by a high dependency on US-based infrastructure and cloud providers, and a lack of large-scale, publicly listed GenAI companies in Europe.

GenAI development in Europe is driven by the need for massive, high-quality, and diverse datasets, with a particular emphasis on multilingual and domain-specific data. Scientific sources highlight that foundation models require vast amounts of unstructured data (text, images, audio, video), but for fine-tuning and specialized applications—common in Europe—smaller, highly curated, and annotated datasets are essential. The literature stresses the importance of linguistic and cultural diversity, as the dominance of English in training data leads to weaker performance in other languages and risks cultural underrepresentation. Stakeholders echo these findings, emphasizing that domain-specific, annotated, and legally usable data are critical for accuracy, compliance, and practical relevance. Structured tabular data and sectoral databases are particularly valued, while modalities like audio and video are less prioritized at this stage.

Legal, technical, and practical barriers limit data access for GenAI companies in Europe. The literature identifies GDPR and copyright as regulatory hurdles, with strict rules on personal data and complex licensing requirements for copyrighted content. Fragmented enforcement and varying interpretations across Member States add to the complexity, especially for SMEs. Technical barriers include the scarcity of high-quality, annotated, and multilingual data, as well as the risk of bias and underrepresentation in training datasets. Stakeholder feedback highlights the difficulty in obtaining domain-specific, multilingual, and multimodal datasets, with SMEs particularly disadvantaged by high compliance costs, lack of standardized frameworks, and reluctance among data holders to share valuable assets due to trust and competitive concerns.

The cost of acquiring, preparing, and managing data is a major operational burden and competitive factor for European GenAI firms. Scientific and market analyses reveal that licensing fees for high-quality datasets, internal data preparation (annotation, cleaning, curation), and regulatory compliance (GDPR, AI Act, copyright) are the most significant cost drivers. The market for AI training data is growing rapidly, with high prices for proprietary content and annotation services. Financial reporting from European GenAI

firms shows that staff costs and external purchases (including data licensing and cloud services) dominate operating expenses, but detailed cost breakdowns are often unavailable due to commercial sensitivity. Stakeholders report that these costs are highly variable and often confidential, with manual curation and annotation remaining unavoidable, especially in sectors with scarce metadata or fragmented data standards.

Common European Data Spaces and related initiatives are widely anticipated to address current data needs and accessibility challenges, but awareness and participation remain limited, especially among SMEs. The literature and stakeholder perceptions agree that data spaces could provide high-quality, curated, and privacy-preserving datasets, particularly for domain-specific and sensitive data. Initiatives like GAIA-X and AI Factories aim to create federated infrastructures for trusted data sharing, but practical engagement is low due to limited knowledge, skills, and resources among SMEs. There is a perceived conservatism within European industry regarding data sharing, and bureaucratic hurdles further limit participation. Stakeholders emphasize the need for robust governance, clear legal frameworks, and incentives for data contribution to make data spaces effective and inclusive.

A multi-faceted approach is needed to overcome data access challenges, combining partnerships, federated learning, synthetic data, open-source models, and targeted public investment. Scientific literature advocates for public investment in high-quality, multilingual, and sector-specific datasets, as well as the development of open-source EU-scale models to reduce dependency on foreign providers. Federated learning and privacy-preserving analytics are seen as promising solutions for enabling collaborative AI development without compromising data sovereignty or privacy. Stakeholders call for clearer legal frameworks, financial incentives for data sharing, and certification schemes (e.g., Europrivacy, GAIA-X labels) to build trust and facilitate compliance. There is also strong support for regulatory sandboxes, targeted support for SMEs, and the integration of data spaces with compute resources to lower entry barriers and accelerate innovation.

Maturity and offer of synthetic data solutions

The European synthetic data market is growing but remains fragmented and new. Most vendors are young companies, established after 2017, and while some have achieved advanced deployments and international reach, many are still in early commercialisation stages. The market is anchored in highly regulated sectors such as finance, insurance, and healthcare, where privacy and compliance are critical drivers for adoption.

Privacy assurance is central to all vendor offerings, but approaches are diverse and there is no single standard for GDPR compliance. Vendors employ a mix of technical safeguards – such as differential privacy, anonymisation, and privacy-by-design engineering – alongside process controls like detailed documentation, privacy metrics, and attack testing. Some rely on external certifications (e.g., Europrivacy), while others use published privacy metrics or expert review. Increasingly, vendors provide “evidence packs” containing reports, metrics, and audit artefacts to support clients’ compliance and governance needs.

Synthetic data is primarily used to enable analytics, machine learning, and data sharing where direct use of real data is restricted. Common applications include fraud detection, risk modelling, healthcare research, and privacy-preserving data sharing. Augmentation of real datasets—such as balancing rare events or simulating scenarios—has become a standard feature. Adoption is strongest in sectors with clear business cases and compliance incentives, but misconceptions about synthetic data and internal governance requirements often slow down projects.

Fragmented assurance practices and legal uncertainty are significant barriers to adoption. Buyers face difficulties comparing solutions due to varied privacy metrics and

the lack of a common evidentiary baseline. Legal ambiguity around what qualifies as “anonymous” synthetic data under GDPR leads to cautious, lengthy procurement processes and slows cross-partner projects.

Certification costs and market pressure present additional challenges. Formal certifications are valued, especially in healthcare, but are expensive and resource-intensive, creating barriers for SMEs. The US market is more dynamic, with lighter privacy rules and faster adoption, prompting some European vendors to shift resources abroad in search of growth opportunities.

Awareness gaps and misconceptions about synthetic data persist among buyers and stakeholders. Many still conflate robust synthetic data with generic “fake data” or LLM-generated content, which elongates sales cycles and complicates procurement. Education and clear communication remain essential parts of vendor engagement.

The provision of structured privacy and utility reports (“evidence packs”) is becoming a market norm. This supports auditability and trust, and is increasingly expected by buyers, particularly in sensitive sectors. Controller-side execution—processing within the client’s perimeter—is consolidating as a baseline expectation, further reducing compliance risk.

There is a clear push for standardisation and clearer regulatory guidance. Vendors and buyers alike call for explicit GDPR criteria for anonymity, standardised evaluation practices, and recognition of transparent privacy metrics. Modular or proportionate certification schemes are also recommended to reduce costs and barriers for smaller providers.

Sectoral growth is expected to remain strongest in finance and healthcare, with telecoms and public sector adoption dependent on clearer legal guidance and more affordable certification. Hybrid use cases, where synthetic data augments real data, and automation of privacy evaluation are likely to become more common as the market matures.

5.1.2 Recommendations

The following policy recommendations are directly linked to the data needs of GenAI companies in Europe, as identified in the preceding analysis. They are designed to address and overcome the key challenges outlined above—including barriers to data access, legal uncertainty, and the scarcity of high-quality, multilingual, and domain-specific datasets. By implementing these targeted measures, the European Union can foster a more competitive, innovative, and inclusive GenAI ecosystem that is aligned with European values and strategic objectives.

Accelerate the operationalisation of European Data Spaces and AI Factories

To address persistent data access barriers for GenAI companies, the report recommends moving from conceptual frameworks to fully operational Common European Data Spaces (CEDS) and AI Factories. This requires targeted investment in technical infrastructure, governance, and support services, with a focus on making these resources accessible to SMEs. The report highlights the need for interoperability, robust governance (including clear IP licensing and access controls), and incentives for data contribution. By prioritizing the practical rollout of data spaces and AI Factories, the EU can foster high-quality, domain-specific datasets and reduce reliance on non-EU providers, supporting a more competitive and innovative GenAI ecosystem.

Clarify and harmonize the legal framework for data access and use

Legal uncertainty—especially around GDPR, copyright, and fragmented enforcement—remains a major obstacle for GenAI development in Europe. The report calls for swift, practical guidance on the implementation of the AI Act and GDPR, standardized licensing

models, and the promotion of certification schemes (e.g., Europrivacy, GAIA-X labels). It also recommends the adoption of transparent privacy-evaluation metrics and consistent enforcement across Member States. These measures will lower compliance costs, reduce legal risk, and enable SMEs to innovate with greater confidence, while ensuring privacy and IP rights are respected throughout the EU.

Direct public investment toward high-quality, multilingual, and domain-specific datasets.

The report identifies the lack of high-quality, multilingual, and sectoral datasets as a critical bottleneck for European GenAI. It recommends that public funding be allocated to the creation and curation of open datasets in underrepresented languages and domains, and that public-private partnerships be used to unlock access to sensitive or privately held data. The development of open-source EU-scale models and integration of these datasets into data spaces and AI Factories are also emphasized. Strategic investment in these areas will enable the development of more accurate, inclusive, and relevant AI models, supporting Europe's digital sovereignty and competitiveness.

Maturity and offer of synthetic data solutions

Clarify GDPR scope and adopt a more permissive approach. Companies consistently call for clearer, more practical, and favourable rules for synthetic data under GDPR. They urge authorities to define when synthetic data qualifies as “anonymous” and what evidence is sufficient, providing explicit guidance on lawful bases for synthesis and sharing, especially in cross-partner projects. Standardised evaluation practices—such as recognising transparent privacy metrics and attack tests—are needed to normalise a minimum evidentiary baseline and make compliance more portable across projects.

Ensure consistent enforcement and raise awareness. There is a strong need for consistent GDPR enforcement across sectors and for raising awareness of robust privacy engineering. Companies highlight the importance of myth-busting and authoritative explanations to distinguish mature solutions from superficial claims, ensuring privacy is treated as a design discipline rather than a checkbox.

Support innovation and fair competition for EU companies. Practical incentives—such as outcome-linked funding for evaluation and R&D, and proportionate, accessible certification schemes—are needed to help EU vendors compete internationally, particularly against US firms. Policy should level the playing field by reducing structural disadvantages, supporting methodological breadth, and recognising evaluation artefacts in public sector projects.

5.2 Task 2: Using automated data reporting mechanisms for demonstrating regulatory compliance

5.2.1 Conclusions

Agriculture

The adoption of automated data reporting mechanisms in European agriculture is progressing, yet the sector remains characterized by a **strong reliance on traditional compliance methods**. While digitalization is advancing, most farms continue to depend on **paper-based records** and spreadsheets for meeting regulatory requirements under frameworks such as the CAP, the Deforestation Regulation, and the Plant Protection Products Regulation. These standard techniques are valued for their adaptability and low direct costs, but they impose a significant administrative burden, are labour-intensive, and are prone to errors and inefficiencies as compliance demands grow more complex and frequent.

Advanced digital tools, such as **digital logbooks and Farm Management Information Systems (FMIS)**, are increasingly present, particularly among larger and more technologically mature farms. These systems enable structured, real-time data capture, integration with geospatial and sensor data, and improved traceability for compliance. However, **their uptake is constrained by high licensing and subscription costs, digital skill gaps, and persistent interoperability issues with national CAP portals and farm machinery**. Fully automated solutions, including blockchain-based traceability platforms, **remain marginal and largely experimental**, with adoption limited by technical, financial, and trust-related barriers.

The analysis reveals that the transition from manual to digital and automated compliance tools is gradual and uneven. **Barriers to adoption are multifaceted**: resource intensity and labour demands remain high for standard tools, while advanced systems introduce new challenges related to cost, digital skills, and integration. **The digital divide is particularly pronounced**, with smaller and older farms facing significant obstacles in adopting advanced solutions, risking exclusion from increasingly digitalized compliance regimes. Larger operations, by contrast, are better positioned to benefit from streamlined digital processes and early access to innovation.

Cost-benefit assessments indicate that advanced digital logbooks currently offer the best balance between efficiency gains and manageable costs, making them accessible to a broad range of farms. **FMIS deliver the most comprehensive compliance support** but are financially viable primarily for large farms. Fully automated tools promise transformative benefits in traceability and real-time monitoring, but their high costs and technical complexity limit their practical relevance for most farms at present.

Despite the promise of digitalization, the persistence of standard methods underscores the need for scalable, interoperable, and user-friendly solutions that can bridge the digital divide and ensure inclusive compliance across the sector. **Farmers are more likely to adopt digital tools that directly support core agronomic activities**, with compliance functionalities often seen as a by-product rather than the main incentive for digitalization. The findings suggest that, under current conditions, advanced digital logbooks represent the most effective and scalable solution, with FMIS and fully automated tools holding promise for the future as barriers are progressively addressed. The evolution of automated data reporting in agriculture will depend on targeted support for digital skills, investment in rural infrastructure, and the development of interoperable platforms that lower entry barriers and foster broad participation.

Automotive

The automotive sector in Europe is navigating a complex landscape of regulatory compliance and reporting, shaped by an extensive array of EU regulations covering type-approval, emissions, chemicals, and sustainability. Despite the sector's high degree of digitalization in manufacturing and supply chain management, **compliance and reporting activities remain dominated by standard and advanced techniques**, with manual data entry, spreadsheets, and specialized software suites as the prevailing tools. While advanced enterprise resource planning (ERP) systems and sector-specific compliance **platforms are increasingly used to manage and automate parts of the reporting process**, full automation—such as robotic process automation or blockchain-based certification—remains rare and largely confined to pilot projects.

The transition toward more automated compliance is hindered by several persistent challenges. The sheer volume and complexity of legal obligations, fragmented reporting requirements across Member States, and the **lack of harmonized digital data standards** create significant administrative burdens for manufacturers. **Integration of compliance tools with existing IT infrastructure** is often difficult, and the absence of unified reporting platforms means that companies must navigate a patchwork of systems

and formats. Financial and technical barriers, particularly for smaller suppliers, further limit the adoption of advanced and automated solutions.

Cost-benefit assessments indicate that while advanced compliance software can deliver significant efficiency and accuracy gains, **these benefits are offset by high capital and operating costs, ongoing maintenance, and the need for specialized training. Automated techniques promise further efficiency and real-time monitoring**, but their high implementation costs and the continued need for human oversight limit their practical impact. **Stakeholders consistently emphasize that the main challenge** lies not in completing templates, but in **collecting, validating, and integrating relevant data from diverse sources**.

Overall, the automotive sector demonstrates a **cautious but steady progression toward digital and automated compliance**, with advanced tools offering clear operational benefits for larger firms. However, the persistence of manual and semi-automated methods underscores the need for harmonized standards, interoperable platforms, and targeted support to ensure that the benefits of automation are accessible across the entire value chain. The future of compliance in the sector will depend on continued investment in digital infrastructure, regulatory harmonization, and the development of scalable solutions that can accommodate the sector's complexity and diversity.

Healthcare

The healthcare sector in the EU is experiencing a **gradual digital transformation in regulatory compliance and reporting**, driven by the introduction of new regulations such as the Medical Devices Regulation (MDR), In Vitro Diagnostic Medical Devices Regulation (IVDR), and the Clinical Trials Regulation (CTR). While **paper-based methods have largely been phased out, spreadsheets and in-house digital tools remain widely used** for managing technical documentation, incident reporting, and regulatory submissions. **Specialised compliance software suites are increasingly adopted**, particularly by larger organisations, offering centralised management of regulatory data, risk assessments, and audit documentation.

Despite these advances, **the sector faces significant challenges in achieving fully automated compliance**. High licensing and integration costs, technical complexity, and persistent interoperability issues with national and EU-level platforms limit the widespread adoption of advanced and automated tools. Smaller providers, in particular, struggle with the financial and human capital demands of digital compliance systems. The lack of harmonised data standards and the continued reliance on manual data entry and validation further constrain efficiency gains.

Cost-benefit analyses reveal that specialised compliance software delivers substantial improvements in efficiency and accuracy, streamlining documentation and audit readiness. However, **these benefits are tempered by ongoing expenses** for licensing, training, and system integration. **Fully automated solutions**, while promising in terms of scalability and real-time monitoring, **remain experimental and resource-intensive**, making them accessible primarily to large, technologically advanced organisations.

The findings underscore the importance of interoperability, digital skills development, and targeted investment in digital infrastructure to support the sector's transition. While advanced tools are increasingly seen as essential for managing the growing complexity of healthcare compliance, the sector remains at an intermediate stage, with most organisations operating between spreadsheet-based workarounds and partially integrated compliance platforms. The **evolution towards automation will depend on regulatory clarity**, investment in digital skills, and the development of user-friendly, interoperable solutions that can meet the diverse needs of healthcare providers across the EU.

Chemicals

The European chemicals sector operates within one of the most demanding regulatory environments, with companies required to comply with a wide range of EU legislation, including REACH, CLP, the Biocidal Products Regulation, and the Corporate Sustainability Reporting Directive. **Compliance and reporting activities are still predominantly managed through standard techniques**, such as spreadsheets and manual data entry, especially among small and medium-sized enterprises. **Advanced tools, including ERP systems and specialised compliance software, are more common among larger firms**, supporting the integration of regulatory data across business functions and enabling partial automation of reporting tasks.

Despite the availability of advanced digital solutions, **the sector faces persistent barriers to full automation**. The complexity of regulatory requirements, the need for harmonisation across supply chains, and the lack of interoperability with public authorities' systems limit the effectiveness of automated tools. **Smaller firms, in particular, struggle with the technical and financial demands of digital compliance**, while larger companies benefit from centralised databases and more sophisticated IT infrastructure.

Cost assessments indicate that **standard techniques are labour-intensive and prone to errors**, while **advanced tools shift costs towards training**, licensing, and system integration. **Automated solutions offer significant efficiency gains** and real-time monitoring but **require substantial upfront investment and ongoing maintenance**. The lack of harmonised reporting standards and regulatory acceptance of automated outputs further constrains adoption.

The sector's experience highlights the **need for harmonised digital reporting requirements**, regulatory acceptance of automated solutions, and targeted support for digital literacy and infrastructure, particularly for SMEs. While advanced and automated tools hold promise for improving efficiency and accuracy, their widespread adoption will depend on overcoming technical, organisational, and regulatory barriers, and on ensuring that digital solutions are accessible and practical for companies of all sizes.

Energy

The energy sector in the EU is undergoing a gradual transition towards digital and automated compliance, shaped by a complex regulatory landscape that includes the Energy Efficiency Directive, REMIT, the EU Emissions Trading System, and the Industrial Emissions Portal Regulation. **Standard compliance methods**, such as spreadsheets, manual data entry, and paper-based records, **remain prevalent, particularly for smaller operators**. **Advanced compliance** software and integrated digital platforms are **increasingly used by larger firms** to manage data collection, reporting, and audit readiness, but full automation is still rare.

The adoption of advanced and automated tools is hindered by **several persistent challenges**. Fragmented reporting channels, divergent templates and identifiers, and evolving technical specifications create significant administrative burdens and increase the risk of errors. **Integration with legacy IT systems**, high upfront and ongoing costs, and the need for specialised digital skills further limit the uptake of digital solutions, especially among smaller providers. **Regulatory uncertainty** regarding the acceptance of automated outputs and the lack of harmonised standards add to the complexity.

Cost-benefit analyses show that while **advanced and automated tools can deliver significant efficiency gains**, real-time monitoring, and improved accuracy, these benefits are offset by high capital and operating costs, technical complexity, and the need for ongoing training and support. **Labour remains a major cost driver** for standard methods, while advanced solutions shift costs towards infrastructure, licensing, and integration.

The findings emphasise the importance of **harmonised reporting standards**, investment in digital infrastructure, and targeted support for digital skills development to enable a

broader transition to automation. While advanced compliance tools are increasingly recognised as essential for managing the sector's regulatory complexity, their adoption remains uneven, with smaller operators at risk of being left behind. The future of compliance in the energy sector will depend on continued investment in interoperable platforms, regulatory clarity, and the development of scalable solutions that can accommodate the sector's diversity and evolving regulatory demands.

5.2.2 Recommendations

Sector-specific recommendations are detailed in the main body of this report, reflecting the challenges and opportunities within agriculture, automotive, healthcare, chemicals, and energy. The summary below draws out the most salient cross-sectoral lessons, offering a strategic perspective on how the EU can address common barriers and accelerate digital transformation in compliance and reporting across these key industries.

Advance interoperable digital infrastructure and standards

A recurring challenge across all sectors is the fragmentation of digital systems and the lack of interoperability between platforms used for compliance, reporting, and audit readiness. The EU should prioritise coordinated investment in interoperable digital infrastructure, including the adoption of common data standards, open APIs, and machine-readable reporting templates. This is particularly urgent for agriculture, healthcare, chemicals, automotive, and energy, where legacy systems and bespoke solutions impede efficient data exchange and increase administrative burden. Special support should be directed towards SMEs, ensuring that digital tools are accessible and scalable for organisations of all sizes.

Harmonise and clarify regulatory frameworks for digital compliance

Legal uncertainty and inconsistent enforcement of digital compliance requirements remain significant obstacles, especially in sectors governed by complex EU and national regulations such as CAP (agriculture), MDR/IVDR (healthcare), REACH (chemicals), REMIT and ETS (energy), and type-approval and sustainability directives (automotive). Policy actions should be focused on providing clear, harmonised guidance on the interpretation and implementation of these frameworks, promote the use of standardised reporting templates and certification schemes (such as Europrivacy and GAIA-X labels), and ensure that digital and automated outputs are recognised by authorities across Member States. This will reduce compliance costs, lower legal risk, and give organisations the confidence to invest in digital solutions.

Invest in digital skills and capacity-building for inclusive adoption

The digital divide is a persistent barrier, particularly for smaller operators and older workforces in agriculture, healthcare, and chemicals. The EU and Member States should invest in targeted training programmes, digital literacy initiatives, and technical support to build capacity for the adoption of advanced compliance tools. By empowering all actors – regardless of size or digital maturity – to participate in the digital transition, the EU can ensure that the benefits of automation and digitalisation are shared across the entire value chain.

Support scalable, modular, and sector-specific digital solutions

A one-size-fits-all approach is not effective given the diversity of operational contexts and regulatory requirements across sectors. Policy actions should encourage the development and uptake of scalable, modular digital solutions that can be tailored to the needs of both large enterprises and SMEs. In agriculture and healthcare, this means promoting user-friendly digital logbooks and compliance software; in automotive, chemicals, and energy, it involves supporting the integration of advanced ERP systems, compliance modules, and automated reporting tools that can adapt to evolving regulatory demands. Public funding

and procurement should prioritise solutions that are interoperable, open, and capable of scaling as regulations and technologies evolve.

Foster Public-Private Partnerships and Open Data Ecosystems

Unlocking the full potential of digital compliance and reporting requires collaboration between public authorities, industry, and technology providers. Public-private partnerships and the development of open data ecosystems should be promoted, particularly in the chemicals, energy, and automotive sectors, where complex supply chains and the need for data sharing are critical for regulatory compliance and sustainability objectives. This approach involves incentivising the creation and curation of high-quality, sector-specific datasets, supporting the advancement of open-source compliance tools, and enabling secure data sharing through trusted intermediaries and data spaces.

Embed proportionality and flexibility in digital regulation

Digital compliance requirements should be risk-based and proportionate, with phased implementation and support for micro-enterprises, especially in agriculture and chemicals, to avoid excluding smaller actors from digital transformation. Regulatory sandboxes, pilot projects, and exemptions for micro-enterprises can help ensure that digital transformation is inclusive and does not inadvertently penalise those with fewer resources.

5.3 Task 3: International aspects on the data economy

5.3.1 Conclusions

The **global analysis** indicates that the EU is in a strong position to establish a distinct model of data governance, separate from US market-driven and Chinese state-driven approaches. Its regulatory influence, combined with sector-specific initiatives (like the EU Deforestation Regulation and Battery Regulation), gives it leverage to shape global data-sharing norms. These provide practical leverage for embedding EU standards abroad. At the same time, the EU faces persistent challenges: a lack of trust in data sharing across borders, geopolitical tensions, and the rapid pace of technological change, particularly in AI. While the EU promotes a rights-based, human-centric approach as an alternative to both the US's market-driven model and China's state-centric system, success will depend on moving from principles to implementation. This requires aligning regulatory ambition with technical infrastructures and ensuring that the EU's "third way" is not just a political narrative but a globally credible framework.

Within the EU's regulatory ambition, the handling of personal and non-personal **sensitive and highly sensitive data** is central to strengthening safeguards, enabling international cooperation, and ensuring secure cross-border transfers. While sensitive personal data is extensively defined and regulated under the GDPR and comparable laws in the eight selected countries, the categories of non-personal sensitive or highly sensitive data remain largely undefined, both in the EU and globally. This gap presents an opportunity for the EU to take the lead and set the standard internationally by adopting clear definitions of non-personal sensitive and highly sensitive data as well as their related safeguards. At the same time, the EU should continue to refine safeguards for personal sensitive data and explore new frameworks to ensure balance between protection and innovation in data use.

The **overall conclusion of the use cases** is that the EC should prioritise establishing unified technical and regulatory frameworks across its digital initiatives – spanning e.g., battery passports, PNR data exchanges, and federated digital twins – to boost the EU's leadership in ethical and privacy-preserving innovation, strategic autonomy, and global standard-setting. By harmonising interoperability standards (e.g., Gaia-X, IDSA) with robust safeguards for data sovereignty, and IP, the EU can foster resilient supply chains,

secure cross-border collaboration, and competitive industries. Accelerating Public-Private Partnerships, scaling SME support, and embedding privacy-enhancing technologies will promote inclusive participation and operational efficiency, while proactive diplomacy and certification schemes will help to export EU norms globally. Dynamic governance, stakeholder engagement and regulatory sandboxes are critical to balancing agility with compliance, enabling the EU to lead in sustainable digital transformation, secure industrial ecosystems, and rights-respecting innovation which are the key pillars for advancing the Green Deal and strengthening economic resilience.

5.3.2 Recommendations

The key findings and recommendations arising from the analysis of the international dimension of data governance and cross-border collaboration opportunities with non-EU countries are summarised below. It should be noted that these policy recommendations are presented in greater detail in section 4.4.

Recommendations based on the global analysis

Leverage regulatory gravity

The EU should leverage and exploit its regulatory influence by expanding the “Adequacy Alliance” through sector-specific agreements (e.g., Brazil’s finance or South Korea’s smart cities). In parallel, the EU should promote Gaia-X and IDSA-compliant architectures as default solutions, funding open-source pilots, as well as scale EU-controlled cloud and AI platforms to reduce reliance on external providers.

Pioneer new governance models

As part of its data strategy, the EU should pioneer new and alternative data governance models through the development of hybrid frameworks inspired by countries such as Japan to bridge GDPR-compliant and innovation-driven states, in addition to the development of pilot “data trustees” models to address B2B/B2G trust gaps. Abroad, the EU could lead a global AI governance with likeminded partners such as South Korea and Japan, harmonising ethics standards with innovation policies. Internally, the EU could adapt its GDPR for innovation through targeted regulatory sandboxes in AI and health tech.

Transform friction into partnership

The search for a European data model has led to occasional frictions with third countries, a challenge that can be turned into opportunity with the promotion of international partnerships. As such, the EU should launch a “Traceability for Sustainable Trade” initiative with partners such as Brazil, offering technical support and co-funding. Furthermore, the EU should advance multilateral data diplomacy via G7/OECD to promote GDPR-inspired norms as well as develop transnational data spaces in key sectors like healthcare, using hybrid approaches to balance local storage with cross-border R&D.

Prioritise PETs for security cooperation

The EU should initiate PET-based pilots with likeminded countries such as the UK and Australia for critical information systems including the next-generation Passenger Name Record (PNR) systems. In addition, the EU should balance rights and competitiveness in trade deals by allowing limited deregulation in non-sensitive sectors while enforcing GDPR in critical areas.

Scale and structure R&D funding

Strategy without the right funding tends to have a rather limited impact. To avoid this, the EU should increase its long-term investment in AI and data spaces through funding tools such as Horizon Europe, Digital Europe and future programmes. Research in key areas such as generative AI, PETs, federated learning and cross-border data infrastructures

should be prioritised. A prioritisation that should be promoted by supporting pilot projects and testbeds to bridge research and deployment and internationalising R&D partnerships with likeminded countries to diversify expertise and markets.

The actions presented above should be implemented following a structured strategic timeline:

- Short Term: Finalise adequacy agreements and expand GDPR-aligned data spaces with likeminded global partners.
- Medium Term: Co-design AI ethics frameworks with likeminded partners across the globe to balance existing dominant models.
- Long Term: Establish the EU as the global hub for secure, rights-based innovation via initiatives such as GAIA-X and PET-driven standards.

Recommendations based on the analysis of Sensitive and Highly-Sensitive Data Categorisation

Reinforce the legal concept of sensitive personal data internally and abroad

The EU should reinforce the protection of sensitive personal data both internally and abroad, mainly through its support for the international alignment of protection measures between the EU and those countries that currently have a formal categorisation of sensitive personal data. To this end, the adequacy decisions stand out as a key legal and diplomatic tool to precisely promote international alignment as well as cooperation. Internally, new safeguard frameworks should be explored to reinforce protection including a clear and auditable EU legal framework for de-identification and certified anonymisation.

Promote the legal codification of non-sensitive personal data internally and abroad

The identified lack of comprehensive legal codifications of non-sensitive personal data both internally and abroad should be addressed with the establishment and adoption by the EU of a formal cross-cutting definition on sensitive non-personal data. In line with this, the possibility of developing a General Non-Personal Data Protection Regulation with formal definitions and categorisations should be explored. If this option were to be pursued, the policy-making process could be enhanced if the development of the Regulation were coordinated with likeminded global partners, or at least the establishment of common definitions and categorisations.

Reinforce the protection of highly sensitive data in the EU

As with sensitive non-personal data, highly sensitive data remains largely uncoded and undefined. In such context, the EU should establish and adopt a formal definition on special categories of non-personal data identified as highly sensitive. In addition to this, the protection of this type of data should be reinforced incorporating frameworks such as the given by the EUCI into the general protection of non-personal sensitive and highly sensitive data. The definition of specific requirements and safeguards to be applied to data controllers and processors when handling non-personal sensitive and highly sensitive data should also be established.

Recommendations based on the use case analyses

The six use cases converge on three pillars: a **unified regulatory and technical framework, interoperable standards** driven by multi-stakeholder governance, and **close collaboration with industry, academia and standardisation bodies**. Achieving these aims hinges on balancing EU-wide consistency with Member State flexibility, and on combining top-down policy action with bottom-up innovation support.

The recommendations for the **Cross-Border Digital Battery Passport** call for an accelerated roll-out of a single ecosystem that embeds interoperable standards for battery passports and digital twins within Mobility and Green Deal Data Spaces. Safeguards for

intellectual property and data sovereignty should be institutionalised through secure federated sharing protocols and standardised IP agreements, drawing on national pilots such as the German Battery Passport Project. Targeted funding and capacity-building programmes must scale digital infrastructure for SMEs, while a dynamic governance platform with regulatory sandboxes and public-private partnerships iteratively refines standards. Finally, the EU should export its frameworks via trade agreements and multilateral collaborations, harmonising core requirements yet permitting local adaptation.

For **Federated Learning in Rare Disease Genomics**, the priority is to pilot the model within a regulatory sandbox to enable rapid progress and ensure later alignment with the European Health Data Space. Strategic policy direction must harmonise data standards, offer incentives for participation and reinforce infrastructure through Horizon Europe and partnerships such as South Korea's Pillar II. Support for collaborative, bottom-up projects will drive stakeholder engagement, laying the groundwork for positioning the EU as a leader in ethical, data-sovereign AI for health, and possibly extend the use case on privacy-preserving AI to cancer genomics, mental health and paediatrics.

In the scope of **safety data exchange for Autonomous Vehicles**, recommendations centre on governance and trust rather than purely technical breakthroughs. A "data trustee" model should clarify usage rules, enforce non-disclosure agreements and operate a neutral data intermediary – potentially under EASA or the Mobility Data Space – to allow safety-critical scenario reconstruction without exposing proprietary algorithms. Engagement with South Korea and Japan will leverage their established AV data-management agencies and reinforce the EU's convening role. Regulatory sandboxes will facilitate experimentation with non-punitive data sharing in a competitive global market.

The **Deforestation-Free Agricultural Supply Chain** use case urges the EU to build on existing traceability systems and legal frameworks, aligning the timing of the EU Deforestation Regulation with domestic schemes such as Brazil's. Coordinated financial and technological support from the EU and major commodity traders should incentivise farmer compliance, while an EU intermediary agency enhances data sharing among enforcement and confidence authorities. Parallel cooperation with China can channel lessons from the EU's traceability, monitoring and evaluation efforts into Chinese environmental governance. Such bilateral engagement paves the way for mutual recognition of sustainability standards.

Next-Generation Passenger Name Record exchange must rest on a harmonised legal and technical framework embedding privacy-enhancing technologies – anonymisation protocols, strict access controls and independent judicial oversight. Revised agreements with the US, Canada and Australia should formalise these safeguards, while pilots of secure multiparty computation and federated learning demonstrate real-time GDPR-compliant data sharing. A multidisciplinary task force, regulatory sandboxes and public awareness campaigns will maintain transparency, and continuous impact assessments with stakeholder feedback will keep the framework adaptive. Promotion through multilateral forums will cement Europe's leadership in ethical security innovation.

Finally, **Federated Digital Twins for Aerospace Supply Chains** require scaling national initiatives like Germany's COOPERANTS within an EU-wide framework underpinned by Gaia-X and IDSA interoperability standards. Targeted funding should support predictive maintenance systems and visibility tools via public-private partnerships, and regulatory sandboxes must test high-stakes data-sharing models while incentivising SME participation. Proactive diplomacy with NATO allies and key aviation markets will align global standards, and an industry-state task force should oversee certification schemes. Investment in workforce training on data-space technologies completes the blueprint for European leadership in resilient aerospace ecosystems.

Annex 1 Task 1a – Reference list

Source	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
Balland, P.-A., R. Balling, V. Borre, B. van der Borg, C. De Bruyckere, M. de Fombelle, A. Godefroy, A. Klenner, T. Lenz, V. Maini, A. Minne, F. Nagle, A. Neff, P. Neuhäusler, F. Novelli, V. Strop, R. W. Van der Veen, M. Vest, F. W. Van der Veen, S. Wiegand, J. Woerner, R. W. Van der Veen, & J. Wolslegel. (n.d.). <i>Bridging the AI Gap in Europe: How Europe can overcome its structural dependencies on the US and China in generative AI</i> .	✓	✓	✓	✓	✓	✓
Bandi, A., Adapa, P.V.S.R. and Kuchi, Y.E.V.P.K., 2023. The power of generative ai: A review of requirements, models, input-output formats, evaluation metrics, and challenges. <i>Future Internet</i> , 15(8), p.260.	✓					
Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Large Language Models Be Too Big? <i>Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency</i> , 610–623.						✓
Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). <i>On the opportunities and risks of foundation models</i> . arXiv preprint arXiv:2108.07258.	✓		✓	✓		
Bonawitz, K., Eichner, H., Grieskamp, W., Huba, D., Ingerman, A., Ivanov, V., ... & Roselander, J. (2019). Towards Federated Learning at Scale: System Design. <i>SysML</i> .						✓
Chen, J., & Mueller, J. (2024). CLEAR: A Confidence-Based Data Curation Pipeline for LLM Instruction Tuning. <i>arXiv preprint arXiv:2401.07767</i> .	✓					
Confederation of European Data Protection Organisations. (2023). <i>Generative AI and data protection: The DPO's perspective</i> .			✓	✓		✓
Confederation of European Data Protection Organisations AI and Data Working Group. (2023). <i>Generative AI in practice: Guidance for Data Protection Officers</i> .			✓			
De Wilde, P., Arora, P., Buarque de Lima Neto, F., Chin, Y., Thinyane, M., Stinckwich, S., ... & Marwala, T. (2024). <i>Synthetic Data for AI: Benefits and risks in the Global South</i> . United Nations University Institute in Macau.			✓	✓	✓	✓

Deloitte. (2024). <i>A look at the accounting for generative AI.</i>	✓	✓		✓		✓
Díaz-Rodríguez, N., Del Ser, J., Coeckelbergh, M., de Prado, M. L., Herrera-Viedma, E., & Herrera, F. (2023). <i>Towards a trustworthy AI in the European Union: A comprehensive analysis of the European Artificial Intelligence Act.</i> arXiv preprint arXiv:2303.04584.				✓	✓	✓
EIT Digital. (2024). <i>Generative AI in Europe: Challenges and opportunities for digital sovereignty.</i>	✓	✓	✓	✓	✓	✓
ERT, European Round Table for Industry. (n.d.). <i>Manifesto: Boosting Europe's industrial competitiveness and strategic autonomy with Generative AI.</i>	✓	✓	✓	✓	✓	
European Commission. (2024). <i>Guidance for Researchers on Generative AI.</i>			✓			
European Copyright Society. (2025). <i>Text and Data Mining (TDM) and the EU Copyright Framework after the AI Act.</i>			✓	✓		✓
European Data Protection Board. (2024). <i>Information on the processing of personal data in the context of AI models.</i>	✓	✓	✓			
European Data Protection Supervisor. (2024). <i>Generative AI and the EUIs.</i>	✓	✓				
Feuerriegel, S., Hartmann, J., Janiesch, C. et al. (2024). <i>Generative AI. Bus Inf Syst Eng</i> 66, 111–126.	✓		✓			
Fonteneau, F., & Anvar, E. (2025). <i>Generative AI for official statistics: Challenges and opportunities.</i> OECD Artificial Intelligence Working Paper, No. 1.		✓	✓	✓	✓	✓
Google. (2023). <i>Unlocking Europe's AI Potential: Powering Innovation & Growth.</i>	✓		✓	✓	✓	✓
Guo, X., & Chen, Y. (2024). <i>A Survey of Synthetic Data Generation using Large Language Models.</i> arXiv preprint arXiv:2402.17658.			✓	✓	✓	✓
High-Level Group for the Modernization of Official Statistics. (2025). <i>Using Generative Artificial Intelligence (GenAI) in Official Statistics: Guidelines, Use Cases, Challenges and Future Perspectives.</i>	✓	✓	✓	✓		
Hutchinson, P., Goll, F., & Mügge, C. (2024). <i>Generative AI in Europe: Challenges and Opportunities for European Startups.</i>	✓	✓	✓	✓	✓	✓

Implement Consulting Group. (2024). <i>Generative AI in Europe: State of play, opportunities and challenges</i> .	✓	✓	✓	✓	✓	
Jo, A., 2023. <i>The promise and peril of generative AI</i> . <i>Nature</i> , 614(1), pp.214-216.	✓					
Le Picard, H. (2025). <i>Generative AI: Europe's New Artificial Dependency</i> . Bruegel Policy Brief.	✓			✓	✓	✓
Manduchi, L., Pandey, K., Meister, C., Bamler, R., Cotterell, R., Däubener, S., ... & Fortuin, V. (2024). <i>Leveraging Domain Knowledge for Deep Generative Models: A Survey</i> . arXiv preprint arXiv:2403.08581.		✓		✓		
Massey, J., Simperl, E., Narayan, V., Airan, A., & Kapoor, A. (2024). <i>Co-generation in AI: Who owns what?</i> . University of Southampton.		✓	✓	✓	✓	✓
Microsoft Corporation. (2024). <i>Responsible AI in the Public Sector: Guidance on Using Generative AI Solutions in a Compliant Manner</i> .	✓	✓	✓			✓
Monaco, D., & Nurski, L. (2025). <i>Generative AI and the future of work: A global outlook</i> . ILO.	✓		✓	✓		
Myers, D., Mohawesh, R., Chellaboina, V.I. et al. (2024). Foundation and large language models: fundamentals, challenges, opportunities, and social impacts. <i>Cluster Comput</i> 27, 1–26.	✓	✓	✓	✓		
Organisation for Economic Co-operation and Development. (2024). <i>The OECD AI Principles: Next steps and key challenges</i> .	✓	✓	✓	✓		
Organisation for Economic Co-operation and Development. (2025). <i>Data scraping for AI: Issues and policy considerations</i> .	✓	✓	✓	✓	✓	✓
Outeda, C. C. (2024). <i>The European AI Act: Navigating the regulatory landscape</i> .				✓	✓	
Piccirilo, F. (2025). <i>Ethical AI Framework Enhanced for LLMs: A Case Study on the EU AI Act</i> .	✓	✓	✓	✓	✓	
Publications Office of the European Union. (2023). <i>Common European Data Spaces: A new approach to data sharing</i> .	✓	✓	✓	✓	✓	
Román, D. A. (2025). <i>AI Factories and Data Spaces: A Synergistic Approach</i> . EuroHPC Joint Undertaking.	✓	✓	✓	✓	✓	✓

Ruokonen, M. (2025). <i>Applied AI for Europe's Growth and Productivity</i> . BusinessEurope.	✓	✓			✓	✓
Schola Europaea, Office of the Secretary-General, Pedagogical Development Unit. (2025). <i>A framework for the use of generative artificial intelligence in the European Schools</i> .			✓	✓		
Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024). <i>The Curse of Recursion: Training on Generated Data Makes Models Forget</i> . arXiv preprint arXiv:2305.17493.	✓	✓	✓	✓	✓	✓
Stanford Institute for Human-Centered AI. (2025). <i>Artificial Intelligence Index Report 2024</i> .	✓	✓	✓			
Tammes-Peters, H. (2025). <i>Guidelines for the use of generative AI in public administration</i> .	✓	✓	✓	✓		✓
University of Turin Law School & Nexa Center for Internet & Society. (2025). <i>Generative AI and Copyright: A study on the impact of generative AI on creative industries</i> .			✓			
Wachter, S., Mittelstadt, B., & Floridi, L. (2017). Why a Code of Ethics for AI won't work. <i>Minds and Machines</i> , 27(3), 425-442.			✓	✓		

Annex 2 Task 1b – Mapping of synthetic solution

Name	Location	Creation of synthetic data	Domains of activity	Type of services	Data source	Commercial maturity	GDPR compliance	Corporate data	Estimated customers / users	Synthetic generation category
MOSTLY AI	Vienna, AT	Generative models for tabular/text; optional Differential Privacy (DP) training; privacy-by-design safeguards (rare-category & extreme-value protection, overfitting prevention)	Finance/banking, insurance, telecom/marketing analytics; enterprise data access & testing	Data Intelligence Platform (self-hosted/cloud), no-code UI & API/SDK; open-source Synthetic Data SDK	Trains on customer datasets in their environment; outputs synthetic data with no 1:1 linkage	Founded 2017; productized platform + SDK; public refs (Swiss Post, Erste Group); DP feature added 11/2024	Privacy-by-design; DP option with epsilon accounting; safeguards against overfitting, rare categories, extremes	Employees: 35+ (site); turnover: not disclosed	Not publicly disclosed; named refs include Swiss Post & Erste Group	B (using real-world data); B1/B2 with B3 support
Sarus	Paris, FR	Privacy layer for analytics/ML on non-visible data; DP-protected outputs; DP-trained synthetic data; automatic privacy-first query rewriting	Finance, healthcare, public sector research / LLM with privacy controls	On-prem/cloud privacy middleware; secure querying; DP safeguards; optional synthetic datasets	Operates on source systems without exposing rows; returns DP-protected aggregates or DP-fitted synthetic	Founded 2019; YC-backed seed; PETs focus	Differential Privacy for outputs; zero-trust/least privilege; re-identification risk bounded	Employees: 11–50 (Crunchbase); turnover: not disclosed	Not publicly disclosed; sector pilots and customer quotes on site	B (using real-world data); B1/B2; B3 supported
Syndata	Stockholm, SE	Generative AI (single & multi-table), presents 'Random'/'Realistic' with evaluation dashboard	Healthcare research sharing, finance, internal/external sharing & test data	Synapp SaaS connects to DBs, generates privacy-preserving copies; guidance on dataset selection	Requires original datasets; produces synthetic copies for collaboration/sharing	Founded 2020; accelerator-backed; lean team	Positions synthesis for compliant sharing; evaluation metrics; no formal certs stated	Employees: 3 (PitchBook); turnover: not disclosed	Not publicly disclosed; public test datasets available	B (using real-world data); B1/B2 with B3
Syntheticus	Zurich, CH	Generative AI across tabular/text/image/multimodal; Differential Privacy & Confidential	Finance, healthcare, public sector & enterprise multi-	Syntheticus Suite (Core + Modules: Profile/Generate/Protect/Validate);	Learns from organizational data; outputs synthetic	Founded 2021; in Google & Microsoft for Startups	PET stack (DP + TEEs), SSO; aligns to EU compliance	Employees: 1–10 (Crunchbase); turnover: not	Not publicly disclosed	B (using real-world data);

STUDY ON THE NEXT DATA FRONTIER:
GENERATIVE AI, REGULATORY COMPLIANCE AND INTERNATIONAL DIMENSIONS (NO. 2024-020)

		Computing options	type data	deployable on-prem/private/public cloud/edge	datasets for analysis/sharing	programmes	expectations	disclosed		B1/B2 and B3
YData	Lisbon, PT	Generative models + anonymisation & PII detection; rebalancing/augmentation in Fabric	Finance (credit/fraud), telecom, healthcare, retail/utilities	YData Fabric: data catalog/profiling, automated PII, synthetic generation & governance	Uses customer datasets to generate/augment synthetic versions	Founded 2019; seed funding 2021; expanding to NAV EU	Privacy-by-design workflows; automated PII detection & masking; GDPR/CCPA-aligned handling	Employees: ~34 (PitchBook); turnover: not disclosed	Not publicly disclosed	B (using real-world data): B1 (patient) with B2/B3 as applicable
AETION	Barcelona, ES (EU tech hub) / US HQ	De-identification + synthetic data generation within Aetion Generate; AI-powered privacy risk assessment	Healthcare & life sciences (RWE from RWD; RCT augmentation)	Software + expert services: risk modeling, de-identification, synthetic data services	Requires real-world clinical/claims/EHR data; produces privacy-preserved synthetic & de-identified datasets	EU hub growth 2024; Generate 3.6 launched 12/2024; established RWE vendor	HIPAA/GDPR-aligned; expert determination; governance/audit logging	Employees/turnover: not publicly disclosed (site)	Not publicly disclosed; Top-20 biopharma quote on synthetic trial data	B (using real-world data): B1/B2 with B3
AINDO	Trieste, IT	Generative AI for fully synthetic tabular enterprise data; bias removal/compatibilization/completion	Healthcare, finance, public administration/R&D	Synthetic Data Platform with enterprise features; Europrivacy/ISO 9001; NIST ranking claim	Synthetic Data Platform with enterprise features; Europrivacy/ISO 9001; NIST ranking claim	Founded 2018; Europrivacy certification (22 Jul 2024)	Europrivacy (GDPR Art. 42 scheme); privacy-by-design local deployment	Employees/turnover: not publicly disclosed	Not publicly disclosed; healthcare deployments implied	B (using real-world data): B1/B2 (anonymized outputs; limited B3)
TRUATA	Dublin, IRE	Anonymisation & privacy-enhanced analytics (clear rooms); not primarily generative SD vendor.	Retail/consumer analytics; Mastercard bank analytics	TAS (Anonymization Service), PEAP (bank analytics), Data Clean Room (Databricks/Delta Sharing)	Operates on enterprise data; returns anonymized outputs/aggregates/model code inside controlled envs	Established 2018; ISO 27701 referenced; strong Mastercard linkage	GDPR-aligned anonymization; controller-to-controller trust structure; risk assessment & governance	Employees/turnover: not publicly disclosed; Dublin office published	Not publicly disclosed; PEAP targets Mastercard qualifying banks	B (using real-world data): B1/B2 (anonymized outputs; limited)

STUDY ON THE NEXT DATA FRONTIER:
GENERATIVE AI, REGULATORY COMPLIANCE AND INTERNATIONAL DIMENSIONS (NO. 2024-020)

B3)

OCTOPIZE	Nantes, FR	Avatar method: anonymous synthetic 'avatars' aligned to EDPB anonymisation criteria; peer-reviewed,	Healthcare, social sciences./public statistics; expanding to telco/insurance/mobility/banking	Avatar software & services for anonymization/synthetic datasets	Uses real datasets to generate anonymous avatars intended as non-personal data	Founded 2018; CNIL evaluation (2020); strong French/EU health-data profile	Claims outputs are non-personal for secondary use (aligned to EDPB criteria); CNIL evaluation noted	Employees/turnover: not publicly disclosed	Not publicly disclosed; sector case mentions/public partnerships	B (using real-world data): B1 (health) and B2; supports B3
Brighter AI	Berlin, DE	Deep Natural Anonymisation (synthetic overlays) + Precision Blur for faces/plates; full-body blur option	Automotive/ADAS, smart cities, video analytics/surveillance	Software (cloud/private/edge) for image/video anonymization	Operates on real visual streams; creates synthetic overlays; not scene-from-scratch	Founded 2017; acquired by Milestone Systems (Apr 2025)	Supports GDPR/CCPA/API/PIPL; non-reversible overlays to mitigate re-identification	Employees: ~24 pre-acquisition (media); turnover: not publicly disclosed	Not publicly disclosed; 'trusted by leading enterprises'	B (using real-world data): B1 (personal images); augmentation
Sherpa.ai	Erandio, ES	Federated Learning with Secure Aggregation and Differential Privacy options; does not create datasets absent real inputs.	Healthcare, finance/insurance, public sector/industry	SaaS FL platform + open-source framework; governance/auditability	Uses distributed data at source; exchanges model parameters only	Operating since 2012 (company lineage); high-profile advisors; multi-sector deployments	Data stays local; DP & Secure Aggregation; EDPS/AEPD TechDispatch commentary cited by vendor	Employees/turnover: not publicly disclosed	Not publicly disclosed	B (using real-world data): training on B1/B2 data in situ; no SD product
Anyverse	Madrid, ES	Fully simulated, physics-accurate synthetic imagery (RGB/LiDAR/Radar) with pixel-level ground truth; no real personal data.	ADAS/AD, in-cabin monitoring, defence/robotics CV	Synthetic data platform & apps (ADAS & AD; In-Cabin); scenario design UI	Generates from scratch (simulation); no existing datasets required	Founded ~2018; OEM partnerships; expo presence; Madrid HQ	Effectively outside GDPR scope for personal data (no personal data processed)	Employees/turnover: not publicly disclosed	Not publicly disclosed; deployments via partners	A (without real-world data); simulation-based
Oblivious	Dublin, IRE	Privacy infrastructure (Differential Privacy analytics + Confidential	Cross-border data sharing/compliance; secure	PET middleware & enclave deployment; auditing/attestation	Operates on existing datasets; enables privacy-	Founded 2020; ISO 27001 and SOC 2 posture	DP-safe query middleware + TEEs (Intel SGX/Nitro/Azure	Employees/turnover: not publicly	Not publicly disclosed	B (using real-world data):

Computing);	analytics for finance/health/pu blic sector	n tooling	first collaboration; no generation without inputs	highlighted	CC); zero-trust design & auditability	disclosed	PETs to analyze B1/B2 in situ; no SD product

Annex 3 Task 2: Legal analysis

A3.1 Agriculture

<u>EU Legislation</u>	<u>Specific article</u>	<u>Reporting and compliance obligations</u>	<u>Type of reporting activities</u>
Deforestation Regulation (Regulation (EU) 2023/1115)	<i>Article 9</i>	Maintain records of information on supply chain	Upon request of the competent authorities, provide documentation on the type, quantity, origin and non-infringement of products.
	<i>Article 10</i>	Maintain records of risk assessment	Upon request of the competent authorities, provide documentation showing an annual risk assessment has been carried out.
	<i>Article 11</i>	Maintain records of risk mitigation measures	Upon request of the competent authorities, provide documentation showing an annual review of the risk mitigation measures and procedures has been done.
	<i>Article 12</i>	Annual public reporting on due diligence system	Publicly publish report yearly as widely as possible, including online, on the due diligence system and steps taken to meet information, risk assessment and risk mitigation goals. Upon request of the competent authorities, provide documentation related to due diligence, including for the past 5 years. The due diligence statement must contain at least the information in Annex II, namely: 1. Operator's name, address and, in the event of relevant commodities and relevant products entering or leaving the market, the Economic Operators Registration and Identification (EORI) number in accordance with Article 9 of Regulation (EU) No 952/2013. 2. Harmonised System code, free-text description, including the trade name as well as, where applicable, the full scientific name, and quantity of the relevant product that the operator intends to place on the market or export. For relevant products entering or leaving the market, the quantity is to be expressed in kilograms of net mass and, where applicable, in the supplementary unit set out in Annex I to Regulation (EEC) No 2658/87 against the indicated Harmonised System code or, in all other cases, expressed in net mass specifying a percentage estimate or deviation or, where applicable, volume or number of items. A supplementary unit is applicable where it is defined consistently for all possible subheadings under the Harmonised System code referred to in the due diligence statement. 3. Country of production and the geolocation of all plots of land where the relevant commodities were produced. For relevant products that contain or have been made using cattle, and for such relevant products that have been fed with relevant products, the geolocation shall refer to all the establishments where the cattle were kept. Where the relevant product contains or has been made using commodities produced in different plots of land, the geolocation of all plots of land shall be included in accordance with Article 9(1), point (d). 4. For operators referring to an existing due diligence statement pursuant to Article 4(8) and (9), the reference number of such due

			diligence statement. 5. The text: 'By submitting this due diligence statement the operator confirms that due diligence in accordance with Regulation (EU) 2023/1115 was carried out and that no or only a negligible risk was found that the relevant products do not comply with Article 3, point (a) or (b), of that Regulation.' 6. Signature
Regulation (EU) No 2021/2115	<i>Article 12 and 14</i>	Conditionality requirement for receiving aid	Farmers and other beneficiaries of CAP must meet requirements related to climate and environment, public and plant health, animal welfare, and working conditions. If these requirements are not met, farmers are subject to administrative penalties. Accordingly, under national schemes, farmers will have to report. For aid, farmers must respect Statutory Management Rules (SMR), as well as Good Agricultural and Environmental Conditions (GAEC). These can be found in Annex III to Regulation 2021/2115. The table of SMR and GAEC obligations can be found subsequent to this table.
	<i>Article 29-30</i>	Complimentary income support for sustainability and young farmers	Farmers and other beneficiaries of CAP may receive complimentary income support if they are a young farmer or for sustainability. Accordingly, under national schemes, they will have to report age and sustainability.
	<i>Article 70</i>	Payments for environmental, climate-related and other management commitments	Farmers and other beneficiaries can receive payments for voluntary management commitments beneficial to the environment. Accordingly, they will have to report these commitments under the national schemes.
Regulation (EU) No 2021/2116	<i>Article 69</i>	Geo-spatial and animal-based applications for interventions	Farmers and other beneficiaries must make an application by means of the geo-spatial application form, or the animal-based application where animals are concerned, provided by the competent authority.
Commission Implementing Regulation (EU) 2022/1173	<i>Article 6</i>	Content of aid applications	All CAP aid applications must include: (a) the identity of the beneficiary; (b) details of the intervention(s) applied for; (c) where appropriate, any supporting documents needed to establish the eligibility conditions and other relevant requirements for the intervention concerned; (d) information relevant to conditionality.
	<i>Article 8</i>	Content of the geo-spatial applications	Geo-spatial aid applications must include: (a) unambiguous identification of agricultural parcels and units of land containing non-agricultural areas considered eligible by the Member State of the holding; (b) clear delineation of the area declared for aid under each intervention on the agricultural parcels and units of land containing non-agricultural areas considered eligible by the Member State, in particular if the area claimed is smaller than the total area of the agricultural parcel; (c) the type, location and, where relevant, size of landscape features relevant for conditionality or interventions; L 183/30 EN Official Journal of the European Union 8.7.2022

			(d) crop on agricultural parcels, where relevant; (e) where relevant, whether the parcel is subject to organic farming, and in particular for the conversion or maintenance of organic farming practices and methods as laid down in Regulation (EU) 2018/848 of the European Parliament and of the Council (8), relevant for support granted for interventions referred to in Articles 31 and 70 of Regulation (EU) 2021/2115 or conditionality; (f) where relevant, information on the use of plant protection products for parcels under interventions for sustainable and reduced use of pesticides under Articles 31 and 70 of Regulation (EU) 2021/2115. Member States may decide to use this information in respect to the requirement of record-keeping of plant protection products laid down in Article 67(1) of Regulation (EC) No 1107/2009; (g) the identification of the payment entitlements in accordance with the identification and registration system provided for in Article 73 of Regulation (EU) 2021/2116 for the purpose of the basic income support for sustainability; (h) for areas claimed for the crop-specific payment for cotton, the variety of cotton seed used and, where applicable, the identification of the approved interbranch organisation of which the beneficiary is a member; (i) for areas used for the production of hemp, the variety of seed used, an indication of the quantities of the seeds used, expressed in kilograms per hectare, and the official labels used on the packaging of the seeds in accordance with Council Directive 2002/57/EC (9), and in particular Article 12 thereof, or any other document recognised as equivalent by the Member State. Where the labels also have to be submitted to other national authorities, Member States may provide for those labels to be returned to the beneficiaries. The labels returned shall be marked as used for an application.
	<i>Article 9</i>	Animal-based interventions application content requirements	Applications for animal-based interventions shall contain at least the following: (a) the number of animals or, as relevant, the number of livestock units, of each type, in respect of which an animal-based intervention is applied for; (b) where relevant, information on the location where the animals will be held in the calendar year covered by the aid application; (c) where the support concerns the bovine animals or sheep and goats, updated information relevant for the intervention on the animals in relation to the system for the identification and registration of animals, in accordance with Article 34(2) of Regulation (EU) 2021/2115.
Plant Protection Product Regulation (Regulation (EC) 1107/2009)	<i>Article 29</i>	Demonstrate compliance with plant protection product norms	When making an application for authorisation for placing a plant protection product on the market, the applicant has to demonstrate in their application that the following requirements in article 29.1 of Regulation 1107/2009 are met: - Active substances, safeners and synergists have been approved; - the specification, pursuant to Article 38, does not deviate significantly from the specification included in the Regulation approving that substance, safener or synergist; and the active substance, safener or synergist has no more harmful effects within the meaning of Article 4(2) and (3) due to its impurities than if it had been produced in accordance with the manufacturing process specified in the dossier that supported the

		approval if the active substance, safener or synergist is produced by a different source, or by the same source with a change in the manufacturing process and/or manufacturing location: - its co-formulants are not included in Annex III of the Regulation; - its technical formulation is such that user exposure or other risks are limited as much as possible without compromising the functioning of the product; - in the light of current scientific and technical knowledge, it complies with the requirements provided for in Article 4(3); - the nature and quantity of its active substances, safeners and synergists and, where appropriate, any toxicologically, ecotoxicologically or environmentally relevant impurities and co-formulants can be determined by appropriate methods; - its residues, resulting from authorised uses, and which are of toxicological, ecotoxicological or environmental relevance, can be determined by appropriate methods in general use in all Member States, with appropriate limits of determination on relevant samples - its physical and chemical properties have been determined and deemed acceptable for the purposes of the appropriate use and storage of the product.
Article 33	Provide information during application for authorization of the product	When making an application for authorisation for placing a plant protection product on the market, the applicant has to include the elements listed in article 33.2 and 33.3 of Regulation 1107/2009, namely: - a list of intended uses in each zone as indicated in Annex I and the Member States where the applicant has made or intends to make an application; - a proposal as to which Member State the applicant expects to evaluate the application in the zone concerned. In the case of an application for use in greenhouses, as post-harvest treatment, for treatment of empty storage rooms and for seed treatment, only one Member State shall be proposed, which evaluates the application taking account of all zones. In this case the applicant shall send the summary or complete dossier as referred to in Article 8 to other Member States on request; - where relevant, a copy of any authorisations already granted for that plant protection product in a Member State; - where relevant, a copy of any conclusion of the Member State assessing equivalence as referred to in Article 38(2); - for the plant protection product concerned, a complete and a summary dossier for each point of the data requirements of the plant protection product; - for each active substance, safener and synergist contained in the plant protection product, a complete and a summary dossier for each point of the data requirements of the active substance, safener and synergist; - for each test or study involving vertebrate animals, a justification of the steps taken to avoid animal testing and duplication of tests and studies on vertebrate

			animals; - the reasons why the test and study reports submitted are necessary for first authorisation or for amendments to the conditions of the authorisation; - where relevant a copy of the application for a maximum residue level as referred to in Article 7 of Regulation (EC) No 396/2005 or a justification for not supplying such information; - where relevant for an amendment of an authorisation an assessment of all information submitted in accordance with point (h) of Article 8(1); - a draft label.
	<i>Article 43</i>	Provide information during application for renewal of authorization of the product	When making an application for renewal of authorisation for placing a plant protection product on the market, the applicant has to include the elements listed in article 43.2 of Regulation 1107/2009 within 3 months of the renewal, namely: - a copy of the authorisation of the plant protection product; - any new information required as a result of amendments in data requirements or criteria; - evidence that the new data submitted are the result of data requirements or criteria which were not in force when the authorisation of the plant protection product was granted or necessary to amend the conditions of approval; - any information required to demonstrate that the plant protection product meets the requirements set out in the Regulation on the renewal of the approval of the active substance, safener or synergist contained therein; - a report on the monitoring information, where the authorisation was subject to monitoring.
	<i>Article 56</i>	Provide information when a product no longer meets the plant protection product norms	The holder of an authorization shall notify the granting Member State if there is any new information suggesting the plant protection product would no longer meet the criteria in this Regulation.
	<i>Article 67</i>	Record keeping	Records of the plant protection products produced, imported, exported, sorted or placed on the market must be kept for 5 years. Professional users must keep records for 3 years of products used. On request by the competent authorities, this information must be made available.
Commission Implementing Regulation (EU) 2023/564 as regards the content and format of the records of plant protection products kept by professional	<i>Article 1</i>	Content	The professional users of plant protection products must provide the information listed in Annex I of this Implementing Regulation. This Annex is added below.
	<i>Article 2</i>	Format	The format must be machine-readable format, meaning it is a file format structured so that software applications can easily identify, recognise and extract specific data, including individual statements of fact, and their internal structure
	<i>Article 5</i>	Entry into force	This Implementing Regulation on the content and format of the record-keeping shall enter into force on 1 January 2026.

users pursuant to Regulation (EC) No 1107/2009			
Corporate Sustainability Reporting Directive (Directive (EU) 2022/2464)	<i>Article 19a</i>	Sustainability report	Companies in scope of the CSRD must include information related to the sustainability aspects of their business and business model in a dedicated section of their annual management report. SMEs may utilize a more limited form of sustainability reporting.
	<i>Article 29a</i>	Sustainability reporting standards	Companies in scope of the CSRD must report using Commission-made standards on the environment, including water use, pollution and ecosystems, as well as social, human rights and governance factors.

A3.2 Automotive

<u>EU Legislation</u>	<u>Type</u>	<u>Specific article</u>	<u>Reporting and compliance obligations</u>	<u>Type of reporting activities</u>
Regulation (EU) 2018/858 on the regulation on the approval and market surveillance of vehicles, systems and components (M, N, O categories)	Mixed	Art. 9(5)	Share records with competent authorities	Manufacturers are required to give the Commission (free of charge and without undue delay) any data that is necessary for the purpose of compliance verification and which is not already available in the EU type-approval certificate and its attachments.
		Art. 13(7)	Maintain records	Manufacturers are required to examine the complaints they receive and keep records of them.
		Art. 13(10)	Sharing data	Manufacturers are required to make available data which is needed for testing by third parties for possible non-compliance.
		Art. 14(1)	Inform relevant authorities	Inform the relevant approval authority whether a vehicle, system, component, separate technical unit, part or equipment placed on the market or entered into service is not in conformity and the data on which the approval was granted is incorrect. The manufacturer will also indicate if any measures are taken.
		Art. 14(3)	Maintain records.	Manufacturers are required to keep the EU type-approval certificates and their attachments for a period of 10 years following their end of validity (5 years for systems, components, or separate technical units).
		Art. 14(4)	Share records with competent authorities	Manufacturers are required to provide national authorities or the Commission with a copy of the EU type-approval certificate or authorisation proving conformity in the language of the authority.
		Art. 23(1)	Submit application for type-approval	Manufacturers are required to submit an application for EU type-approval that includes an information folder (Art. 24) and the additional information mentioned in Art. 25.
		Art. 36	Issue a certificate of conformity	Manufacturers are required to provide a paper-based certificate of conformity for each vehicle sold, following a standardized template. Starting in 2026, they will be permitted to issue the certificate in electronic format as structured data instead of the paper version.
Regulation (EU) 2019/631 setting CO ₂ emission performance standards for new passenger cars and for new light commercial vehicles (M1, N1 categories)	Environment	Art. 4	Sets specific CO ₂ emissions targets	Establishes CO ₂ emission performance standards for new passenger cars and new light commercial vehicles.
		Art. 12(2)	Make available data on real-world CO ₂ emissions and fuel or energy consumption	The Commission is required to ensure that data on real-world CO ₂ emissions and fuel or energy consumption of passenger cars and light commercial vehicles are made available by manufacturers, national authorities or through direct data transfer from vehicles. Reporting obligations are established in the implementing regulation below.

Commission Implementing Regulation (EU) 2021/392 on the monitoring and reporting of data relating to CO ₂ emissions from passenger cars and light commercial vehicles (M1, N1)	Environment	Art. 6	Notification of errors	Manufacturers are required to notify the Commission of any errors found in the submitted data. To this purpose, they are required to upload the complete corrected dataset to the BDR and notify the Commission by e-mail.
		Art. 7	Transmit data	Manufacturers of a base vehicle as referred to in point 1.2.2. of Annex III to Regulation (EU) 2019/631 will need to transmit the data in that point via electronic data transfer to the BDR.
		Art. 8	Transmit data	Manufacturers of new passenger cars (registered in 2021 and 2022) with measured NED CO ₂ < 50 g CO ₂ /km shall report those values to the Commission.
		Art. 9(1)	Real world data collection	Car manufacturers are required to collect real world data (fuel and energy consumption) together with the vehicle's identification number (VIN) of new passenger cars and new light commercial vehicles.
		Art. 9(3)	Submit annual reports	Car manufacturers are required to report to the Commission the data collected during the previous year on 1 April of each year
Regulation (EU) 2017/2400 on the determination of CO ₂ emissions and fuel consumption of heavy-duty vehicles (M3, N2)	Environment	Art. 9(2)	Record test results	Manufacturers are required to record the results of the simulations conducted in accordance with the procedure mentioned in Art. 9(1). The results will be presented using the model provided in Part I of Annex IV.
		Art. 11(1)	Maintain records Share records with competent authorities	Manufacturers will store their records file, the vehicle information file (VIF) and certificates on CO ₂ emissions and fuel consumption for 20 years after the production of the vehicle. They are required to provide approval authorities and the Commission <with the records when requested to do so.
		Art. 16	Submit application for type-approval	Manufacturers are required to apply for certification of CO ₂ emissions and fuel consumption by filling in an information document (the models are provided in the Annexes). The application needs to be accompanied by explanations of the elements of design and relevant test reports and by a statement of compliance issues by an approval authority based on point 2 of Annex IV to Regulation (EU) 2018/858.
		Art. 19(1)	Inform relevant authorities	Manufacturers are required to inform the approval authority of any change to the design or the manufacturing process occurring after certification and potentially affecting CO ₂ emissions and fuel consumption.
		Art. 20(1)	Submit annual reports Maintain records Share records with competent authorities	Manufacturer are required to provide by the 31 December of each year a test report for each medium and heavy lorry. The test reports will be stored for at least 10 years. The manufacturer will make them available to the Commission and the approval authorities upon requests.

		Art. 22(2)	Submit annual reports	Manufacturers are required to provide annual testing reports containing the results from the procedures mentioned in Art. 22(1) to the approval authority who certified the CO ₂ emissions and fuel consumption.
		Art. 22(4)	Share records with competent authorities	Manufacturers are required to provide the approval authority with all the relevant documents, samples and other materials needed to perform verifications.
Regulation (EU) 2019/1242 setting CO ₂ emission performance standards for new heavy-duty vehicles (M2, N2; N1, N2, N3 if out of scope of Regulation (EU) 2019/631; O3, 04), as amended by Regulation (EU) 2024/1610	Environment	Art. 13b	Report data	Manufacturers of heavy-duty vehicles are required to report data on required for the determination of the CO ₂ emissions for each new a heavy-duty vehicle by 30 September of each year.
Regulation (EC) No 1907/2006 – The Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) Regulation	Safety	Article 6 and 7 (registration) Article 41 (evaluation of compliance of by ECHA)	Registration of substances (and compliance check by ECHA)	Companies are required to collect information on the properties and uses of substances they manufacture in amounts that exceed one tonne per year. They also need to assess the hazards and potential risks. This information is communicated to the ECHA via a registration dossier. The ECHA has the responsibility of checking whether the registrations are compliant with the REACH regulation. The evaluation concerns: the testing proposals submitted by registrants; compliance check of the dossier; substance evaluation.
		Art. 7.2	Notify ECHA	Producers are required to notify the ECHA if a substance meets the criteria in Art. 57 and is identified in accordance with Art. 59(1) and is present in article above off a concentration of 0.1% w/w and is present in those articles here the concentration exceeds 0.1% w/w in quantities totalling over 1 tpa.
		Article 14	Prepare Chemical Safety Reports (CSR)	A CSR is required to be completed for all substances subject to registration and produced in quantities of 10 tonnes or more per year per registrant. It documents the chemical safety assessment conducted in accordance with paragraphs 2 to 7 and with Annex I. Registrants are required to keep the CSR available and up to date.

		Article 31 (requirements detailed in Annex II)	Prepare Safety Data Sheets (SDS) (and more generally communicate information along the supply chain, as detailed in the next articles)	Suppliers of a substance or a mixture shall provide the recipient of the substance or mixture with an SDS if the substance meets the criteria for classification as hazardous in accordance with Regulation (EC) No 1272/2008, or is persistent, bioaccumulative and toxic or very persistent and very bioaccumulative (Annex XIII), or included in the list established in accordance with Article 59(1) for other reasons than those mentioned above.
		Article 32	Communicate information	Suppliers of substances are required to communicate specific information down the supply chain for substances that do not require a safety data sheet. This information may be provided either in paper or electronic format and must include the elements set out in Article 32(1).
		Article 33	Communicate information	Suppliers of articles containing a substance meeting the criteria in Art. 57 and identified in accordance with Art. 59(1) in a concentration above 0.1% w/w needs to provide the recipient with sufficient information to allow safe use (as a minimum, the name of the substance)
		Article 34	Communicate information	Actors along the supply chain are required to pass on specific information to the next actor or distributor. This includes any new data regarding hazardous properties of substances, as well as any other information that may raise concerns about the adequacy of the risk management measures outlined in the safety data sheet (SDS).
		Article 62	Submit authorization applications	Manufacturers are required to submit an authorisation request for substances of very high concern included in Annex XIV.
Regulation (EU) 2024/1257 on type-approval of motor vehicles and engines and of systems, components and separate technical units intended for such vehicles, with respect to their emissions and battery durability (Euro 7)	Mixed	Art. 6(8)	Inform authorities in case of non-compliance	Manufacturers are required to inform the approval authority in case of a vehicle / system / component / separate technical unit presents a serious risk or non-compliance with the requirements of the Regulation and provide details.
		Art. 13(2)	Making data available	Manufacturers are required to provide data when requested by the Commission for verifying compliance of vehicles, components and technical units with the Annexes of this Regulation.
		Art. 14	Provide declaration of compliance	Manufacturers are required to provide a declaration of compliance stating that a specific type or group of vehicles, components or technical units are in compliance with the Regulation requirements to the approval authority.
Directive (EU) 2022/2464 on corporate sustainability reporting	Sustainability	Art. 19a	Sustainability report	Companies in the scope of the CSRD must include information related to the sustainability aspects of their business and business model in a dedicated section of their annual management report. SMEs may utilize a more limited form of sustainability reporting.

		Art. 29a	Sustainability reporting standards	Companies in scope of the CSRD must report using Commission-made standards on the environment, including water use, pollution and ecosystems, as well as social, human rights and governance factors.
Directive 2006/40/EC relating to emissions from air conditioning systems in motor vehicles (M1 and N1 categories)	Environment	Art. 4(2)	Share records with competent authorities	Manufacturers must provide approval authorities with the technical documentation related to the installed air-conditioning systems and the gases.
Directive 2005/64/EC on the type-approval of motor vehicles with regard to their reusability, recyclability and recoverability (M1 and N1 categories)	Environment	Art. 5(2)	Communicate technical data	Manufacturers must provide the approval authority with the detailed technical information necessary to perform the calculations and checks described in Annex I related to the nature of the materials used to construct the vehicle and its components.
		Art. 6(7)	Inform relevant authorities	The manufacturer will communicate to the competent if there is any significant change that could impact the relevance of the certificate of compliance.

A3.3 Healthcare

<u>EU Legislation</u>	<u>Specific article</u>	<u>Reporting and compliance obligations</u>	<u>Type of reporting activities mandated by the Regulation</u>
Medical Devices Regulation (MDR, Regulation (EU) 2017/745)	<i>Article 10</i>	Manufacturers must keep the technical documentation, the EU declaration of conformity, and a copy of any relevant certificates during 10 years.	Create and maintain for 10 years and show to the authorities if requested: - Technical documentation on the medical device - EU Declaration of conformity - Copies of any relevant certificates - System of risk management
	<i>Article 31</i>	Manufacturers of medical devices must register themselves in the electronic system for medical device manufacturers	Submit the following information relating to the economic operator and ensure the information entered in the registration system is up to date. 1. type of economic operator (manufacturer, authorised representative, or importer), 2. name, address and contact details of the economic operator, 3. where submission of information is carried out by another person on behalf of any of the economic operators mentioned under Section 1.1, the name, address and contact details of that person, 4. Name, address and contact details of the person or persons responsible for regulatory compliance.
	<i>Article 83</i>	Manufacturers must establish, implement, and maintain a post-market surveillance (PMS) system proportionate to the risk class of the device.	Inform the competent authorities where there is a need for preventive or corrective action is identified. Report serious incidents and field safety corrective actions in accordance with article 87.
	<i>Article 85</i>	Manufacturers of Class I devices must prepare a post-market surveillance report (PMSR).	Manufacturers of class I devices must provide a report to the authorities upon request. This report covers the results and conclusions of analyses of the post-market surveillance data during the post-market surveillance plan.
	<i>Article 86</i>	Manufacturers of Class IIa, IIb, and III devices must prepare a Periodic Safety Update Report (PSUR).	Manufacturers of class IIa, class IIb and class III devices shall maintain a periodic safety update report, which they must make available to the notified body involved in the conformity assessment and upon request to the competent authority. The report must be updated annually for class IIb and class III products, and every 2 years for class IIa devices.

	<i>Article 87</i>	Mandatory reporting of serious incidents and field safety corrective actions (FSCA) to competent authorities.	Manufacturers shall report serious incidents and field safety corrective actions to the competent authorities. Serious public health threats must be reported within 2 days of the manufacturer becoming aware of the threat. In case of death or unanticipated serious deterioration in a person's health, the report shall be provided within 10 days of the manufacturer becoming aware of the serious incident. The report on serious incidents may consist of an incomplete initial report that is followed up by a complete report. Field safety corrective actions must be reported before carrying them out, unless they are urgent.
	<i>Article 88</i>	Requires trend reporting when a statistically significant increase in incident frequency/severity is identified.	Manufacturers must report statistically significant incidents that are not serious incidents that could have a significant impact on the health or safety of patients, users or other persons that are unacceptable when weighed against the intended benefits. Manufacturers must also specify how they are managing these incidents in the post-surveillance plan.
	<i>Article 89</i>	Obligates submission of field safety notices and coordination with competent authorities during vigilance reporting.	Manufacturers must provide upon request all necessary documents for the risk assessment carried out by the authorities.
	<i>Article 93</i>	Competent authority may ask for documentation to carry out their activities.	Manufacturers must provide documentation and information necessary for the competent authorities upon request, including samples of devices or access to devices free of charge.
In Vitro Diagnostic Medical Devices Regulation (IVDR, Regulation (EU) 2017/746)	<i>Article 10</i>	Manufacturers must keep the technical documentation, the EU declaration of conformity, and a copy of any relevant certificates during 10 years.	Create and maintain for 10 years and show to the authorities if requested: - Technical documentation on the medical device - EU Declaration of conformity - Copies of any relevant certificates - System of risk management
	<i>Article 26</i>	Devices must be registered before being placed on the market	Manufacturers must assign a Basic UDI-DI to the device and provide it to the UDI database alongside other core data elements, namely: 1. Basic UDI-DI, 2. type, number and expiry date of the certificate issued by the notified body and the name or identification number of that notified body and the link to the information that appears on the certificate and was entered by the notified body in the electronic system on notified bodies and certificates, 3. Member State in which the device shall or has been placed on the market in the Union, 4. in the case of class B, class C or class D devices: Member States where the device is or is to be made available, 5. presence of tissues, cells, or, their derivatives,

			of human origin (y/n), 6. presence of tissues, cells or their derivatives of animal origin as referred to in Regulation (EU) No 722/2012(y/n), 7. presence of cells or substances of microbial origin (y/n), 8. risk class of the device, 9. where applicable, the single identification number of the performance study, 10. in the case of devices designed and manufactured by another legal or natural person as referred in Article 10(14), the name, address and contact details of that legal or natural person, 11. in the case of class C or D devices, the summary of safety and performance, 12. status of the device (on the market, no longer placed on the market, recalled, field safety corrective Action initiated), 13. indication as to whether the device is a 'new' device. A device shall be considered to be 'new' if: (a) there has been no such device continuously available on the Union market during the previous three years for the relevant analyte or other parameter; (b) the procedure involves analytical technology not continuously used in connection with a given analyte or other parameter on the Union market during the previous three years. 14. indication as to whether the device is intended for self-testing or near-patient testing.
	Article 28	Manufacturers of IVF medical devices must register themselves in the electronic system for medical device manufacturers	Submit the following information relating to the economic operator and ensure the information entered in the registration system is up to date. 1. type of economic operator (manufacturer, authorised representative, or importer), 2. name, address and contact details of the economic operator, 3. where submission of information is carried out by another person on behalf of any of the economic operators mentioned under Section 1.1, the name, address and contact details of that person, 4. Name, address and contact details of the person or persons responsible for regulatory compliance.

	<i>Article 78</i>	Establish and maintain a post-market surveillance system for IVD devices.	Inform the competent authorities where there is a need for preventive or corrective action is identified. Report serious incidents and field safety corrective actions in accordance with article 82.
	<i>Article 80</i>	For Class A and B devices, prepare and update a Post-Market Surveillance Report (PMSR) summarising key findings of the PMS system.	Manufacturers of class A and B devices must provide a report to the authorities upon request. This report covers the results and conclusions of analyses of the post-market surveillance data during the post-market surveillance plan.
	<i>Article 81</i>	For Class C and D devices, compile and submit a Periodic Safety Update Report (PSUR) detailing benefit-risk evaluations and conclusions on device performance and safety	Manufacturers of class C and D devices shall maintain a periodic safety update report, which they must make available to the notified body involved in the conformity assessment and upon request to the competent authority. The report must be updated annually.
	<i>Article 82</i>	Obligates manufacturers to report serious incidents and field safety corrective actions to competent authorities.	Manufacturers shall report serious incidents and field safety corrective actions to the competent authorities. Serious incidents must be reported within 15 days of the manufacturer becoming aware of the incident. Serious public health threats must be reported within 2 days of the manufacturer becoming aware of the threat. In case of death or unanticipated serious deterioration in a person's health, the report shall be provided within 10 days of the manufacturer becoming aware of the serious incident. The report on serious incidents may consist of an incomplete initial report that is followed up by a complete report. Field safety corrective actions must be reported before carrying them out, unless they are urgent.
	<i>Article 83</i>	Submit trend reports when statistically significant increases in the frequency or severity of incidents are identified	Manufacturers must report statistically significant increases in the frequency or severity of non-serious incidents which could have a significant impact on the health or safety of patients, users or other persons that are unacceptable when weighed against the intended benefits. Manufacturers must also specify how they are managing these incidents in the post-surveillance plan.
	<i>Article 84</i>	Perform risk-based analysis of serious incidents, as well as Incident investigation, risk evaluation, and documentation in response to field safety events.	Manufacturers must provide upon request all necessary documents for the risk assessment carried out by the authorities.

	<i>Article 88</i>	Competent authority may ask for documentation to carry out their activities.	Manufacturers must provide documentation and information necessary for the competent authorities upon request, including samples of devices or access to devices free of charge.
Clinical Trials Regulation (CTR, Regulation (EU) No 536/2014)	<i>Article 42</i>	Requires sponsors to report on serious adverse reactions.	Sponsors must report electronically all relevant information on suspected unexpected serious adverse reactions to investigational medicinal products in or following a clinical trial, or serious adverse reactions to the same active substance. In case of fatal or lifethreatening suspected unexpected serious adverse reactions, this report must be done within 7 days of becoming aware of the reaction. In case of non-fatal or non-life-threatening suspected unexpected serious adverse reactions, the report must be done within 15 days of becoming aware of the reaction. The report must be submitted electronically to the database for safety reporting set up by the European Medicines Agency, namely the EudraVigilance database. The report may be initially incomplete, followed up by a complete report, to ensure timely reporting.
	<i>Article 43</i>	Requires annual reporting by the sponsor to the Agency	When testing investigational medicinal products other than placebo, the sponsor shall submit through the EudraVigilance database an annual report on the safety of each investigational medicinal product in a clinical trial for which it is the sponsor. This report shall contain only aggregate and anonymised data. These annual reports must continue until the last clinical trial has been conducted by the sponsor on the investigational medicinal product. In cases where there are multiple investigational medicinal products, the sponsor may submit a single safety report on all investigational medicinal products used in a particular clinical trial.
	<i>Article 52</i>	Requires sponsors to report serious breaches of the Clinical Trial Regulation	Sponsors must notify the relevant Member States about a serious breach of the Clinical Trials Regulation within seven days of becoming aware of the breach. A significant breach is a breach likely to significantly affect the safety and rights of a subject or the reliability and robustness of the data generated in the clinical trial. The report must be done electronically through the Clinical Trials Information System.
	<i>Article 53</i>	Requires sponsors to report unexpected events and third country inspection reports	Sponsors must notify the relevant Member States of all unexpected events which affect the benefit-risk balance of the clinical trial, which are not suspected unexpected serious adverse reactions. Sponsors must also submit all inspections reports of third country authorities concerning the clinical trial. Upon request, the sponsor must submit a translation into an official EU language. These reports must be done electronically through the Clinical Trials Information System.

	<i>Article 54</i>	Requires sponsors to report unexpected events and urgent safety measures	Sponsors must notify the relevant Member States electronically through the Clinical Trials Information System of any unexpected event likely to seriously affect the benefit-risk balance and the urgent safety measures taken
Community Code relating to Medicinal Products for human use (Consolidated Directive 2001/83/EC)	<i>Article 107b (Directive 2001/83/EC)</i>	Mandates marketing authorization holders to record and report all suspected adverse reactions occurring within the EU.	Marketing authorization holders must submit a periodic safety report to the European Medicines Agency. This report must contain summaries of data relevant to the benefits and risks of the medicinal product, including results of all studies with a consideration of their potential impact on the market authorisation, a scientific evaluation of the risk-benefit balance of the medicinal product and all data relating to the volume of sales of the medicinal product. The frequency of these reports shall be specified in the marketing authorisation by the relevant authorities.
Regulation on Medicinal Products for paediatric use (Regulation (EC) No 1901/2006)	<i>Article 34</i>	Applicants must detail measures to ensure efficacy and possible adverse reactions.	Applicants must detail the measures to ensure the follow-up of efficacy and of possible adverse reactions to the paediatric use of a medicinal product. Additional reports may be requested by competent authorities. The details on when these reports must be performed and submitted for review can be included in the market authorisation conditions.

A3.4 Chemical

<u>Legislation</u>	<u>Article</u>	<u>Reporting and compliance obligations</u>	<u>Type of reporting activities</u>
Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) (Regulation (EC) No 1907/2006)	Article 6	Registration of substances	Manufacturers and importers of chemicals are required to collect information on the properties and uses of substances they manufacture in amounts that exceed one tonne per year. They also need to assess the hazards and potential risks. This information is communicated to the ECHA via a registration dossier.
	Article 7.1	Registration and notification of substances in articles	Any producer or importer of articles shall register to the Agency for any substances contained in those articles, if two conditions are met. The two conditions are the following: the substance must be present in the articles in quantities totalling over 1 tonne per producer or importer per year and the substance must be intended to be released under normal or reasonably foreseeable conditions of use.
	Art. 7.2	Notify ECHA	Any producer or importer is required to notify the ECHA if a substance meets the criteria in Art. 57 (carcinogenic, germ cell mutagenic, reproductively toxic, persistent bioaccumulative toxic, very persistent and bioaccumulative substances) and is identified in accordance with Art. 59(1) by the European Chemicals Agency and is present in article above a concentration of 0.1% w/w, and is present in those articles here the concentration exceeds 0.1% w/w in quantities totalling over 1 tpa.
	Article 9	Product and process orientated research and development (PPORD) notification	Manufacturers and importers of substances must notify ECHA of substances under research and development.
	Article 14	Preparing / updating Chemical Safety Reports (CSR)	A CSR is required to be completed for all substances subject to registration and in quantities of 10 tonnes or more per year per registrant. It documents the chemical safety assessment conducted in accordance with paragraphs 2 to 7 and with Annex I. Registrants are required to keep the CSR available and up to date.
	Article 22	Duties of registrants	Following registration, a registrant shall update his registration without undue delay with relevant new information and shall submit it to the Agency in the cases set out in article 22.1.
	Article 26	Inquiry	Manufacturers and importers of substances are required to submit a dossier and inquire from the Agency whether a registration has already been submitted for the same substance.
	Article 31 (requirements detailed in Annex II)	Preparing / updating Safety Data Sheets (SDS)	Suppliers of a substance or a preparation shall provide the recipient of the substance or mixture with an SDS if the substance meets the criteria for classification as hazardous in accordance with Directives 67/548/EEC or 1999/45/EC, or is persistent, bioaccumulative and toxic or very persistent and very bioaccumulative (Annex XIII), or included in the list established in accordance with Article 59(1) by the European Chemicals Agency for other reasons than those mentioned above.

	Article 32	Duty to communicate information down the supply chain	Suppliers of a substance on its own or in a preparation who do not have to supply a safety data sheet in accordance with Article 31 shall provide recipients with the following information: registration number(s), if the substance is subject to authorisation and details of any authorisation granted or denied under Title II in this supply chain, details of any restriction imposed under Title VIII, and any other available and relevant information about the substance that is necessary to enable appropriate risk management measures to be identified and applied including specific conditions resulting from the application of section 3 of Annex XI.
	Art. 33	Communicate information to consumers and professional customers	Producers are required to communicate sufficient information to the recipients of substances from the Candidate List to ensure safe use of the article.
	Article 34	Duty to communicate information on substances and preparations up the supply chain	Any actor in the supply chain of a substance or a preparation shall communicate some information to the next actor or distributor up the supply chain. The information is the following: new information on hazardous properties, regardless of the uses concerned and any other information that might call into question the appropriateness of the risk management measures identified in a safety data sheet supplied to him, which shall be communicated only for identified uses. Distributors must pass on that information to the next actor or distributor up the supply chain.
	Article 36	Obligation to keep information	Manufacturers, importers downstream users and distributors shall assemble and keep available all the information required to carry out duties under the Regulation of at least 10 years after they last manufactured, imported, supplied or used the substance or preparation.
	Article 38	Downstream Users notification	Under specific conditions, Downstream Users have to notify the ECHA of unsupported uses (i.e., uses outside the conditions described in the exposure scenario provided by the supplier). They also have to report classification differences with their suppliers to ECHA. This reporting needs to occur within six months of the receipt of the safety data sheets for a registered substance.
	Article 62	Submit authorization applications	Manufacturers are required to submit an application for an authorisation for substances to the Agency.
	Article 66	Downstream Users notification	Downstream Users that use a substance included in the Authorisation List, for which an authorisation has already been granted, must notify ECHA of their use within three months of the first supply of the substance.
	Entry 78 in Annex 78	Annual reporting to ECHA	Companies placing synthetic polymer microparticles on the market for consumer or professional use must annually report estimated environmental emissions to ECHA
Regulation on classification, labelling and packaging of substances and mixtures (CLP) (Regulation, Regulation (EC) No 1272/2008)	Art. 36, 37	Harmonised classification dossier	Manufacturers and importers of substances can submit dossiers for harmonised classification and labelling for industrial chemicals (not biocides and pesticide actives).
	Art. 40	Notify the ECHA	When placing a hazardous substance on the market, manufacturers are required to notify the ECHA of its classification and labelling within one month after their placing on the market.
	Art. 41	Notify the ECHA	If this notification results in different entries on the inventory for the same substance, the notifiers and registrants shall come to an agreed entry to be included in the inventory. The notifiers shall inform the ECHA accordingly.
	Art. 45	Providing information to responsible bodies	Suppliers must provide information on hazardous mixtures to nationally designated bodies (such as poison control centres).

Biocidal Products Regulation (BPR) [Regulation (EU) No 528/2012]	Art. 4	Active Substance Approval	Companies are required to apply for approval of an active substance by submitting a dossier to the ECHA. The ECHA carries out a validation check, followed by a completeness check and an evaluation by the competent evaluating competent authority. The approval is valid for a maximum of 10 years. To renew the approval, companies must submit another application to the ECHA. The content of applications is specified in Art. 6 and the way to submit it in Art. 7.
	Art. 6	Application Data Requirements	An approval of an active substance must contain a dossier for the active substance (following the requirements of Annex II), a dossier satisfying the requirements of Annex III, and where the substance would normally be excluded from approval, evidence that an exemption from the exclusion is applicable.
	Art. 17	Product Authorisation Applications	Biocidal products can only be legally placed on the EU market if they have been authorised. Three types of authorisations are possible: national, Union, mutual recognition procedure. The procedures for submitting applications are detailed in articles 20-34. All procedures require the submission of an application dossier.
	Art. 47	Notification of unexpected or adverse effects	Holders of authorisation are obliged to notify the competent authority who granted the authorisation if they become aware of information concerning the authorised biocidal product (or the active substance). The following information should be communicated: new data / information on the adverse effects; data indicating the potential of the active substance for the development of resistance); data indicating that the biocidal product is not sufficiently effective.
	Art. 58(3) and 58(5)	Treated Articles Notification	Companies placing a treated article on the market are required to ensure that the label of the product contains specific information. They are also required to provide consumers (upon request) with information on the biocidal treatment of the treated article within 45 days.
	Art. 65(2)	Maintaining documentation	Manufacturers of biocidal products placed on the market are required to maintain (in paper or electronic format) appropriate documentation (safety data sheets and specifications of active substances and ingredients; records of the various manufacturing operations performed; results of internal quality controls; identification of production batches).
	Art. 68	Record-keeping	Authorisation holders must keep records of biocidal products they place on the market for at least 10 years after placing the product on the market, or 10 years after the date on which the authorisation was cancelled or expired (whichever is earlier). Upon request by a competent authority, they must provide these records.
	Art. 69	Labelling	Authorisation holders must classify, package and label biocidal products in accordance with the approved summary of biocidal product characteristics. This labelling may not be misleading, especially in respect to the risks of the product to human health, animal health, the environment or its efficacy.
	Art. 95	Article 95 List Notification	Any person wishing to place on the market an active substance (on its own or in a biocidal product) needs to submit to the ECHA a dossier complying with the requirements of Annex II (or, when appropriate, Annex IIA to Directive 98/8/EC); or a letter of access to a dossier; or a reference to a dossier for which all data protections have expired.

Fertilising Products Regulation (FPR) (Regulation (EU) 2019/1009) Note: only the obligations of manufacturers are listed (importers and distributors are not covered in the analysis)	Art. 6(2)	Draw up technical documentation	Before placing a product on the market, manufacturers are required to compile the technical documentation and carry out the conformity assessment procedure or have it carried out. If compliance was demonstrated using the conformity assessment procedure, then they will write an EU declaration of conformity and affix the CE marking.
	Art. 6(3)	Store documentation	Manufacturers shall keep the technical documentation and EU declaration of conformity for a period of 5 years after the product has been placed on the market.
	Art. 6(4)	Keep registers	If necessary, manufacturers may keep a register of complaints, non-conforming EU fertilising products, and recalls of such EU fertilising products.
	Art. 6(5)	Identification	Manufacturers shall ensure that the packaging of the EU fertilising products which they have placed on the market contains elements allowing their identification (type number, batch number etc.). When the products are supplied without packaging, the required information must be provided in a document accompanying each fertilising product.
	Art. 6(6)	Traceability	Manufacturers shall indicate their name, registered trade name or registered trademark and the postal address at which they can be contacted on the packaging of the EU fertilising product. When the products are supplied without packaging, this information shall be provided in a document accompanying the product.
	Art. 6(7)	Labelling	Manufacturers are required to label products in accordance with Annex III.
	Art. 6(8)	Inform authorities of non-compliance	Manufacturers are required to inform the competent national authorities of the Member States in which they made the EU fertilising product available whenever they have reason to believe that an EU fertilising product which they have placed on the market presents a risk to human, animal, or plant health, to safety or to the environment.
	Art. 6(9)	Provide information	Upon a reasoned request from a competent national authority, manufacturers will provide all the information and documentation (in paper or electronic form) to demonstrate conformity with this regulation.
	Art. 16	EU declaration of conformity	Manufacturers are required to draw up an EU declaration of conformity using the model structure from Annex V, containing the elements specified in the relevant modules set out in Annex IV, and be continuously updated. The manufacturer assumes the responsibility for the compliance of the product placed on the market with the requirements laid down in this regulation.
Corporate Sustainability Reporting Directive (CSRD) (Directive (EU) 2022/2464)	Art. 19a	Sustainability report	Companies in the scope of the CSRD must include information related to the sustainability aspects of their business and business model in a dedicated section of their annual management report. SMEs may utilize a more limited form of sustainability reporting.
	Art. 29a	Sustainability reporting standards	Companies in scope of the CSRD must report using Commission-made standards on the environment, including water use, pollution and ecosystems, as well as social, human rights and governance factors.
Waste Framework Directive (WFD) (Directive 2008/98/EC)		SCIP Notification	Manufacturers, importers, and distributors of articles containing Substances of Very High Concern (SVHCs) above 0.1% weight by weight must submit information to the SCIP database (Substances of Concern In articles as such or in complex objects (Products)), managed by ECHA

Prior Informed Consent Regulation (PIC) (Regulation (EU) 649/2012)	Art. 10	Reporting obligations	Exporters are required to submit an annual report to the Designated National Authorities containing exact quantities of each chemical listed in Annex I that was exported to or imported from non-EU countries, and any additional information requested by a competent authority and required for the implementation of the regulation.
	Art. 8(2)	Notification of exports	Exporters are required to notify the designated national authority whenever they intend to export a chemical referred to in paragraph 1. The notifications shall comply with the information requirements laid down in Annex II.

A3.5 Energy

<u>Legislation</u>	<u>Article</u>	<u>Reporting and compliance obligations</u>	<u>Type of reporting activities</u>
Energy Efficiency Directive (Consolidated Directive 2012/27/EU)	Article 7a (5)	Registration in energy efficiency obligation schemes	Obligated parties (nationally designated) must measure, document and verify on a statistically significant proportion and representative sample of the energy efficiency improvement measures put in place. These reports are then published by Member States yearly.
	Article 8 (4)	Energy audits and energy management systems	Enterprises (not SMEs) are subject to an energy audit carried out in an independent and cost-effective manner by qualified and/ or accredited experts or implemented and supervised by independent authorities under national legislation at least every 4 years from the date of the previous energy audit.
	Article 8 (6)	Energy audits and energy management systems	Enterprises (not SMEs) that are implementing an energy or environmental management system (certified by an independent body according to the relevant European or International Standards) shall be exempted from the requirements of paragraph 4, provided that Member States ensure that the management system concerned includes an energy audit on the basis of the minimum criteria based on Annex VI.
REMIT - Regulation on wholesale Energy Market Integrity and Transparency (Regulation (EU) No 1227/2011)	Article 4	Obligation to publish inside information	Market participants ⁷⁷⁷ shall publicly disclose in an effective and timely manner inside information which they possess in respect of business or facilities which the market participant concerned, or its parent undertaking or related undertaking, owns or controls or for whose operational matters that market participant or undertaking is responsible, either in whole or in part.
	Article 5a	Notification of algorithmic trading	Market participants must notify that they will engage in algorithmic trading to the national regulatory authority and the Agency for the Cooperation of Energy Regulators ('ACER').
	Article 7c	Provision of LNG market data	LNG market participants shall submit daily to ACER the LNG market data, free of charge, through reporting channels established by ACER.

⁷⁷⁷ A **market participant** under Regulation (EU) No 1227/2011 is defined as any person, including transmission system operators, who enters into transactions, including the placing of orders to trade, in one or more wholesale energy markets.

	Article 8	Data collection	(1) Market participants shall provide ACER with a record of wholesale energy market transactions, including orders to trade. The information reported shall include the precise identification of the wholesale energy products bought and sold, the price and quantity agreed, the dates and times of execution, the parties to the transaction and the beneficiaries of the transaction and any other relevant information. (5) Market participants shall provide ACER and national regulatory authorities with information related to the capacity and use of facilities for production, storage, consumption or transmission of electricity or natural gas or related to the capacity and use of LNG facilities, including planned or unplanned unavailability of these facilities, for the purpose of monitoring trading in wholesale energy markets.
	Article 9	Registration of market participants	Market participants entering into transactions which are required to be reported to ACER in accordance with Article 8(1) shall register with the national regulatory authority in the Member State in which they are established or resident or, if they are not established or resident in the Union, in a Member State in which they are active.
	Article 13b	Request for information	ACER may request any person for information necessary for fulfilling the ACER's obligations regarding market abuse.
	Article 15	Notify Agency in case of suspicious events	Persons professionally arranging or executing transactions in wholesale energy products must notify ACER and national regulatory authority upon becoming aware of a suspicious event.
EU Emissions Trading System Directive (EU ETS) (Consolidated Directive 2003/87/EC)	Article 7	Changes relating to installation	The operator ⁷⁷⁸ shall inform the competent authority of any changes planned in functioning, or an extension of the installation which may require updating of the greenhouse gas emissions permit.
	Article 12(3)	Transfer, surrender and cancellation of allowances	Operators of each installation shall surrender a number of allowances equal to the total emission from that installation during the preceding calendar year.
	Article 14	Guidelines for monitoring and reporting of emissions	Operators of an installation shall report the emissions from that installation during each calendar year to the competent authority after the end of that year in accordance with the guidelines and implementing acts by the Commission.

⁷⁷⁸ Under the EU ETS directive (2003/87/EC) 'operator' means any person who operates or controls an installation for activities listed in Annex I (like combustion of fuels, refining of oil, production of coke), or where this is provided for in national legislation, to whom decisive economic power over the technical functioning of the installation has been delegated.

Industrial Emissions Portal / E-PRTR (Regulation (EC) No 166/2006)	Article 5	Reporting by operators	The operator⁷⁷⁹ of each facility that undertakes one or more of the activities specified in Annex I (mineral oil and gas refineries, installations for gasification and liquefaction, thermal power stations and other combustion installations, coke ovens, coal rolling mills and installations for the manufacture of coal products and solid smokeless fuel) above the applicable capacity thresholds specified therein shall report the amounts annually to its competent authority, along with an indication of whether the information is based on measurement, calculation or estimation. They shall also communicate to its competent authority the information identifying the facility in accordance with Annex III unless that information is already available to the competent authority.
	Article 6	Releases to land	The operator shall report waste subject to 'land treatment' or 'deep injection' disposal operations, as specified in Annex IIA to Directive 75/442/EEC, as a release to land only if the waste originates from their own facility.
	Article 9	Quality assurance and assessment	The operator of each facility subject to the reporting requirements set out in Article 5 shall assure the quality of the information that they report.
Regulation on submission and publication of data in electricity markets (Commission Regulation (EU) No 543/2013)	Article 6	Information on total load	TSOs (Transmission System Operators, entities responsible for managing and operating electricity transmission networks) shall calculate and submit data as set out in article 6 to the ENTSO for Electricity for each bidding zone.
	Article 7	Information relating to the unavailability of consumption units	TSOs shall provide information concerning unavailability of consumption units to the ENTSO for Electricity.
	Article 8	Year-ahead forecast margin	For their control areas, TSOs shall calculate and provide for each bidding zone the year-ahead forecast margin evaluated at the local market time unit to the ENTSO for Electricity. The information shall be published one week before the yearly capacity allocation but no later than the 15th calendar day of the month before the year to which the data relates.
	Article 9	Transmission infrastructure	TSOs shall establish and provide information on future changes to network elements and interconnector projects including expansion or dismantling in their transmission grids within the next three years, to the ENTSO for Electricity. This information shall only be given for measures expected to have an impact of at least 100 MW on cross zonal capacity between bidding zones or on profiles at least during one market time unit.
	Article 10	Information relating to the unavailability of transmission infrastructure	TSOs shall calculate and provide information concerning the unavailability of transmission infrastructure to the ENTSO for Electricity.

⁷⁷⁹ Under the E-PRTR Regulation, an operator is defined as any natural or legal person who operates or controls the facility or, where this is provided for in national legislation, to whom decisive economic power over the technical functioning of the facility has been delegated.

Article 11	Information relating to the estimation and offer of cross zonal capacities	For their control areas TSOs or, if applicable, transmission capacity allocators, shall calculate and provide information relating to the estimation and offer of cross zonal capacities to the ENTSO for Electricity sufficiently in advance of the allocation process.
Article 12	Information relating to the use of cross zonal capacities	For their control areas TSOs shall calculate and provide information concerning the use of cross zonal capacities the ENTSO for Electricity.
Article 13	Information relating to congestion management measures	For their control areas TSOs shall provide information relating to congestion management measures to the ENTSO for Electricity.
Article 14	Forecast generation	For their control areas, TSOs shall calculate and provide information relating to forecast generation to the ENTSO for Electricity.
Article 15	Information relating to the unavailability of generation and production units	For their control areas, TSOs shall provide information relating to the unavailability of generation and production units to the ENTSO for Electricity
Article 16	Actual generation	For their control areas, TSOs shall calculate and provide information concerning actual generation to the ENTSO for Electricity
Article 17	Balancing	For their control areas, TSOs or where applicable operators of balancing markets, where such markets exist shall provide information related to balancing to the ENTSO for Electricity

Annex 4 Task 3: Policy fiches

This annex presents all the policy fiches the Study team has submitted under the contract, except pilot fiches.

A4.1 Data Availability and Transparency Act Scheme | Australia

Data Availability and Transparency Act Scheme Australia	
Summary	The Data Availability and Transparency Act Scheme (DATA Scheme) facilitates the sharing of Australian Government data among accredited users, including Australian Commonwealth, state and territory government bodies as well as Australian universities. The Scheme is established by the 2022 Data Availability and Transparency Act, which designates the National Data Commissioner as regulator of the DATA Scheme and provider of advice and guidance on its operation.⁷⁸⁰ The Scheme seeks to increase the availability and use of Government data by government bodies and universities to deliver quality public services, better inform public policies and programs as well as support research and development. Linked to these efforts, the Australian Government has developed a searchable Australian Government Data Catalogue and Dataplace - a platform for the Scheme participants and others to manage data requests and support the administration of the Scheme.
Geographic scope	Australia
Type of measure	Data sharing scheme. In force since 2022
Types of data covered	Personal data, non-personal data. Australian Government data lawfully collected, created or held by a Commonwealth body, or on its behalf
Data accessibility	Data accessibility is restricted to accredited users - Commonwealth, state, territory and government bodies as well as Australian universities. Access is requested to a data custodian through Dataplace - requests outside Dataplace are also allowed. Data custodians – entities within the Australian Government that are entrusted with managing specific datasets – must then consider the request and decide whether to agree to share data. Data sharing must contain certain information including a data sharing agreement and comply with additional requirements in case data sharing involves foreign nationals. On this basis, a data sharing agreement is signed and submitted for registration.⁷⁸¹
Relevant sectors	Public sector. Australian Commonwealth Government bodies, state government bodies, territory government bodies, Australian universities
Objectives	The main objectives of the Scheme are: - Establish a structured framework for data sharing within the public sector. - Promote safe and secure data sharing within the public sector. - Improve public services and better inform public policies and programs. - Support research and innovation. No explicit reference to FAIR Principles – Findable, Accessible, Interoperable, Reusable –, but close alignment with them.
Restrictions/safe-guards	Data sharing restrictions. For national security reasons, some bodies including intelligence and law enforcement entities are excluded from the Scheme as well as their data. Additional data held by Home Affairs and the Australian Transaction Reports and Analysis Centre (AUSTRAC) cannot be shared. Data sharing purposes. Data sharing under the Scheme must be based on one of the three permitted purposes: government service delivery, informing government policies and programs, research and development. Accreditation. Accreditation acts as a gateway into the DATA Scheme and ensures that data service

⁷⁸⁰ Office of the National Data Commissioner (2023). *Introducing the DATA Scheme*. Available online: <https://www.datacommissioner.gov.au/sites/default/files/2023-04/Introducing%20the%20DATA%20Scheme%20-%20April%202023.pdf>

The information presented hereafter refers to the DATA Scheme introduction document unless otherwise referenced.

⁷⁸¹ Office of the National Data Commissioner (2023). *Data sharing under the DATA Scheme*. Available online: <https://www.datacommissioner.gov.au/sites/default/files/2024-05/How%20to%20share%20data%20under%20the%20DATA%20Scheme.pdf>

	providers and users are capable of handling public sector data and minimising risk of unauthorised access or use. Both the Justice Minister and the National Data Commissioner are the accrediting responsible authorities and can impose conditions on accreditation if needed. Data sharing principles. The data sharing principles are the risk management framework that sits at the core of the Scheme to support data custodians in deciding if it is safe to share data. The principles cover the data sharing project, people, setting, data and output. The principles must be applied in such a way that, when viewed as a whole, the risks in sharing, collecting and using data is appropriately mitigated. Foreign nationals access. In the event of a data sharing agreement permitting access to a foreign individual, the responsible user must provide the Australian Security Intelligence Organisation (ASIO) with relevant information including a copy of the signed data sharing agreement and the details of the foreign individual(s) involved. Privacy protections. The Scheme is subject to the Australian Privacy Act. It includes protections that minimise the sharing of personal information and prohibit the re-identification of de-identified data as well as the storage or access of personal information outside Australia. Express consent is required to share biometric data. The Act also contains purpose-specific privacy protections, depending on the data sharing purpose of the project.
Data governance structure	National Data Commissioner. The Commissioner regulates and enforces the Scheme through their regulatory functions. These are: accrediting eligible entities; handling complaints from Scheme entities and others; assessing and investigating Scheme entities, taking enforcement action such as issuing infringement notices and directions, and/or seeking injunctions as well as civil and criminal penalties; transferring matters to another appropriate authority. Accredited data service providers. Commonwealth, state and territory government bodies as well as Australian universities that can provide specialised data services including complex data integration, de-identification and secure access services to support data sharing projects. Entities must apply to be accredited as data service providers. Data custodians. Australian Government bodies linked to accredited data service providers. Data custodians are automatically participants of the DATA Scheme and do not need to apply for this role. Accredited users. Commonwealth, state and territory government bodies as well as Australian universities. Entities must apply to be accredited as users.
Data localization requirements	No specific data localisation requirements. For personal information, access from outside Australia is prohibited.
Available data infrastructures	As a data sharing scheme based on best practices, safeguards and restrictions, no specific data infrastructure. Individual accredited data service providers and users must ensure they have data infrastructures capable of handling public sector data and minimising risks related to unauthorised access or use. To support the DATA Scheme, the Australian Government developed Dataplace, which features the Australian Government Data Catalogue which includes the data inventories of both open and restricted data of Australian Government agencies.⁷⁸²
Effectiveness of enforcement mechanism	Procedural enforcement mechanisms to ensure compliance with the 2022 Data Availability and Transparency Act through technical enforcement, accreditation systems and regulatory powers under the National Data Commissioner. More precisely, the Act includes penalties for unauthorised sharing, collection or use; civil penalty provisions; an accreditation framework for the imposition, renewal, variation or removal (suspension or cancellation) of accreditation; monitoring and investigation powers as well as issuing of directions by the Commissioner; transfer of matters to appropriate authority.⁷⁸³ As of June 2025, there were 31 data users and 13 data service providers accredited as well as 56,031 records available on the Australian Government Data Catalogue and 22 agencies contributing records directly from their data inventory.⁷⁸⁴ There are no publicly reported instances of civil penalties, infringement notices or removals of accreditation.
Comparison to EU initiatives	The DATA Scheme follows a rather restrictive understanding of data sharing and the development of data spaces, as access is limited to public entities and universities. In this sense, despite the access

⁷⁸² Office of the National Data Commissioner (2025). *Dataplace*. Available online: <https://www.dataplace.gov.au/>

⁷⁸³ Department of Finance (2022). *Data Availability and Transparency Act (No.11, 2022) - An Act to authorise the sharing of public sector data, and for related purposes*. Federal Register of Legislation. Available online: <https://www.legislation.gov.au/C2022A00011/asmade/downloads>

⁷⁸⁴ Office of the National Data Commissioner (2024). *DataPoints: June 2025 Edition*. Available online: <https://www.datacommissioner.gov.au/node/317>

	limitations and other particularities, both the EU and Australia have undergone parallel efforts to facilitate data availability, transparency and ultimately, the development of data spaces. In 2022, the EU introduced the European Data Governance Act for cross-sector data sharing and to support the setup and development of the Common European Data Spaces, and the Data Act in 2024, which establishes clear rules for access and use of data within the European data economy.⁷⁸⁵ However, both initiatives have a broader scope than the Australian DATA Scheme in terms of the type of data shared, the size of the data spaces linked and the targeted users of the shared data. By contrast, the DATA Scheme presents a structured and secure framework to share Government data between public entities and academic institutions, which supports public policy development, decisions and research. Taking into account the unparalleled amount of data developed by EU institutions, agencies and bodies, the Australian Scheme presents a potential mechanism for improved data sharing between the EU, national and regional governments.

A4.2 Guidance on privacy and developing and training generative AI models | Australia

Guidance on privacy and developing and training generative AI models Australia	
Summary	The Australian Guidance on privacy and developing and training generative AI models⁷⁸⁶, published in 2024 by the Office of the Australian Information Commissioner in collaboration with the Department of Industry, Science and Resources. It aims at ensuring compliance of generative AI development with Australian law - Privacy Act 1988 and Australian Privacy Principles guidelines (APPs). In particular, the Guidance seeks to reinforce the protection of personal data, and especially sensitive information, during the development and training of generative AI models. The Guidance is part of larger efforts carried out by the Australian government to regulate generative AI, with initiatives such as the Guidance on privacy and the use of commercially available AI products or the Proposals Paper for introducing mandatory guardrails for AI in high-risk settings.
Geographic scope	Australia
Type of measure	Guidelines. In force since October 2024
Types of data covered	Personal data. All types of personal information, defined by the Privacy Act as information or an opinion about an identified individual, or an individual who is reasonably identifiable.⁷⁸⁷ The definition of personal information covers sensitive information, which includes racial or ethnic origin, political opinions, membership of a political association, religious beliefs or affiliations, philosophical beliefs, membership of a professional or trade association, membership of a trade union, sexual orientation or practices, criminal record, health, genetic and biometric information, biometric templates.
Relevant sectors	Public and private sector
Objectives	The main objectives of the Guidance are: - Ensure compliance of developers with the Privacy Act and the Australian Privacy Principles guidelines. - Promote Privacy by Design process, based on embedding good privacy practices into the design specifications of technologies, business practices and physical infrastructures. - Limit the use of personal information for AI development and training. - Reinforce the principle of consent for the collection of sensitive information.
Restrictions/safe-	Accuracy in generative AI models: Developers must take reasonable steps to ensure accuracy,

⁷⁸⁵ European Commission (2024). *European Data Governance Act*. Available online: <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act>; European Commission (2024). *Data Act*. Available online: <https://digital-strategy.ec.europa.eu/en/policies/data-act>

⁷⁸⁶ Office of the Australian Information Commissioner (2024). *Guidance on privacy and developing and training generative AI models*. Available online: <https://www.oaic.gov.au/privacy/privacy-guidance-for-organisations-and-government-agencies/guidance-on-privacy-and-developing-and-training-generative-ai-models#section-other-considerations>

The information presented hereafter refers to the Guidance on privacy and developing and training generative AI models unless otherwise referenced.

⁷⁸⁷ Attorney-General's Department and Department of the Treasury (1988). *Privacy Act 1988 (Act No. 128, 2024)*. Federal Register of Legislation. Available online: <https://www.legislation.gov.au/C2004A03712/latest/text>

guards	taking into account the likely increased level of risk in an AI context, through the use of high-quality datasets and undertaking appropriate testing. Legal compliance for personal data: Developers must consider whether data they intend to use or collect contains personal information, and comply with the obligations set by the Privacy Act and the APPs. Requirement of consent: Many data contain sensitive information and therefore may not be scraped from the web or collected from a third-party dataset without formal consent. Where developers seek to use personal information, they already hold for the purpose of AI model training, and this was not a primary purpose of collection, they must either collect a formal consent for use or establish that this secondary use would be reasonably expected by the individual. Privacy by Design: Developers subject to the Privacy Act must take reasonable steps to implement practices, procedures and systems that will ensure they comply with the APPs and any binding registered APP code, and are able to deal with related inquiries and complaints. When developing or fine-tuning a generative AI model, developers should consider the potential risks at the planning and design stage through a “privacy by design” approach. Privacy obligation regarding accuracy: Developers must ensure that personal information collected is accurate, up-to-date and complete, and that personal information used and disclosed is accurate, up-to-date, complete, relevant and having regard to the purpose of the use or disclosure.
Access control mechanisms	As a guidance on data privacy for the development and training of generative AI, no specific access control mechanisms to personal data.
Compliance burden	The Guidance on privacy and developing and training generative AI models aims to facilitate compliance of AI developers with the Privacy Act and the APPs. The main challenge identified is the lack of enforceable legislation governing responsible AI use, with the Australian Government currently developing mandatory regulatory safeguards.
Enforcement mechanism	The Guidance is intended to apply to all kinds of generative AI, including both Multimodal Foundation Models (MFM) and Large Language Models (LLMs), and considers the obligations that apply to entities subject to the Privacy Act. The enforcement mechanisms therefore fall under the Privacy Act and its specific provisions. The Privacy Act applies to Australian Government agencies, organisations with an annual turnover of more than \$3 million and other organisations. The Act also applies to acts or practices engaged in outside Australia by entities with an Australian link.
Effectiveness of enforcement mechanism	There is an increasing focus of enforcement and enquiry cases towards generative AI on the basis of the Privacy Act and the APPs. In 2021 Australian Information Commissioner and Privacy Commissioner found that Clearview AI, Inc. breached the Privacy Act by scraping biometric information from the web and disclosing it through a facial recognition tool. The determination ordered Clearview AI to cease collecting the data and destroying existing images and templates collected from Australia.⁷⁸⁸ Since 2024, the Office of the Australian Information Commissioner is carrying out a preliminary enquiry into allegations that patient data from radiology firm I-Med was used without consent to train healthcare technology company Harrison.ai’s AI models.⁷⁸⁹
Alignment with international frameworks	The Australian Guidance on privacy and developing and training generative AI models is closely aligned with international efforts to regulate generative AI and build trustworthy data governance frameworks. In this sense, the Australian Privacy Commissioner signed a joint declaration to establish data governance that fosters innovative and privacy-protective AI together with data protection authorities from Korea, Ireland, France and the UK. ⁷⁹⁰ Furthermore, the Australian Guidance has sought guidance in similar documents such as the UK’s Information Commissioner’s Office’s consultation series on generative AI, the French “AI how to sheets” and the Canadian principles for responsible development and use of generative AI. In this sense, the Guidelines have sought - where possible - to align with international guidance on privacy for generative AI, including OECD ⁷⁹¹ , in order to reduce the burden on Australian businesses operating globally ⁷⁹² .

⁷⁸⁸ Office of the Australian Information Commissioner (2021). *Clearview AI breached Australians’ privacy*. Available online: <https://www.oaic.gov.au/news/media-centre/clearview-ai-breached-australians-privacy>

⁷⁸⁹ Lund, Siska (2024). *I-MED’s Data Headache: Unauthorised Access & Unauthorised Disclosure Double Whammy*. Privacy 108. Available online: <https://privacy108.com.au/insights/imed-data-headache-double-whammy/>

⁷⁹⁰ Office of the Australian Information Commissioner (2024). *Joint statement on building trustworthy data governance frameworks to encourage development of innovative and privacy-protective AI*. Available online: <https://www.oaic.gov.au/news/media-centre/joint-statement-on-building-trustworthy-data-governance-frameworks-to-encourage-development-of-innovative-and-privacy-protective-ai>

⁷⁹¹ OECD (2024). *AI, Data Governance and Privacy: Synergies and areas of international co-operation*. *OECD Artificial Intelligence Papers*, No. 22. Available online:

Comparison to EU initiatives	The EU and Australia share significant legal parallels, with rules such as EU GDPR and Australia's Privacy Act mirroring each other despite certain differences in their provisions. Additionally, the EU 2024 Artificial Intelligence Act further reinforced the protection of data protection and privacy in contexts of generative AI. ⁷⁹³ Despite the important legal developments, the EU does not yet have a structured guidelines document that directly links the development and training of generative AI models to privacy requirements - mainly the GDPR. While the EU has provided several documents in this regard including EDPB's Opinion 28/2024 on personal data protection related to AI models ⁷⁹⁴ or EDPS' Orientations for EULs using Generative AI ⁷⁹⁵ , none of these provide clear, comprehensive guidance applicable to the entire public and private sectors.

A4.3 Open Industry | Brazil

Open Industry Brazil	
Summary	The Open Industry Initiative is a public-private partnership launched in 2021 by Brazil's Ministry of Economy and the National Confederation of Industry (CNI). It is supported among others by the Institute NEO from University of the state Rio Grande do Sul and FIEMG, the Industry Federation from the state Minas Gerais ⁷⁹⁶ . Key components of this initiative are to establish 12 regional hubs to provide SMEs with access to shared digital infrastructure (e.g., cloud platforms, 5G testbeds). Open Industry also aims for tax breaks for R&D investments and subsidised loans through the Brazilian Development Bank. Since the launch of this initiative there were several impacts made, such as more than 200 startups partnered with industries to co-develop solutions (e.g., AI-driven crop monitoring in agribusiness). ⁷⁹⁷ It created 35,000 tech-related jobs (2021–2023) and reduced industrial CO ₂ emissions by 8% in participating companies (2022–2023). Early adopters reported 15-20% of efficiency gain through predictive maintenance. The key success of the Open Industry is that there is public-private alignment assuring that CNI industry networks and federal funding is rapidly scaling. The initiative has established global partnerships, such as collaboration with Plattform Industrie 4.0 Germany and partnership with World Economic Forum (WEF), which boosted the Open Industry with technical expertise. However, despite of the impacts and key successes, the Open Industry also identified challenges. For example, only 12% of SMEs integrated Industry 4.0 tools by 2023 due to the high costs and skills gap. A second challenge is that there is also limited 5G coverage outside major cities of Brazil limiting the IoT deployment. ⁷⁹⁸
Geographic scope	Brazil
Type of measure	Industry Agreements
Types of data covered	Personal and non-personal data is covered.
Data accessibility	Open data includes aggregated industry performance metrics, while restricted data is patient records or proprietary R&D data. Open Industry initiative's collaboration with international partners involves data sharing under certain conditions, like General Data Protection Regulation (GDPR) compliance.

https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf

⁷⁹² Office of the Australian Information Commissioner (2024). *Can personal information be used to develop or train GenAI?* Available online: <https://www.oaic.gov.au/news/blog/can-personal-information-be-used-to-develop-or-train-a-generative-ai-model>

⁷⁹³ European Parliament and the Council (2024). *Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence and amending Regulations and Directives (Artificial Intelligence Act)*. Official Journal of the European Union. Available online: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:L_202401689

⁷⁹⁴ European Data Protection Board (2024). *Opinion 28/2024 on certain data protection aspects related to the processing of personal data in the context of AI models*. Available online: https://www.edpb.europa.eu/system/files/2024-12/edpb_opinion_202428_ai-models_en.pdf

⁷⁹⁵ European Data Protection Supervisor (2024). *Generative AI and the EUDPR. First EDPS Orientations for ensuring data protection compliance when using Generative AI systems*. Available online: https://www.edps.europa.eu/system/files/2024-06/24-06-03_genai_orientations_en.pdf

⁷⁹⁶ Open Industry Official website: <https://www.openindustry.org.br/>

⁷⁹⁷ OECD Taxation Working Papers no.27. available online: https://www.oecd.org/content/dam/oecd/en/publications/reports/2016/09/fiscal-incentives-for-r-d-and-innovation-in-a-diverse-world_g17a286d/5jlr9stckfs0-en.pdf

⁷⁹⁸ Hannover Messe 2025. Video Recording on Data Spaces and interoperability in Brazilian industrial environment. Available online: <https://www.hannovermesse.de/event/data-spaces-and-interoperability-in-brazilian-industrial-environment/pan/11354>

	The use of blockchain for traceability in agribusiness (see section on infrastructure), which have open aspects but with controlled authenticity for access. Lastly, the legal frameworks governing restrictions are Lei General de Proteção (LGPD), sector-specific laws, and international agreements. Compliance mechanisms like certifications and audits ensure that data restrictions are enforced.
Relevant sectors	Manufacturing, Agriculture, Energy and Logistics/Infrastructure
Objectives	The main objectives are: To accelerate adoption of Industry 4.0 technologies in targeted sectors. To foster collaboration between industries, startups, and research institutions. To promote circular economy practices and reduce carbon footprints in sectors like automotive, chemicals, and agribusiness. To train 500,000 workers in digital skills by 2025. To prioritise industrial modernisation in underserved regions (e.g., Northeast Brazil).
Restrictions/safeguards	The Open Industry in Brazil includes specific restrictions to protect national interests and safeguards to ensure equitable growth and compliance. ⁷⁹⁹ The initiative balances economic openness with strategic protections for sovereignty, labour, and the environment. The restrictions of the initiative are most limited to foreign companies in sensitive sectors such as nuclear energy, healthcare, postal services, domestic aviation, and aerospace. There are also restrictions in industrial participation requirements in certain sectors for instance in mining has to mandate minimum local procurement (60% in offshore projects). Another restriction is for foreign companies in specific/strategic sectors such as defence. In this scenario, the companies must share proprietary technology with at least one Brazilian partner.
Data governance structure	This initiative has a mix of top-down and multi-stakeholder governance structure. The government sets regulations and standards, but collaboration between private sectors and civil society suggests a multi stakeholder approach. The Ministry of Economy and CNI are the primary actors. Other key stakeholders include the Brazilian Development Bank (BNDES), industry associations (e.g. ABIMAQ, ANFAVEA), SMEs, academia, and international partners like Germany's Plattform Industrie 4.0 or WEF. The Ministry of Economy set policies and provides funding and CNI role is to coordinate industry efforts. BNDES role is to handle financial support whereas industry leaders such as Vale or Petrobras have a role to drive sector-specific adoption. Furthermore, state governments (e.g. Sao Paulo, Minas Gerais) implements regional Industry 4.0 hubs and support the Open Industry initiative. The Open Industry initiative covers operational, sustainability, product, workforce, innovation, and regulatory data. For example, operational data from IoT devices in manufacturing is governed by industry standards, while sustainability data has stricter government regulations. Personal data requires compliance with the LGPD.
Data localisation requirements	Open Industry initiative in Brazil does not universally mandate data storage within national borders, but specific sectors and data types are subject to localisation rules under broader Brazilian laws. LGPD does not require data localisation by default, but cross-border transfers must comply with safeguards, including contractual clauses. Energy, telecommunications, and defence industries require local storage under sector-specific regulations. ^{800 801} Moreover, there are some localisation mandates for specific sectors: Healthcare – under legal basis of Lei no 13.709/2018 (LGPD and SUS norms the jurisdiction requirement is that patient data must be stored locally. Government contracts – under legal basis Decree 10.306/2020 (IT Governance) the jurisdiction requirement is that data generated under federal contracts require local storage. Critical infrastructure – operational data from power grids or oil refineries must reside in Brazil. There are exceptions to localisation requirements. For instance, data shared with foreign partners can be stored abroad if encrypted and compliant with LGPD's cross-border transfer rules. Multinational cloud providers such as AWS can host data if they comply with LGPD and use Brazilian data centres for certain sectors such as finance. Lastly, there is an exception for non-sensitive data, where aggregated operational data may be stored offshore.
Available data infrastructures	The core infrastructure of Open Industry consists of 12 regional hubs, which provide shared access to advanced manufacturing services including digital twins, simulation platforms for optimising production lines. Cloud services are also used: national platforms include a government cloud for

⁷⁹⁹ CNI (2024), *Integrity and Compliance Policy Paper*. Available online: https://g20.utoronto.ca/b20/2024-B20_Policy_Paper_Integrity-&-Compliance2.pdf

⁸⁰⁰ Brazil's General Data Protection Law / Lei Geral de Proteção de Dados (LGPD) – an overview. Available online: [https://usercentrics.com/knowledge-hub/brazil-lgpd-general-data-protection-law-overview/#:~:text=The%20Lei%20Gera%20de%20Prote%C3%A7%C3%A3o%20de%20Dados,Dados%20\(ANPD\)%20or%20National%20Data%20Protection%20Authority.](https://usercentrics.com/knowledge-hub/brazil-lgpd-general-data-protection-law-overview/#:~:text=The%20Lei%20Gera%20de%20Prote%C3%A7%C3%A3o%20de%20Dados,Dados%20(ANPD)%20or%20National%20Data%20Protection%20Authority.)

⁸⁰¹ Digital Business Laws and Regulations Brazil 2024-2025. Available online: <https://iclg.com/practice-areas/digital-business-laws-and-regulations/brazil>

	public sector integration (SEPRO Cloud) and a unified portal (gob.br) for accessing industrial datasets. Open Industry also has sector-specific data spaces, such as AgriData, a shared platform for agribusiness stakeholders to exchange IoT farm data and Energy Data Space, which integrates smart grid data from renewable projects.⁸⁰² Open Industry has established interoperability frameworks: Technical standards, which includes ISO/IEC 30141 adopted for cross-sectors IoT device integration (e.g., mining sensors to cloud platforms) and OPC UA, an industrial automation protocol used in automotive and chemical plants for machine-to-machine communication. National Frameworks, which includes Brazilian Data Governance Framework, aligning sectoral data-sharing rules with LGPD (e.g., anonymized health data in pharmaceuticals), and Conecta Industria, which is a federated data exchange platform linking Brazilian Industry 4.0 hubs, SMEs, and research institutions.⁸⁰³ The initiative has public data APIs: Brazilian Institute of Geography and Statistics APIs for regional economic indicators and National Petroleum Agency APIs for real-time fuel production data. Furthermore, open-source APIs for predictive maintenance have been developed (e.g., vibration data from machinery). Open Industry also offers blockchain solutions for traceability. For instance, AgriBlock Chain is developed by Embrapa to track soy and beef exports to the EU.⁸⁰⁴
Effectiveness of enforcement mechanism	The enforcement mechanism includes regulatory mandates, financial incentives and multi-stakeholder supervisions. Regulatory mandates include sectorial agencies, for instance ANFAVEA for automotive sector to enforce technical standards through periodic inspections. Moreover, the initiative is strictly supervised by the Brazilian Institute of Environment (IBAMA) for projects in sensitive regions, where it is required to provide environmental impact assessment. The Open Industry initiative also has created groups like CNI-led coalitions of audit is to share infrastructure, including Industry 4.0 hubs and publish compliance scores. Moreover, the Ministry of Economy publishes dashboards tracking regional progress (e.g. 5G rollout, SME adaptation rates). The 2023 National Confederation of Industry (CNI) survey revealed that only 12% of Brazilian SMEs had adopted Industry 4.0 tools, reflecting significant barriers to digital transformation.⁸⁰⁵ Key challenges included economic constraints (e.g., high implementation costs), limited technical expertise, and insufficient organizational readiness to integrate advanced technologies like IoT and AI.⁸⁰⁶ Despite government efforts, such as funding for smart factories and digital maturity assessments, adoption remains uneven, with smaller companies lagging behind larger enterprises due to resource limitations. The survey underscores the need for targeted support, including workforce training, financial incentives, and ecosystem collaboration, to bridge the Industry 4.0 gap in Brazil's industrial sector.
Comparison to EU initiatives	Brazil's Open Industry Platform and EU initiatives share core objectives of fostering digital ecosystems for SMEs through collaborative frameworks, data interoperability, and workforce upskilling. Both emphasize API-driven integration (e.g., Brazil's open banking and the EU's Digital Single Market) and aim to democratise access to Industry 4.0 tools. Initiatives like Brazil's SENAI 4.0 and the EU's Digital Innovation Hubs similarly target skill gaps, while ecosystem-building efforts (GAIA-X in Europe vs. Open Industry Brazil) reflect parallel strategies to enhance SME competitiveness. However, both face challenges like uneven adoption, with Brazil's 12% Industry 4.0 uptake among SMEs mirroring lagging participation in some EU regions despite stronger EU-wide averages. Key differences lie in regulatory maturity and priorities. The EU enforces stringent, harmonised policies (GDPR, Green Deal Industrial Plan) that prioritise sustainability and cross-border data governance, whereas Brazil's framework remains fragmented, focusing on basic digitalisation and infrastructure upgrades. Funding models diverge sharply: the EU leverages centralised programmes

⁸⁰² Open Industry (2025), *2nd Webinar of the "Data Spaces Knowledge Journey"*. Available online: <https://www.openindustry.org.br/conhecimento/webinar-sobre-data-spaces?c=webinars>

⁸⁰³ GovTech Intelligence hub (2025), *Conecta gov.br: Brazil's National Interoperability Platform for Automated Data Exchange*. Available online: <https://www.govtechintelhub.org/case-study-details/conecta-gov.br-brazil's-national-interoperability-platform-for-automated-data-exchange/aJYTG000000Yxl4AE>

⁸⁰⁴ AL-INVEST Verde (2022), *Traceability initiatives in the Beef and Leather Value Chains in Brazil*. Available online: https://alinvest-verde.eu/wp-content/uploads/2023/02/Iniciativas_rastreabilidade_ENG.pdf

⁸⁰⁵ Aniceski et al. (2024), *The Four Smarts of Industry 4.0 and barriers for technology deployment: A TOE perspective*. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S0360835224004662>

⁸⁰⁶ Collins et al. (2024), *Assessing the Readiness of Small and Medium-Sized Enterprises (SMEs) for Industry 4.0 Adoption in Manufacturing*. Available online: https://www.researchgate.net/publication/387790548_Assessing_the_Readiness_of_Small_and_Medium-Sized_Enterprises_SMEs_for_Industry_40_Adoption_in_Manufacturing

	such as Horizon Europe, DEP, while Brazil relies on regional pilot projects and public-private partnerships. Additionally, the EU mandates standardised APIs and sustainability metrics across sectors, contrasting with Brazil's sector-specific approaches and slower integration of environmental goals. These disparities highlight Brazil's developmental challenges against the EU's advanced, policy-driven ecosystem.

A4.4 Data access and data sharing measures - brief summary | Canada

Data access and data sharing measures - brief summary Canada	
Note about this fiche	The identified data access and data sharing measures in Canada largely refer to high-level policy initiatives. Thus, it was agreed with DG CNECT to make short summaries of two of the identified initiatives, considered interesting but not suitable for self-standing information fiches. The two selected initiatives are: Canadian Sovereign AI Compute Strategy, and Linkable Open Data Environment (LODE).
Introduction	Canada's current federal data and privacy regulation include three main components: Personal Information Protection and Electronic Documents Act (PIPEDA), which applies to private-sector personal data nationwide, except in provinces with "substantially similar" laws (Alberta, British Columbia and Quebec). Privacy Act, which governs federal government collection/use of personal data. Canada's Anti-Spam Legislation (CASL), which regulates commercial electronic messages (email, SMS). The data regulation in Canada has been under revision for several years. The proposed Digital Charter Implementation Act (Bill C-27 from 2022), would replace the existing federal privacy legislation (PIPEDA) and create the first AI law in Canada. However, due to the minority status of the federal government it remains uncertain if and when the bill will come into force.⁸⁰⁷
Canadian Sovereign AI Compute Strategy	The Canadian Sovereign AI Compute Strategy (from now on "the Strategy") aims to strengthen Canada's position as a global leader in artificial intelligence (AI). It provides €1.27 billion (\$2 billion) investments, over five years, in the Canadian AI ecosystem and infrastructure in order to empower national businesses and researchers, and to drive innovation and competitiveness in this highly strategic area. The Strategy is divided in three lines of investments: 1) Mobilising private sector investment – up to €445 million (\$700 million); 2) Building public supercomputing infrastructure – up to €636 million (\$1 billion); and 3) AI Compute Access Fund – up to €191 million (\$300 million).^{808, 809} For this study the first investment line is considered to be of particular interest. It is implemented through a Call for Proposals under the AI Compute Challenge, which will fund projects that are aiming to establish "fully integrated <u>AI data-centre solutions</u> ready for commercial deployment". Eligible applicants to the AI Compute Challenge will be companies, consortiums, and academic-industry partnerships pursuing commercially viable AI compute infrastructure projects. In particular, the Call for Proposal will give priority to the following objectives:⁸¹⁰ Expand national AI compute capacity: Support the creation or growth of new AI data centres, high-performance computing clusters, and testbeds, while ensuring sensitive data remains secure and stored within Canada. Enhance access and affordability: Offer Canadian businesses and innovators affordable, domestic computing power for large-scale AI tasks, along with access to datasets, advanced technologies, and skilled professionals.

⁸⁰⁷ Osler (December 2024), *Privacy and data-related law reform to continue in 2025*, available at: <https://www.osler.com/en/insights/reports/2024-legal-outlook/privacy/>

⁸⁰⁸ Government of Canada, Innovation, Science and Economic Development (updated May 2025), *Canadian Sovereign AI Compute Strategy*, available at: <https://ised-isde.canada.ca/site/ised/en/canadian-sovereign-ai-compute-strategy>

⁸⁰⁹ Government of Canada, Innovation, Science and Economic Development (December 2024), *Canada to drive billions in investments to build domestic AI compute capacity at home*, available at: <https://www.canada.ca/en/innovation-science-economic-development/news/2024/12/canada-to-drive-billions-in-investments-to-build-domestic-ai-compute-capacity-at-home.html>

⁸¹⁰ Government of Canada, Innovation, Science and Economic Development (updated December 2024), *AI Compute Challenge*, available at: <https://ised-isde.canada.ca/site/ised/en/canadian-sovereign-ai-compute-strategy/ai-compute-challenge>

	<i>Support Canadian AI leadership:</i> Strengthen local AI companies by ensuring their computing needs are met with Canadian-based, regulated infrastructure. <i>Promote innovation and sustainability:</i> Encourage the use of sustainable energy and efficient technologies, and stimulate the local supply chain through the use of Canadian-developed AI hardware and software. Several parallels can be drawn between the Canadian Sovereign AI Compute Strategy and EU's AI Continent Action Plan, the latter focusing on five key areas of which two are: <i>Computing infrastructures</i>, and <i>Access to high-quality data</i>.⁸¹¹ Data centres – in EU linked to AI Factories⁸¹² and AI Giga Factories – play a key role in both of these initiatives, both in the access of quality data and in the expansion of AI compute capacity. There is also a strong link to data spaces, which intend to enable secure data sharing and connections between different sectors. In the EU common European data spaces⁸¹³ are being developed in 14 different sectors/domains: agriculture, cultural heritage, energy, finance, green deal, health, industry (manufacturing), language, media, mobility, public administrations, research and innovation, skills, and tourism. Analysing the future opportunities, EU is already collaborating with Canada on digital transformation. The Canada–EU Digital Partnership⁸¹⁴ was launched in November 2023, as an official framework for cooperation on connectivity, digital identity, data governance and data spaces, AI, semiconductors, quantum, and cyber security. A recent joint report (May 2025) from the Information and Communications Technology Council of Canada (ICTC) and the AI, Data and Robotics Association (Adra) of EU outlines six collaboration areas:⁸¹⁵ <i>AI in business transformation</i> – e.g. AI testbeds, regulatory sandboxes, and best practice sharing. <i>Promoting ethical use of AI</i> – e.g. harmonizing regulation, and aligning on privacy protection. <i>Building AI knowledge and skills</i> – e.g. AI talent exchanges, reskilling and upskilling, and alignment of curricula. <i>Just transitions to an AI-enabled economy</i> – e.g. labour rights and sustainable AI development, and international cooperation on labour standards. <i>AI innovation hubs</i> – e.g. leverage Canada's and the EU's ongoing initiatives for joint research on AI-quantum development, cybersecurity, and AI-driven climate solutions. <i>Public-private partnerships</i> – e.g. developing co-designed AI policies, AI in infrastructure projects, and public-private AI collaboration models.
Linkable Open Data Environment (LODE)	The vast majority of open datasets disseminated by Statistics Canada come in the form of aggregated data. However, the Linkable Open Data Environment is an exploratory initiative aimed at enhancing the use and harmonisation of <u>open microdata</u> from municipal, provincial, and federal sources. It focuses on integrating non-sensitive microdata, such as civic addresses, building locations, and public facility listings, into a unified and accessible framework. The main outcomes are a set of <u>open databases</u> accessible through a single open data license, as well as open source tools used to process the data, and collaborations in an open space. The <u>LODE Viewer</u> (see the figure below) is an interactive map-based tool that supports the discovery and use of open microdata by allowing users to visualise and explore integrated georeferenced datasets.⁸¹⁶ Figure: Linkable Open Data Environment Viewer

⁸¹¹ The three other areas being, Development and adoption of AI, AI skills and talents, and Regulatory compliance and simplification.

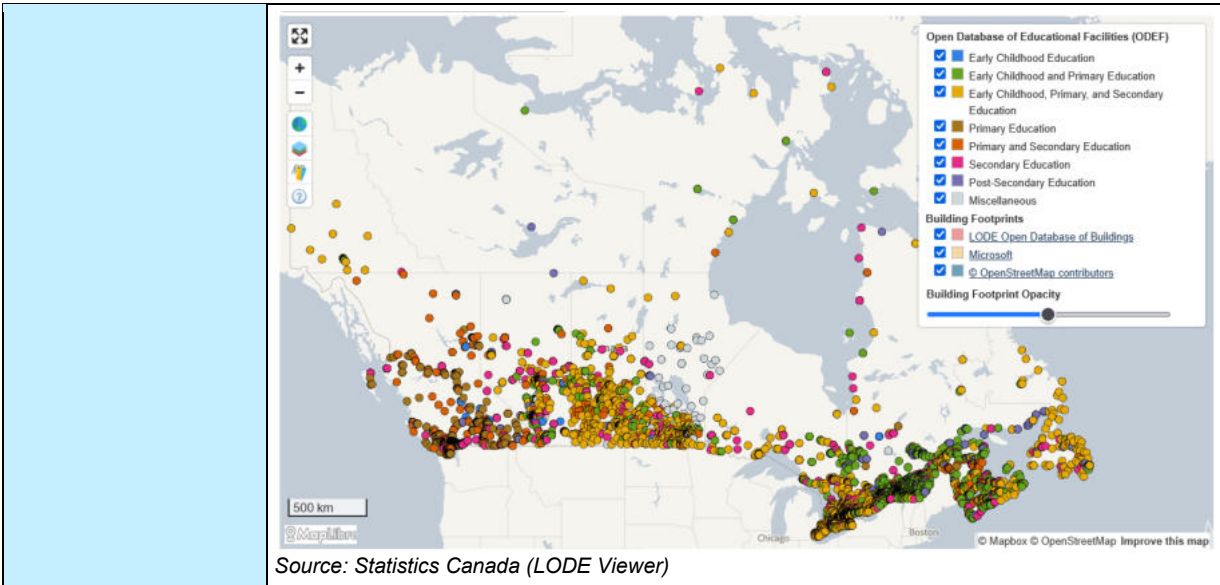
⁸¹² European Commission website, <https://digital-strategy.ec.europa.eu/en/policies/ai-factories>

⁸¹³ European Commission website, <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>

⁸¹⁴ <https://digital-strategy.ec.europa.eu/en/policies/partnerships#:~:text=On%2024%20November%202023%2C%20the,agreed%20to%20work%20together%20on>

⁸¹⁵ ICTC-CTIC (May 2025), *EU-Canada Collaboration on AI Could Lead to Transformational Economic Shifts*, available at: <https://ictc-ctic.ca/news-and-events/news-articles/eu-canada-collaboration-on-ai-could-lead-to-transformational-economic-shifts#:~:text=1,international%20cooperation%20on%20labour%20standards>; Full report at: <https://ictc-ctic.ca/media/1473/download>

⁸¹⁶ Statistics Canada website (updated July 2024), *The Linkable Open Data Environment*, available at: <https://www.statcan.gc.ca/en/lode>



A4.5 Data regulation overview | China

Data regulation overview China	
Introduction	In the People's Republic of China (PRC), personal information protection and data security are governed through a complex legal framework spread across multiple laws and regulations. The core components of this (still evolving) framework are three national laws:⁸¹⁷ the <i>Cybersecurity Law (CSL)</i> (2017),⁸¹⁸ was the first national law to address cybersecurity and data privacy protection; the <i>Data Security Law (DSL)</i> (2021),⁸¹⁹ addresses data security across a broad category of data (not just personal information); and the <i>Personal Information Protection Law (PIPL)</i> (2021),⁸²⁰ the first comprehensive national law on personal information protection – does not replace, but instead enhances and clarifies earlier laws and regulations. They are accompanied by numerous measures and regulations on national, regional and industry level, e.g. The Measures for the Security Assessment of Outbound Data Transfers (effective from September 2022); The Measures for the Standard Contract for the Outbound Transfer of Personal Information, (effective from June 2023); and The Network Data Security Management Regulation (effective from January 2025). From the international dimension perspective, the <i>Provisions on Promoting and Regulating the Cross-Border Flow of Data</i>,⁸²¹ (hereinafter, "Cross-border data transfer regulation"), in force since March 2024, is of particular interest. China has previously established three routes for cross-border transfer, so called "Data Export Safeguard Measures", namely 1) data export security assessments performed by the Cyberspace Administration of China (CAC), 2) filing of Standard Contractual Clauses (SCCs) for personal information exports, and 3) Personal Information Protection Certification by a CAC-accredited organisation. The process is complex and resource-intensive, often requiring legal, technical, and

⁸¹⁷ DLA PIPER (January 2025), *Data protection laws in China*, available at: <https://www.dlapiperdataprotection.com/?c=CN>

⁸¹⁸ Full text (unofficial translation) of the Cybersecurity Law available at: https://www.dataguidance.com/sites/default/files/en_cybersecurity_law_of_the_peoples_republic_of_china_1.pdf

⁸¹⁹ Full text of the Data Security Law available at: https://en.spp.gov.cn/2021-06/10/c_948426.htm

⁸²⁰ Full text of the Personal Information Protection Law available at: https://en.spp.gov.cn/2021-12/29/c_948419.htm

⁸²¹ Full text of the Cross-border data transfer regulation available at: <https://www.chinalawtranslate.com/en/Provisions-on-Promoting-and-Regulating-the-Cross--Border-Flow-of-Data/>

	administrative investments to develop risk mitigation plans, draft legal agreements and implement technical safeguards.⁸²² The Cross-border data transfer regulation introduced various rules aiming to ease the compliance burden on businesses. For instance, these rules feature newly defined exemption scenarios for certain cross-border transfers, revised thresholds for when the Data Export Safeguard Measures apply, and streamlined documentation and filing procedures.⁸²³ China's Free Trade Zones (FTZs) are a strategic tool for the country's economic reform and market opening. A key goal of the FTZs is to attract investment, open up trade and financial sectors, promote innovation by easing regulations and offering government support, and act as a testbed for new regulations and emerging industries. According to the Ministry of Commerce, in 2024, China's 22 FTZs brought in €251 billion (US\$282.5 billion) in foreign investment, representing 24.3% of the nation's total. The FTZs have become experimental grounds for easing cross-border data transfer (CBDT) rules. In this context a positive list (whitelist) refers to specifies types of data that can be freely exported without additional compliance. A negative list, on the other hand, refers to specifies data types and volumes that cannot be freely exported, while any data not on the list can be exported without extra compliance steps.⁸²⁴ The Cross-border data transfer regulation specified that China's FTZs can formulate catalogues with positive lists of general data that companies can freely, or with minimal restrictions, export from the zone. For example, in May 2024, the Shanghai FTZ released a whitelist for the automotive, biopharmaceutical and mutual fund sectors. Several FTZs, in particular Beijing, Shanghai, Tianjin, Zhejiang and Hainan, have taken a distinct approach by issuing data negative lists, either restricting or prohibiting data transfer, including 17 sectors, such as automobiles, pharmaceuticals, retail, civil aviation, reinsurance, the deep-sea industry, and the seed industry. China Briefing presents in a recent article⁸²⁵ key takeaways from an official Questions & Answer document issued by China's Cyberspace Administration (CAC), related to the Cross-border data transfer regulation, adding to the practical interpretation of the regulation. For example, the Q&A document provides practical guidance on security assessments, standard contracts, and certification mechanisms, clarifying when each is required for data exports. It also explains that important data can be exported if a security assessment finds no risk to national security, and outlines scenarios where pre-declared important data may not require assessment.
Geographic scope	People's Republic of China and international dimension (cross-border data transfer and processing). For example, the PIPL applies to personal data processing activities within the PRC. But it also has extra-territorial effect on processing of PRC residents' data outside of PRC:⁸²⁰ for providing products or services to PRC residents; for analytics or evaluation of behaviour of PRC residents; or for any other reasons as required by law or regulations.
Type of measure	Data protection and data governance laws and regulations
Types of data covered	Personal and non-personal. Both public and private sectors.
Relevant sectors	Covers all types of data and sectors.
Objectives	In brief, the objectives of the PRC's three key national data laws are:^{818, 819, 820} CSL: To safeguard network security, maintain national cyberspace sovereignty, and ensure the security of critical information infrastructure and data. DSL: To establish a framework for data classification, promote secure data handling, and protect data relevant to national security, economic development, and the public interest. PIPL: To protect individuals' personal information, regulate its processing, and strengthen individual privacy rights, while ensuring responsible use by organizations.
Restrictions/safe-	This section provides a brief summary of key restriction and safeguards related to the three core laws of

⁸²² Deloitte (June 2024), *The Way Forward: Your Practical Guide to China Cross-border Data Transfer Compliance*, available at: https://www.deloitte.com/cn/en/services/consulting-risk/perspectives/practical-guide-to-china-cbdt-compliance.html?icid=toggle_cn_en

⁸²³ Bird & Bird (January 2025), *China Data Protection and Cybersecurity: Annual Review of 2024 and Outlook for 2025 (I)*, available at: [https://www.twobirds.com/en/insights/2025/china/china-data-protection-and-cybersecurity-annual-review-of-2024-and-outlook-for-2025-\(i\)](https://www.twobirds.com/en/insights/2025/china/china-data-protection-and-cybersecurity-annual-review-of-2024-and-outlook-for-2025-(i))

⁸²⁴ China Briefing (May 2025), *China's Pilot Free Trade Zones – New opinions signal further regulatory easing*, available at: <https://www.china-briefing.com/news/chinas-ftzs-new-opinions-signal-regulatory-easing/>

⁸²⁵ China Briefing (April 2025), *China Clarifies Cross-Border Data Transfer Rules: Key Takeaways from Official Q&A*, available at: <https://www.china-briefing.com/news/china-clarifies-cross-border-data-transfer-rules-official-qa/>

guards	China's data regulation framework. The PIPL⁸²⁰ requires informed consent for data collection and processing, and involves strict rules for sensitive and cross-border data. Large data processors and Critical Information Infrastructure Operators (CIIOs) must store data in China and pass CAC security assessments for data exports. The main safeguards include to grant individuals the right to access, correct and delete their data. Data minimisation and purpose limitation are other means of safeguarding sensitive data. High-risk processing requires designation of data protection officers and impact assessments. The DSL⁸¹⁹ defines data restrictions based on the level of importance according to a mandatory data classification into general, important and core data. Core data must be stored domestically while cross-border data transfers need approval. Foreign data access requests require government authorisation. Secure data handling standards and procedures throughout the data lifecycle, along with obligatory incident reporting and risk assessments, are the most prominent safeguards detected in the DSL. According to CSL,⁸¹⁸ CIIOs must store personal and important data in China. Cross-border transfers of important data is restricted. Regarding safeguards, the law mandates: security assessments for cross-border transfers; obligation to implement technical and organisational security measures; and regular monitoring through security reviews and incident reporting.
Access control mechanisms	The PIPL include several access control mechanisms, such as: purpose and scope control; Consent-based access; Role-based access restriction (both technical measures and organisational rules); and Personal Information Impact Assessments (PIIA) and Audits. The access control means identified for the DSL are: Graded (tiered) access based on data classification; Risk-based access controls (e.g. user authentication, least-privilege principles) aligned to data classification; Incident response and monitoring (mandatory reporting of breaches); and Government oversight (inspections). The CSL handles access control through, e.g. Multi-Level Protection Schemes (network operators system security levels classification); Identity verification and Real-name registration; and Security reviews for cross-border access.
Compliance burden	Foreign companies operating in China face a significant compliance burden under the country's data regulations, particularly regarding cross-border data transfers. Key requirements include:^{821, 825} Data localisation: Critical Information Infrastructure Operators (CIIOs) and companies handling sensitive or large volumes of personal data must store data in China. Security assessments: Transfers of "important data" or large-scale personal data abroad require a security review by the CAC. Standard contracts: Companies with data transfers below regulatory thresholds must use CAC-approved Standard Contractual Clauses and file them within strict timelines. Classification and documentation: Companies must classify data by sensitivity and maintain detailed internal assessments, records, and incident response plans. Frequent updates: Companies must track evolving local rules, such as FTZ-specific negative lists and new exemptions, which vary by region. In short, foreign companies must implement robust data governance frameworks, monitor regulatory changes and maintain legal and operational readiness to meet complex and evolving Chinese data compliance obligations. Note on "important data" The DSL mandates that business data should be categorised based on its importance to national security and public interest. To transfer "important data" abroad, companies must conduct an internal security review and seek approval from the Cyberspace Administration of China (CAC).⁸²⁶ As there is no definitive list published many foreign companies have been struggling with assessing which of their data is considered "important data". The Network Data Security Management Regulations (January 2025), offers a <u>general definition of important data</u> as "<i>data in a specific field, specific group, specific region, or reaching a certain accuracy and scale, which may directly endanger national security, economic operation, social stability, public health and safety once it is tampered with, destroyed, leaked, illegally obtained, or illegally used.</i>" The Q&A on the Cross-border data transfer regulation clarifies that important data can be exported under certain scenarios, where a security assessment verifies that the data transfer does not endanger national security or public interests.^{825, 827}
Enforcement mechanism	Violations to the above-mentioned laws can result in stringent penalties.^{817, 826} PIPL: Fines of up to RMB 50 million or up to 5% of a company's turnover from the previous financial year; Suspension of business operations; and/or Personal liability. DSL: Fines up to RMB 10 million; Criminal liability for endangering national security; and/or Blacklisting of non-compliant entities

⁸²⁶ Ernst & Young (2022), *How China's data privacy and security rules could impact your business*, available at: https://www.ey.com/en_jp/insights/forensic-integrity-services/how-chinas-data-privacy-and-security-rules-could-impact-your-business

⁸²⁷ Brussels Privacy Hub, Yueming Zhang (2024), *Personal Data Protection and Data Transfer Regulation in China*, available at: <https://brusselsprivacyhub.com/wp-content/uploads/2024/04/Personal-Data-Protection-in-China.pdf>

	CSL: Administrative fines; Revocation of licenses; and/or Criminal penalties for serious violations
Effectiveness of enforcement mechanism	China's enforcement framework for data regulation compliance is robust, involving heavy fines, criminal penalties, mandatory security reviews and the power to suspend or revoke business operations. However, the absence of an independent data protection authority and a decentralised enforcement model have been criticised. ⁸²⁷ Also, enforcement actions by the CAC or by regional or sectoral regulators are seldom published, making it difficult to gauge the true scope and effectiveness of the enforcement mechanisms. China focuses more on upfront controls, like requiring security reviews and keeping data stored within the country, rather than relying mainly on penalties after the fact. This approach helps prevent rule-breaking but can also slow down lawful data transfers. ⁸¹⁷
Alignment with international frameworks	PIPL differs significantly from the OECD Privacy Guidelines and other international frameworks like the EU General Data Protection Regulation (GDPR), primarily due to its emphasis on national security and state control over data, which distinguishes it from the more individual-centric and rights-based approach of the GDPR. ⁸²⁸ PIPL, as well as GDPR, adopts the OECD's eight basic privacy principles ⁸²⁹ (collection limitation, purpose specification, use limitation, etc.) as a structural template for consent, data minimisation and impact assessments.
Comparison to EU initiatives	The CSL, DSL and PIPL, as the three main pillars of the Chinese data regulation framework, collectively regulate both personal and non-personal data, with a strong focus on data localization, national security and government control. The Chinese state has broad powers to restrict, monitor and intervene in data transfers, especially those deemed "important" or related to critical information infrastructure. The EU data regulation framework is rights-based, emphasizing individual privacy, transparency and accountability. The EU GDPR and the Data Act focus on empowering individuals and organizations, ensuring data portability and supporting the free flow of data within the EU while protecting fundamental rights.⁸²⁸ While sharing similarities with GDPR, for example in terms of data protection principles, focus on consent and a tier-based system of fines, the PIPL has distinct features, including stricter data localisation requirements and a broader definition of sensitive data. The GDPR allows data processing based on legitimate interests or contractual necessity, provided strict conditions are met. This offers businesses greater flexibility and supports consistent compliance across EU Member States. In contrast, China's PIPL places strong emphasis on explicit consent and central government approvals, which can hinder cross-border operations, especially due to requirements for state-approved infrastructure and oversight.^{817, 828} In China non-personal data considered as "important data" and data linked to CIIOs and relevant to national security are tightly controlled in the DSL and CSL, with strict rules for localisation and export. In EU the Data Act and the Free Flow of Non-Personal Data Regulation put emphasis on free flow and fair access, with exceptions for security. They put no localization requirements, except for justified public security reasons⁸³⁰

A4.6 Action Plan for the Development of Trusted Data Spaces (2024-2028) | China

Action Plan for the Development of Trusted Data Spaces (2024-2028) China	
Summary	The Action Plan for the Development of Trusted Data Spaces (2024-2028) outlines China's national strategy to build a secure, interoperable, and value-driven data ecosystem. It aims to establish over 100 trusted data spaces of five different types, across sectors and borders (see section on "Relevant sectors" below), enabling compliant and efficient data sharing while safeguarding personal and national data security. ⁸³¹ These data spaces will be operated by leading enterprises, consortiums of companies, city governments, and authorised third-party service providers. ^{832, 833}

⁸²⁸ IE International Policy Review, Elena Fernández-Novel Escobar, *How do the European Union's GDPR and China's PIPL regulate cross-border data flows?*, available at: <https://ipr.blogs.ie.edu/2025/01/27/how-do-the-european-unions-gdpr-and-chinas-pipl-regulate-cross-border-data-flows/>

⁸²⁹ OECD (1980), *OECD Privacy Guidelines: Part Two "Basic Principles of National Application"*, OECD LEGAL 0188, available at: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188>

⁸³⁰ Tech GDPR (March 2025), *Data protection digest 16 Feb – 2 Mar 2025: Data Act to strengthen EU digital market, vigilance over US data transfers*, available at: <https://techgdpr.com/blog/data-protection-digest-4032025-data-act-to-strengthen-eu-digital-market-vigilance-over-us-data-transfers/>

⁸³¹ National Data Administration (NDA) (November 2024), *Action Plan for the Development of Trusted Data Spaces (2024-2028)*, available at: https://sjj.hubei.gov.cn/bmdt/sjjyw/gj/202411/t20241125_5426763.shtml

	The plan is guided by a market-oriented approach and aligned with key national policies and data regulations. Apart from legal compliance, it emphasizes technological innovation, and collaborative governance to unlock the full potential of data as a strategic resource.⁸³¹ Pilot work has recently been launched, for which the goal is to develop empirical models that can be replicated and promoted nationwide, enabling large-scale circulation and utilisation of data resources and accumulating operational experience for a national data infrastructure.⁸³³ No information has been found on the expected economic impact of the Action Plan or details on the total number of datasets foreseen to be shared.
Geographic scope	People's Republic of China (PRC) and international dimension (promotion of cross-border data spaces).
Type of measure	National Action Plan, released in November 2024 by the National Data Administration (NDA).
Types of data covered	Personal and non-personal. For example, sensitive manufacturing industry information such as design documents and operational data. ^{831, 832}
Data accessibility	Trusted data spaces aim to facilitate safe, efficient, and rule-based data circulation across sectors, regions and borders. Data in a data space can be of both open and restricted character. Restricted data can only be accessed through members of the data space.⁸³¹ Applicants eligible for the pilot projects include independent enterprise corporations registered in PRC, as well as national innovation platforms, such as national laboratories, medical centres and scientific data centres.⁸³³ The plan specifies that <i>Access verification review</i> of a trusted data space should cover registration, identity audit and identity authentication capabilities. It is advisable to integrate identity verification mechanisms such as public security, taxation and market supervision. In addition, a trusted data space should conduct normative reviews on the access of data resources, data products, data services, etc., for interoperability, security and trustworthiness.^{831, 834}
Relevant sectors	Five types of data spaces are identified in the plan, including also some references to specific sectors and/or thematics:^{831, 833} Enterprises – trusted data space for state-owned enterprises and leading enterprises in coordination with upstream and downstream enterprises. The goal is to provide small and, medium-sized enterprises with accessible data products and services, stimulating innovation with data. Industries – trusted data spaces for key industries; e.g. energy and new energy vehicles; AI-driven research in development of new drugs and new materials; agricultural production and product traceability. Service industry: e.g. financial insurance, trade and logistics, medical health, meteorological services, spatiotemporal information, and carbon footprint management industries. Cities - trusted urban data space focusing on urban planning and construction, transport planning, healthcare, ecosystem protection and key population service guarantees. Individuals - steadily and prudently explore the development of personal trustworthy data space, managed with legal frameworks, e.g. for financial, medical, and insurance data flows. Cross-border – efficient and secure cross-border data flow mechanism, supporting e.g. international scientific research, international trade and international supply chain cooperation.
Objectives	The vision is to build a national, integrated, secure, and value-generating data infrastructure based on a network of trusted data spaces that facilitate safe and efficient data sharing across sectors, regions and borders. The plan specifies three key tasks for the establishment of 100 trusted data spaces by 2028, namely:^{831, 835} Capacity building actions – including trusted control capability; resource interaction capabilities; and value co-creation capability. Cultivation and promotion actions – supporting the establishment of the five types of trusted data spaces outlined above (Enterprises, Industry, Cities, Individuals and Cross-border). Foundation-setting actions – including development of key data space common and technical standards; research in core technologies (e.g. trusted control technology, smart contracts and high-performance computing); development of a common service system reducing the threshold for the

⁸³² The Jamestown Foundation (December 2024), *Trusted Data Spaces': Beijing's Latest Step Toward Digital Dominance*, available at: <https://jamestown.org/program/trusted-data-spaces-beijings-latest-step-toward-digital-dominance/>

⁸³³ Webull (April 2025), *The National Data Administration organizes and carries out pilot work on the innovation and development of trusted data space in 2025*, available at: <https://www.webull.com/news/12604160240698368>

⁸³⁴ Outlook New Era (April 2025), *Promote the construction of a trustworthy data space and accelerate the release of data value*, available at: https://www.lwxsd.com/pcen/info_view.php?tab=mynews&VID=64634

⁸³⁵ China Daily (November 2024), *China to establish over 100 trusted data matrixes by 2028*, available at: <https://www.chinadaily.com.cn/a/202411/23/WS67419042a310f1265a1cf399.html>

	construction and use of trusted data spaces; strengthen the standardized management of data spaces (e.g. compliance management guidelines); and expanding international cooperation on trusted data spaces.
Legal or regulatory provisions	All parties of the trusted data space must abide by the provisions of the Chinese data regulation framework, notably the Cybersecurity Law (CSL), the Data Security Law (DSL), the Personal Information Protection Law (PIPL). These data spaces are subject to oversight by the government or an authorised third party to ensure compliance and accountability. In particular the term <i>Trusted Data Space Supervisors</i> refers to the government authorities or third-party entities authorised to supervise the performance of trusted data space management, and are responsible for guiding, supervising and regulating the activities of trusted data spaces to ensure the compliance of trusted data space operations. ^{831, 833}
Restrictions/safe-guards	The action plan presents five areas of safeguards linked to the success of its implementation. ⁸³¹ Strengthening overall coordination: National and regional authorities are tasked with coordinating across departments to implement trusted data spaces, aligning efforts with local conditions and promoting secure, efficient data usage. Increasing financial support: Increased investment is encouraged through public funds, local government support, and private capital, including venture funding and data-based financial innovation, to back infrastructure, R&D, pilots, and services. Strengthening training of personnel: Educational institutions and enterprises are urged to collaborate on training professionals in areas such as data compliance, analysis, and services. Support for innovation ecosystems, including labs and incubators, is also emphasized. Strengthening the benchmarking lead: The plan calls for the development of model projects and typical applications to guide national efforts. Mechanisms for dynamic evaluation and feedback are to be established to support iterative improvements. Promotion of exchange and cooperation: Industry alliances, standard bodies, and associations will promote collaboration in rule-making, R&D, operations, and market development. Self-regulation and public education are also encouraged to manage risks and foster cross-domain data exchange.
Data governance structure	One of the objectives with the initiative is to establish governance models based on transparent rule-making, rights and responsibility frameworks. As part of the Action Plan China also seeks to influence international standard and data governance development, e.g. by collaborating in multilateral frameworks like the EU, G20, BRICS, and the Shanghai Cooperation Organisation, and actively participating in standard-setting organisations such as the International Organisation for Standardisation (ISO), the International Telecommunication Union (ITU) and the International Union of International Criminal Engineers (IEC). The goal is to promote the adoption of China's technologies, rules, and certification systems for trusted data spaces worldwide. ^{831, 833}
Data localisation requirements	The Action Plan does not explicitly use the term "data localisation", but it clearly supports the principle of controlling and managing data flows within national borders, especially for sensitive data types. The plan also promotes building cross-border trusted data spaces, but only under strict security and privacy conditions. Cross-border data flows must comply with CSL, DSL and PIPL Key compliance mechanisms mentioned related to cross-border data transfer include: ⁸³¹ Support the free trade pilot zone to issue/implement data outbound management list (negative list) (data that cannot leave the country). Cross-border data flow assessment frameworks. Exit management systems for personal and important data. The plan indicates that cross-border data flow mechanisms shall be based on multilateral frameworks, such as the Shanghai Cooperation Organisation, BRICS and Group of 20. ⁸³⁶
Available data infrastructures	There is an ambitious push to build centralised, standardised, and government-aligned infrastructures for data sharing, compliance and value creation on national-level. Envisaged infrastructure elements include: 1) National public basic services platform; 2) Computing power and circulation infrastructure; and 3) Public data space support services. While there is no precise, standalone investment figure for the trusted data spaces themselves, they are considered a significant part of China's broader data infrastructure initiative, with a total investment projected to reach around €246 billion (2 trillion yuan) over the next five years. These investments do not refer only to the trusted data spaces themselves but also target the underlying architecture for data collection, aggregation, transmission, processing, circulation and security. ^{837, 838}

⁸³⁶ BIIA (December 2024), *China aims for more than 100 'trusted data spaces' by 2028 under national action plan*, available at: <https://www.biaa.com/china-aims-for-more-than-100-trusted-data-spaces-by-2028-under-national-action-plan/>

⁸³⁷ ChinaDaily (January 2025), *National data infra design underway*, available at: <https://www.chinadaily.com.cn/a/202501/09/WS677f21d6a310f1265a1d9c94.html>

⁸³⁸ ChinaDaily (May 2025), *Special bonds earmarked for data infrastructure*, available at: <https://www.chinadaily.com.cn/a/202505/07/WS681ac674a310a04af22bddef.html>

Effectiveness of enforcement mechanism	As mentioned above, trusted data spaces must fully comply with national laws, specifically CSL, DSL and PIPL. These laws already carry formal enforcement mechanisms (e.g. fines, audits, license revocations), so trusted data spaces are expected to integrate those requirements.¹ The Action Plan also emphasizes building institutional mechanisms for:⁸³¹ Classification and grading of data. Security reviews. Export control evaluations. Cross-border flow declarations.
Comparison to EU initiatives	China's Action Plan is aimed at creating a national infrastructure for secure, compliant, and value-generating data sharing across sectors. In comparison, EU's Common European Data Spaces initiative,⁸³⁹ aims to create sovereign, interoperable and sector-specific data spaces that foster innovation and EU-wide digital sovereignty. China's Action Plan is state coordinated, following a centralised governance model with top-down design, national-level standards, infrastructures, and auditing platforms. The EU data spaces initiative adapts a federated and decentralised governance approach, in which each sector and member state retains flexibility under a common EU legal and technical framework. In relation to the cross-border stance, China's Action Plan is highly restrictive and cross-border flows must comply with pre-authorised export controls. The EU data spaces initiative encourages cross-border data flows within the EU. Also, data transfers outside the EU are allowed, but under GDPR adequacy or other safeguards. It will be important for the EU to follow the development of the Chinese initiative, particularly when the work starts with the definition and implementation of cross-border data spaces.

A4.7 Japan Act on the Protection of Personal Information (APPI) | Japan

Japan Act on the Protection of Personal Information (APPI) Japan	
Summary	The Act on the Protection of Personal Information is the primary legislation that governs the collection and processing of personal data by both the public and private sectors in Japan.⁸⁴⁰ It equally applies to the transfer of personal data to third countries. As the basic act for the protection of personal information, the APPI establishes the Personal Information Commission (PPC) as a regulatory body in charge of issuing guidance on the application and interpretation of the Act and its requirements. Since its enactment in 2003, the APPI has been amended three times, including the integration of the former Data Protection Law for Public Sector as one of the main changes conducted. The Act is subject to triennial reviews, the latest being completed in 2025. The 2025 review is expected to include exemptions to the consent requirement for AI development.
Geographic scope	Japan
Type of measure	Law. In force since May 2003 - amended in 2015, 2020, 2021 and 2025 (draft stage)
Types of data covered	Personal information, pseudonymised personal information, anonymised personal information, <i>other information</i> serving as reference to a person
Relevant sectors	Public sector including national Government, local governments and administrative entities. Private sector
Objectives	The main objectives of the APPI are: • Protection of the rights and interests of individuals in relation to their personal information. • Ensure a smooth and proper management of the processes or services of administrative entities.

⁸³⁹ European Commission website, *Common European Data Spaces*, available at: <https://digital-strategy.ec.europa.eu/en/policies/data-spaces>

⁸⁴⁰ Kanpō. The Official Gazette (2003). *The Act on the Protection of Personal Information* (Act No. 57, 2003). Available online: <https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en>

The information presented hereafter refers to the latest version 2021 of The Act on the Protection of Personal Information (Act No. 57, 2003) unless otherwise referenced.

	• Creation of a governmental basic policy with regard to the proper handling of personal information. • Clarification of the responsibilities of the national and local governments and establishment of obligations for businesses and administrative entities that handle personal information. • Establishment of the Personal Information Protection Commission (PPC). No explicit reference to FAIR Principles - Findable, Accessible, Interoperable, Reusable -, but close alignment with them.
Legal or regulatory provisions	National Government is to take the necessary measures to ensure the proper handling of personal information held by its organs. Support local governments and businesses in implementing the required measures. Ensure the processing of complaints between businesses and individuals. Local governments are to take the necessary measures to ensure the proper handling of personal information held by its organs. Support local businesses and residents in personal information handling. Provide mediation for complaint processing between businesses and individuals as well as guarantee an appropriate and prompt processing of the complaint. Administrative entities must clearly indicate the purpose of use before directly acquiring personal information (exceptions apply). Administrative entities must guarantee the proper acquisition of personal information, information accuracy, measures for managing the security of personal information and the report of leaks. Businesses handling personal information must specify the purpose of use, not use it in any way fomenting or inducing unlawful acts and guarantee a proper acquisition. Unless the purpose of use has already been disclosed, a business must promptly notify the person of that purpose of use or disclose this to the public once it has acquired personal information. Businesses must guarantee the accuracy of data, proper management of the security of personal data, supervision of employees or entrusted persons handling personal data and report leaks.
Restrictions/safeguards	Personal information cannot be handled beyond the scope necessary for achieving the purpose of use and without obtaining the individual's consent to do so in advance. Personal information must not be provided to a third party without obtaining consent in advance, except specific cases including: cases based on laws and regulations, cases in which there is a need of protection and it is difficult to obtain the consent, cases in which there is a special need to improve public wellbeing or cooperate with a public entity, cases in which an academic research institution and needs to provide the personal data for the academic research purpose. Transfers of personal data to third countries are allowed under specific conditions.⁸⁴¹ Transfers are allowed to third countries deemed to have a personal information protection system recognised to have equivalent standards to that in Japan. Currently, the PPC recognises as such the EU - including EEA member states - and the UK. For those transfers to foreign countries not recognised as having an adequate system, Japanese businesses must obtain the identifiable person's consent. Businesses must provide beforehand information on the personal information protection system of the foreign country and the measures the third party takes for the protection of personal information.⁸⁴² In February 2025 and as part of the triennial review of the APPI, proposed introducing exemptions to the consent requirements under two circumstances: when collecting sensitive personal data; when transferring personal data to third parties, for the creation of statistical information and generative AI development. The proposed amendments are based on the assessment that <i>statistical information excludes specific individual relationships and poses minimal risk to individual rights and interests</i>. In turn, the proposal highlights the requirement to ensure the data is used solely for statistical or AI purposes.⁸⁴³
Access control mechanisms	Data covered under the APPI must be handled following consent requirement, purpose limitation, data transfer controls, restrictions on third-party sharing, proper security measures and breach notifications.

⁸⁴¹ Personal Information Protection Commission (2019). *Supplementary Rules under the Act on the Protection of Personal Information for the Handling of Personal Data Transferred from the EU and the United Kingdom based on an Adequacy Decision*. Available online: https://www.ppc.go.jp/files/pdf/Supplementary_Rules_en.pdf

⁸⁴² Kanpō. The Official Gazette (2003). *The Act on the Protection of Personal Information (Act No. 57, 2003)*. Available online: <https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en>

⁸⁴³ Personal Information Protection Commission (2025). *Personal Information Protection Commission*. Available online: https://www.ppc.go.jp/files/pdf/seidotekikadainitaisurukangaekatanitsuite_r6.pdf

Compliance burden	Since the 2020 and 2021 amendments, the Japanese regulatory environment regarding the protection of personal information has become stricter. In addition to the increase of penalties for violations of the APPI, it is now required to issue notifications of serious data breaches to the PPC. In relation to the transfer of personal information outside Japan, it is now required to provide a considerable amount of information regarding data protection regulations and data processing security measures of the receiving country to the concerned individual. It is equally required to inform individuals about any changes in regulations or security measures. Additionally, if a firm in Japan relies on data handlers or processors overseas, it must audit them regularly to ensure their compliance with Japanese regulation.⁸⁴⁴ On the contrary, the PPC is considering easing certain provisions of the regulation as part of the ongoing triennial review and amendment process of the APPI. The potential changes include relaxing the breach reporting obligations to the PPC for organisations that have obtained prior certification from recognised third-party bodies as well as allowing organisations to use personal data without consent for the purpose of generative AI education. The proposed amendments come at a time when the business community has expressed its concerns about the burden of the APPI's most restrictive provisions in corporate activities. Similarly, the advancement of AI technology is leading to reconsider its balance with the protection of personal information.⁸⁴⁵
Enforcement mechanism	The PPC is the appointed body for the supervision and enforcement of the compliance of the APPI. It is in charge of carrying out reports and on-site inspections, issuing recommendations and orders as well as providing guidance and advice. In the event of a leakage, loss, damage or other situation concerning the security of personal information, any <i>personal information handling business operator</i> shall report it to the PPC when the compromised data has been identified by the PPC as those of which there is a large possibility of harming and individual's rights and interests. The APPI includes penal provisions for individuals of up to two years of imprisonment and up to 1 million YEN, based on the severity of the violation committed. For corporations, fines are set not to exceed 100 million YEN.
Effectiveness of enforcement mechanism	According to the 2024 APPI Interim Report published by the PPC (only available in Japanese), 12,120 cases of data breaches and related incidents were reported by the PPC, of which 96% affected fewer than 1,000 individuals per incident while a 0.5% affected more than 50,000. 84% of the incidents reported affected only a single individual. The most common type of information breached involved sensitive personal information (89.7%), while 86.3% of the incidents were caused by human errors including incorrect deliveries, misdirected mailings, improper disposal or loss.⁸⁴⁶ Additionally, the PPC issued 176 requests for reports, 115 advice, one recommendation, one order and conducted 26 onsite inspections between 2022 and 2023.⁸⁴⁷ By contrast, the imposition of fines was marginal and there is no public information available regarding criminal penalties imposed. The main problems identified by both the 2024 Interim Report and the business community are related to the overreporting of minor incidents, the limited use of administrative fines by the PPC and the lack of clarity in enforcement practices identified by the businesses.⁸⁴⁸
Alignment with international frameworks	The APPI is closely aligned with the main international data protection frameworks including the OECD Privacy Guidelines, the WTO Principles and the EU GDPR, in order to facilitate

⁸⁴⁴ Blair, Gavin (2022). *Stricter policies, safer data. Amendments to Japan's Act on the Protection of Personal Information come into force*. Eurobiz Japan. Available online: <https://eurobiz.jp/feature/stricter-policies-safer-data/?utm>

⁸⁴⁵ Nagae, Shunsuke (2024). *Amendments to the Act on the Protection of Personal Information of Japan Proposed for 2025*. Clifford Chance. Available online: <https://www.cliffordchance.com/insights/resources/blogs/talking-tech/en/articles/2024/06/amendments-to-the-act-on-the-protection-of-personal-information-set-for-2025.html>

⁸⁴⁶ Kawase, Toki. *Unstoppable Personal Information Leaks, a 1.5-Fold Increase in Reiwa 5 (2023) Compared to the Previous Year: Explaining the Latest Trends*. Monolith Law Office. Available online: <https://monolith.law/en/general-corporate/ppc-annual-report-2023?utm>

⁸⁴⁷ Linklaters (2024). *Data Protected - Japan*. Available online: <https://www.linklaters.com/insights/data-protected/data-protected---japan?utm>

⁸⁴⁸ BSA (2024). *The Software Alliance on the Interim Report on the Discussion of the So-called Every-Three-Year Review of the Act on the Protection of Personal Information*. Available online: <https://www.bsa.org/files/policy-filings/en07292024bsareviewppc.pdf>

	cross-border data flows. In particular, the APPI follows the OECD Privacy Guidelines through its limitations of data collection and use, the requirement of data quality and security safeguards as well as the right of individuals to access, correct and delete personal data.⁸⁴⁹ With regard to the WTO Principles, the APPI aims at facilitating cross-border data transfers and avoids unnecessary data localisation requirements by prioritising the implementation of adequate data protection measures both in Japan and abroad.⁸⁵⁰
Comparison to EU initiatives	The APPI has increasingly aligned with EU GDPR through its latest amendments, designed to further facilitate international data flows and cooperation. Despite this, the two regulations maintain significant differences. The GDPR acts as a more comprehensive law in its scope, legal bases and enforcement provisions, with a stronger penalty framework and a broader range of data subject rights.⁸⁵¹ For its part, the APPI continues to prioritise reporting and advice over strong enforcement mechanisms and penalties, although this may be subject to amendment in 2025. The main convergence between both regulations is found in cross-border data transfers. In 2019, Japan and the EU recognised each other as providing an adequate personal information protection framework.⁸⁵² The 2025 proposed amendments could however change the regulatory direction of the Act, if deregulation of the consent requirement is finally introduced for AI purposes. In fact, this introduces a larger debate over the balance between data protection, especially personal information, and the development of new technologies including AI. In this sense, Japan seems to have chosen to subject data protection, under certain conditions, to the development and learning of generative AI models.

A4.8 Japan Data Exchange (JDEX) | Japan

Japan Data Exchange (JDEX) Japan	
Summary	The Japan Data Exchange is a private-sector-led initiative consisting of a membership-based marketplace of over 260 entities for data sharing, collaboration and trading, connecting data providers with acquirers. It provides one-stop support for data registration, search, viewing, marketing, matching, sample acquisition, negotiation and contracts. It is designed to be connected to different public clouds, and it counts with various security measures ⁸⁵³ .
Geographic scope	Japan. Open to international collaboration
Type of measure	Initiative for data sharing - data marketplace. In force since 2020
Types of data covered	Economy, finance and business, real estate, consumer trends, healthcare, vehicle driving and teaching data. Non-personal data.
Data accessibility	Restricted data. Controlled-access data marketplace based on licensing agreements, usage

⁸⁴⁹ OECD Council (2016). *Recommendation of the Council concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data*. Available online: <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0188#backgroundInformation>

⁸⁵⁰ World Trade Organisation (2004). *Understanding the WTO*. Available online: https://www.wto.org/english/thewto_e/whatis_e/tif_e/understanding_e.pdf

⁸⁵¹ European Parliament and the Council (2016). *Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>

⁸⁵² Personal Information Protection Commission (2019). *The framework for mutual and smooth transfer of personal data between Japan and the European Union has come into force*. Available online: <https://www.ppc.go.jp/en/aboutus/roles/international/cooperation/20190123/>

⁸⁵³ Japan Data Exchange JDEX. *Services*. Available online: <https://j-dex.co.jp/>

The information presented hereafter refers to the official Japan Data Exchange JDEX Services webpage unless otherwise referenced.

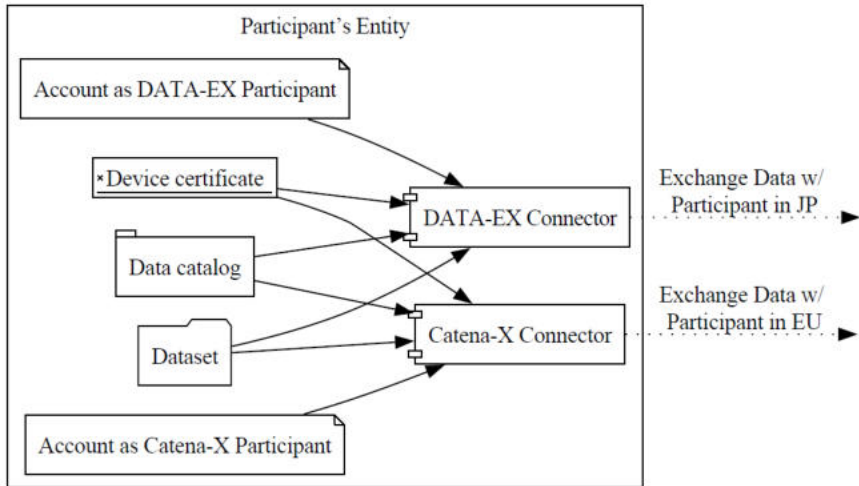
	restrictions and compliance obligations between data providers and acquirers.
Relevant sectors	Public sector, academia, finance sector, real estate sector, healthcare sector, automotive sector, education sector.
Objectives	The main objectives of JDEX are: Cross-industry and cross-border data collaboration and trade. Establish a member-based marketplace for data sharing, collaboration and trading. Conduct ideathons to generate ideas regarding the usage of JDEX customer's data. Provide comprehensive support for data business focused on finding insights for an optimisation of the work. Support of metadata management through data catalogues. Allow customers to set up private trading venues within JDEX and conduct transactions with specific parties. Provide support from marketing to inside sales. No explicit reference to FAIR Principles - Findable, Accessible, Interoperable, Reusable -, but close alignment with them.
Restrictions/safeguards	Data providers are enabled to retain control over data through various security and control settings for trusted data transactions as well as to conduct data exchanges in two ways, either within the JDEX Data Marketplace or within "private spaces" on JDEX. This ensures the possibility to share and exchange data with a selected group of organisations, in isolation of the rest of the participants community. ⁸⁵⁴ Additionally, JDEX complies with local and international data regulations and conforms to Trusted Data Transaction - the future European harmonised standard. ⁸⁵⁵
Data governance structure	Decentralised data governance structure based on data sovereignty , which allows data providers to deploy any data exchange use case and configure the terms of use and service for each transaction. This model guarantees continuous control over the access and use of data exchanged within JDEX. ⁸⁵⁶
Data localization requirements	No specific data localisation requirements. Decentralised platform based on data sovereignty, which entails self-determination for data providers on the terms of use and service of shared data as well as its control. Data providers must comply with the national and international data protection regulations, based on the type of data and origin.
Available data infrastructures	Data infrastructure provided by Data Exchange Solution from Dawex based on⁸⁵⁷: Standard data transfer mechanisms for technical interoperability, including APIs, Data Transfer Agents (DTA) or Eclipse Data Space Components Connectors (EDC connectors). Standards data formats to ensure syntactic interoperability. Core APIs based on Open API standards. Extensive data connectors library for data storage, data modelling and data integration solutions. Support of industrial standards and tools such as OPC-UA or Asset Administration Shell. Standard protocols and frameworks for key concepts such as identity, role and permissions management. Documentation portal and service accounts with fine-grained privilege management. The data infrastructure ensures technical, syntactic and semantic interoperability to connect

⁸⁵⁴ Dawex (2024). Case Study: Kanematsu empowers Japan's data economy with Japan Data Exchange. Available online: https://www.dawex.com/hubfs/1-Marketing-Content-Assets/Dawex-Case-Study-JDEX-Data-Marketplace_EN.pdf?hsCtaTracking=b135e288-ffab-486c-9873-9d0f4c3c7e58%7C71aa1b42-e979-414f-851b-052918c66844

⁸⁵⁵ Dawex. *Data Exchange Solution*. Available online: <https://www.dawex.com/en/data-exchange-solution/>

⁸⁵⁶ Ibid.

⁸⁵⁷ Ibid.

	data providers' and data acquirers' IT systems and business applications to the data space.
Effectiveness of enforcement mechanism	Architectural and procedural enforcement mechanisms to ensure compliance and data sovereignty through security and control settings, certification systems as well as configurable terms of use and service. Due to the private nature of JDEX, information regarding the effectiveness of compliance and enforcement is currently not available publicly.
Comparison to EU initiatives	Both JDEX and EU initiatives such as Catena-X, GAIA-X or IDSA aim to enable reliable data exchanges based on trust frameworks, interoperability between sectors and platforms as well as cross-border data flows. Yet both models differ from each other in key aspects regarding governance, technology standards, infrastructure and authentication frameworks⁸⁵⁸. JDEX operates as a private-driven commercial data marketplace, while initiatives such as GAIA-X and IDSA are led by public-private consortiums with a relevant involvement of the European Commission. In the same line, while JDEX is focused on the development of international data flows following the Data Free Flow with Trust (DFFT) concept, its European counterparts have been shaped by an emphasis on digital sovereignty and strong data localisation requirements. Potential Interoperable architecture between DATA-EX and Catena-X:  Source: Towards Interoperable Data Spaces: Comparative Analysis of Data Space Implementation between Japan and Europe

A4.9 Ouranos Ecosystem | Japan

Ouranos Ecosystem Japan	
Summary	The Ouranos Ecosystem is a series of initiatives launched by the Ministry of Economy, Trade and Industry (METI) in collaboration with the Information-technology Promotion Agency (IPA), the Digital Architecture Design Center (DADC) and the New Energy and Industrial Technology Development Organisation (NEDO). ⁸⁵⁹ It is focused on the design, research, development and promotion of an architecture that enables collaboration and interoperability between various information processing systems operated by different stakeholders. To this end, the Ecosystem has been the result of public-private cooperation between industry, academia and government.
Geographic scope	Japan. Open to international collaboration
Type of measure	National initiative for data sharing. In force since April 2023 - to be fully implemented

⁸⁵⁸ Ishihara, Shun and Taka Matsutsuka (2025). *Towards Interoperable Data Spaces: Comparative Analysis of Data Space Implementations between Japan and Europe*. ArXiv. Available online: <https://arxiv.org/html/2501.15738v1#S5.F5>

⁸⁵⁹ Ministry of Economy, Trade and Industry (2023). *Overview of the Ouranos Ecosystem*. Available online: https://www.meti.go.jp/policy/mono_info_service/digital_architecture/ouranos.html

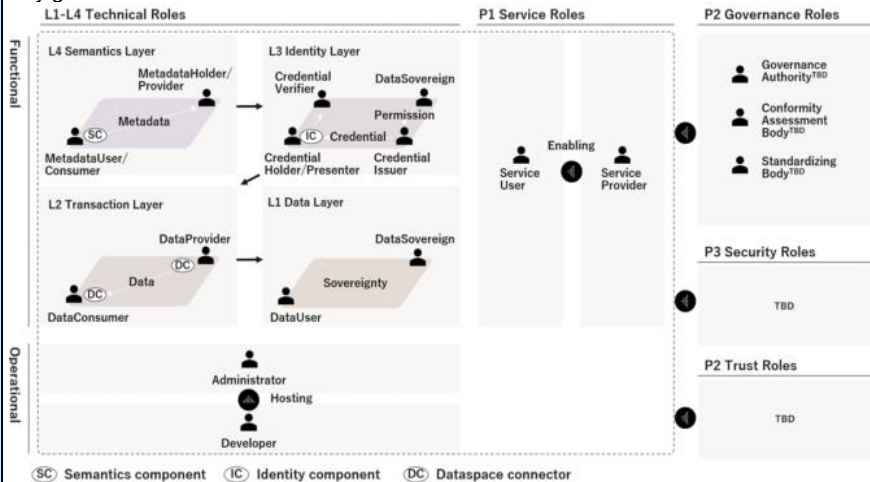
Types of data covered	Product Carbon Footprint (PCF), supply chain, semantic models, cross-border trade, transportation and logistics, energy and decarbonisation. Non-personal data. ⁸⁶⁰
Data accessibility	Not open data. Accessibility based on trustworthiness, secure access, authenticity and identity verification. ⁸⁶¹
Relevant sectors	Public sector, academia, automotive sector, railway sector, chemical sector, technological sector, financial sector. ⁸⁶²
Objectives	Cross-industry and cross-border data collaboration. Development of data exchange infrastructures including 4D Spatiotemporal Data Exchange Infrastructure and Supply Chain Data Exchange Infrastructure. Promotion of digital transformation (DX) in the fields of passenger and freight transport as well as commercial distribution and money flow. Establishment of a Public Interests Digital Platform. International interoperability and standardisation. No explicit reference to FAIR Principles - Findable, Accessible, Interoperable, Reusable -, but close alignment with them.⁸⁶¹
Legal or regulatory provisions	Ouranos Ecosystem Dataspaces Reference Architecture Model (ODS-RAM) Data Free Flow with Trust (DFFT) concept EU Battery Regulation - Regulation (EU) 2023/1542. Through linkage between Japan Ouranos Ecosystem and EU Catena-X.⁸⁶³
Restrictions/safeguards	Data usage. Usage control governs the data based on contract conditions encoded in machine-readable policies. Data providers can impose purpose-based, temporal and actor-specific restrictions through usage policies and contractual controls.⁸⁶¹ Sovereignty Protocol: It defines the specification for self-determination and its guarantee regarding the conditions of storage and use that shall apply when granting permission for data use. Identity and trust framework. Users must go through an onboarding process via identity providers and trust anchors.⁸⁶¹ Identity Layer (L3): It resolves the issue of authentication and authorisation. L3 requires the exchange of verifiable credentials and cross-sectional data mobility with the necessary level of confidentiality. Additionally, it ensures the authentication of participants in dataspace, including level assessment, thereby providing flexible control based on storage and usage conditions. Identity and Trust Protocol: It defines the specifications for the identity, authentication and authorisation of dataspace participants. Conformity assessment. Users must comply with policies and may require an official certification to join or operate the Ecosystem.⁸⁶¹ Certification of Interoperable Data Infrastructure Management Entity (CIDIME): It covers the entity that operates and manages systems for interoperating data across multiple stakeholders and meets the established criteria, thereby externally guaranteeing a certain level of public interest of the entity. Traceability. All data transactions are logged and auditable in order to oversee, manage, detect anomalies and optimise operations of various activities.⁸⁶¹ Contractual governance and data sovereignty. Sovereignty is constantly retained through embedded policies and technical controls. It includes usage control, identity verification and revocability - key for managing consent in data sharing.⁸⁶¹ Heuristic Contracting Protocol: It defines the specification for service contracts that allow newly discovered service users and service providers to electronically enter into new contracts.
Data governance structure	Decentralised data governance structure based on data sovereignty , which enables participating data spaces to maintain their own rules and governance while ensuring compatibility, and data

⁸⁶⁰ Ministry of Economy, Trade and Industry (2025). *The Ouranos Ecosystem Trust Study Group Report*. Available online: https://www.meti.go.jp/english/press/2025/0328_004.html

⁸⁶¹ Ministry of Economy, Trade and Industry, Digital Architecture Design Center, Information-technology Promotion Agency (2025). *Whitepaper: Ouranos Ecosystem Dataspaces Reference Architecture Model*. Available online: <https://www.ipa.go.jp/en/digital/architecture-guidelines/individual-link/d662m3000000hhu8-att/ouranos-ecosystem-dataspaces-ram-white-paper.pdf>

⁸⁶² Ministry of Economy, Trade and Industry (2023). *Japan's Initiatives for Interoperable Data Infrastructure Officially Named "Ouranos Ecosystem"*. Available online: https://www.meti.go.jp/english/press/2023/0429_001.html

⁸⁶³ Catena-X (2025). *Catena-X and Ouranos Ecosystem Successfully Demonstrate Data Space Interoperability*. Available online: <https://catena-x.net/news/catena-x-and-ouranos-ecosystem-successfully-demonstrate-data-space-interoperability/>

	owners to retain full control over the use of data. Key governance roles under the ODS-RAM:  Source: Whitepaper: Ouranos Ecosystem Dataspaces Reference Architecture Model⁸⁶¹ Definition of common rules and policies to ensure that actors interacting in a dataspace follow consistent principles as well as to manage, supervise and operate across the ecosystem. Verification of compliance with governance policies including trade secret protection, personal information protection and public interest compliance through the Certification of Interoperable Data Infrastructure Management Entity (CIDIME). This certification is provided by the Japanese Government. Establishment of a common identifying rule across dataspaces through the Unified Meta Identifier (UMI), which acts as a global identifier across dataspaces governed by a shared protocol. Enforcement of governance through the definition of machine-readable contracts and usage control technologies. This allows users to agree on data usage contracts and dynamic access policies.
Data localization requirements	No specific data localisation requirements. Decentralised ecosystem with data sovereignty, which entails self-determination for data providers regarding the conditions of storage and use that shall apply when granting permission for data use.⁸⁶⁰ Data providers are allowed to enforce usage control policies regardless of the storing or processing of the data. Thus, the Ecosystem emphasises governance, trustworthiness, traceability and transparency over localisation requirements.
Available data infrastructures	Two specific data infrastructures to support secure and interoperable data sharing across sectors. 4D Spatiotemporal Data Exchange Infrastructure, focused on mobility, urban air mobility (UAM), infrastructure and disaster management through geospatial and temporal data. Supply Chain Data Exchange Infrastructure, designed for goods traceability, carbon footprint tracking, lifecycle assessments and automated contracting and payments.⁸⁶¹ Both infrastructures rely on sovereign data stores, metadata registries and federated identity services, with technical specifications and reference implementations included in the ODS-RAM Whitepaper.
Effectiveness of enforcement mechanism	Architectural and procedural enforcement mechanisms to ensure compliance through technical enforcement, certification systems and usage policy enforcement frameworks. Due to the early implementation and testing phase of the Ouranos Ecosystem, information regarding compliance and enforcement has not yet been released.
Comparison to EU initiatives	The Ouranos Ecosystem purposely mirrors EU data sharing initiatives such as GAIA-X, IDSA and in particular Catena-X, not only in its technical and regulatory design, but also to promote potential cooperation opportunities.⁸⁶⁴ In this sense, the Ouranos Ecosystem shares with the previous initiatives core principles including data sovereignty, federated identity, usage policy enforcement and open-source architecture. Despite this, the former is more government-led - with the CIDIME certification and technical enforcement mechanisms - while EU initiatives are consortium-driven and are characterised by voluntary compliance frameworks. In March 2025, a Proof of Concept (PoC) led by the IPA and Catena-X proved the interoperability between two independent data spaces. In particular, this first use case focused on Product Carbon Footprint (PCF) data exchange for batteries.⁸⁶³

⁸⁶⁴ Hayashi, Eiki (2024). Japan, Europe to link industrial data platforms for EV battery materials. Nikkei. Available online: <https://asia.nikkei.com/Spotlight/Supply-Chain/Japan-Europe-to-link-industrial-data-platforms-for-EV-battery-materials>

A4.10 MyData Roadmap | South Korea

MyData Roadmap South Korea	
Summary	South Korea's MyData Roadmap aims to establish a user-centric data ecosystem by expanding personal data control and portability across sectors. In 2024 legal amendments and pilot projects to lay the groundwork were launched. MyData involves 44.5 million public financial records and targets integration with non-financial datasets such as health and legal data. ⁸⁶⁵ More than 140 financial institutions are participating in MyData 2.0 with an ambitious plan to reach 10 million users by the end of 2025. ⁸⁶⁶ According to different sources, MyData initiative already generated economic value and the impact is expected to increase with the projected growth. MyData service providers generated over €1.6 billion (₩2.5246 trillion) in total revenue which includes €52 million (₩72 billion) from credit information integration services. ⁸⁶⁷ MyData 2.0 estimated the reduction of fintech development costs by 30% through standardised APIs and lowering barrier to market entry. ⁸⁶⁸ Overall, the initiative has been successful as it represents a transformative step toward a data-driven economy, with implementation scaling from financial pilots in 2024 to nationwide deployment by the end of 2025. The roadmap aims to expand MyData services beyond finance into healthcare, public services, and logistics, covering 8 sectors (medical, finance, public, distribution/logistics, culture, communication, education, and transportation). ⁸⁶⁹
Geographic scope	South Korea
Type of measure	MyData is a legally backed policy framework combining amendments to existing laws, new regulations, and sector-specific guidelines. Core legal framework: Personal Information Protection Act (PIPA).
Types of data covered	MyData Roadmap covers a wide range of personal data types across multiple sectors: financial data, healthcare data, public administration data, telecommunications data, and energy data. MyData supports structured data formats, data types and APIs for interoperability. ⁸⁷⁰
Relevant sectors	Healthcare, finance, communications, and retail
Objectives	Empower individuals to manage and monetise ⁸⁷¹ their data securely. Promote data portability rights and ethical AI integration. Boost the national data economy, targeting a 50 trillion KRW market by 2027.
Legal or regulatory provisions	MyData regulatory framework is fully operational, with core laws, such as PIPA and sectoral guidelines, in force. Affected parties must comply with strict licensing, technical standards, and cross-border data rules. Enforcement is robust in finance and healthcare; however, the challenges remain in harmonising regulations to support SMEs.
Restrictions/safe-guards	South Korea's MyData Roadmap imposes restrictions and implements safeguards to ensure ethical, secure, and user-controlled data sharing. MyData services are initially restricted to licensed

⁸⁶⁵ Regulations Asia (2024), FSC Korea Unveils Plans to Enhance 'MyData' Service. Available online: <https://www.regulationasia.com/fsc-korea-unveils-plans-to-enhance-mydata-service/>

⁸⁶⁶ Development Asia (2024), Transforming the Republic of Korea's Banking Industry Through MyData 2.0. Available online: <https://development.asia/insight/transforming-republic-koreas-banking-industry-through-mydata-20>

⁸⁶⁷ Government Official website (2024), What experience can we draw on in the development and utilization of personal data in South Korea? Available online: https://www.szzg.gov.cn/2024/szzg/xyzx/202410/t20241012_4906869.htm

⁸⁶⁸ Insurance Asia, (2024), FSC announces enhanced MyData 2.0 in financial services. Available online: <https://insuranceasia.com/insurance/news/fsc-announces-enhanced-mydata-20-in-financial-services>

⁸⁶⁹ W. Choi, et. Al. (2021), Development of a MyData Platform Based on the Personal Health Record Data Sharing System in Korea. Available online: <https://www.mdpi.com/2076-3417/11/17/8208>

⁸⁷⁰ Coalition (2020), *MyData Ecosystem*. Available online: <https://coalition.org/business/resources/mydata-ecosystem/#:~:text=MyData%20seeks%20to%20empower%20individuals%20and%20organisations,has%20supported%20work%20to%20further%20describe%20MyData>

⁸⁷¹ The monetisation of MyData initiative operates through structured mechanisms involving data transfer fees, user-centric revenue models and services fees (subscription and APIs licencing). SMEs receive 50% fee reductions to lower entry barriers.

	operators in finance, healthcare, and public sectors, with gradual expansion to SMEs and energy ⁸⁷² . Moreover, MyData operators must use standardised APIs for data sharing, blocking external third parties from accessing raw datasets. There are strict approval requirements from the Personal Information Protection Commission (PIPC) for transferring personal data overseas, with mandatory anonymisation or pseudonymisation. ⁸⁷³ Critical personal data such as healthcare and finance must be stored domestically unless approved for offshore processing under strict conditions. Furthermore, MyData services used risk assessments for AI systems to prevent bias, discrimination or misuse. ⁸⁷⁴
Access control mechanisms	South Korea's MyData Roadmap employs advanced techniques for user authentication and authorisation to ensure secure, user-controlled data sharing. MyData has standardised API-based Authentication ⁸⁷⁵ and Multi-Factor Authentication ⁸⁷⁶ . APIs are categorised into: identity verification APIs, blockchain-based decentralised identity, and QR code/OTAC (One-Time Authentication Code). ⁸⁷⁷ Moreover, MyData uses OAuth 2.0 Framework, which securely delegates access rights between MyData operators such as banks or healthcare providers and third-party services without exposing user credentials. It is worth to mention, that only government-authorised MyData operators can handle authentication and data transmission, ensuring compliance with technical and ethical standards ⁸⁷⁸ .
Data anonymisation techniques	South Korea's MyData Roadmap employs several anonymisation techniques to balance data utility and privacy, guided by regulatory frameworks like the Personal Information Protection Act (PIPA) and sector-specific guidelines. ⁸⁷⁹ For pseudonymisation, MyData replaces direct identifiers such as names, resident registration numbers with codes, ensuring data remains traceable only with additional secure information. ⁸⁸⁰ Two techniques used for pseudonymisation are partial hiding of identifiers, e.g. by replacing digits in phone numbers with asterisks, and substituting sensitive data with non-sensitive equivalents by using random strings for medical record IDs. Moreover, for finance sector MyData operators must use Financial Services Commission (FSC)-certified pseudonymisation for credit histories and transaction data. ⁸⁸¹ Whereas for the health sector, MyData uses blockchain-based consent logs track anonymised data, ensuring compliance with the Medical Service Act. For anonymisation, MyData is irreversibly removing all identifiers so data cannot be linked to individuals by ensuring each record in a dataset is indistinguishable from at least k-1 others (e.g. grouping geographic data into broader regions). ⁸⁸²
Enforcement mechanism	South Korea's MyData Roadmap is enforced through a structured regulatory framework involving multiple bodies, inspections, audits, and penalties. The primary regulator is PIPC which oversees compliance of PIPA. There is sectorial regulators such as Financial Services Commission and Ministry of Health and Welfare. Moreover, there is Audit Borad who conducts independent audits of public institutions' data practices. MyData operators such as banks and healthcare providers

⁸⁷² Kim and Chang (2023), *Personal Information Protection Commission Announces its Plans for 2023*. Available online: https://www.kimchang.com/en/insights/detail.kc?sch_section=4&idx=26646

⁸⁷³ Peter Oladimeji (2023), *South Korea data protection law (PIPA): Everything you need to know*. Available online: <https://www.didomi.io/blog/south-korea-pipa-everything-you-need-to-know>

⁸⁷⁴ MLex (2025), *South Korea's privacy regulator unveils 2025 policy roadmap for AI era*. Available online: <https://www.mlex.com/mlex/data-privacy-security/articles/2282936/south-korea-s-privacy-regulator-unveils-2025-policy-roadmap-for-ai-era>

⁸⁷⁵ SSenStone, *Milipass's MyData Authentication service*. Available Online: <https://blog.ssenstone.com/en/blog/case3>

⁸⁷⁶ Financial SecurityInsitute (2024) *MyData*. Available online: <https://www.fsec.or.kr/bbs/522>

⁸⁷⁷ Woshipm (2025), *My data, my decision: South Korea's "right to personal information portability" - MyData model research and reference*. Available online: <https://www.woshipm.com/evaluating/5699743.html>

⁸⁷⁸ J. Hgwang, et. al. (2024), *A Comprehensive Analysis of Security Measures for MyData in South Korea Based on AHP*. Available online: https://www.researchgate.net/publication/378919573_A_Comprehensive_Analysis_of_Security_Measures_for_MyData_in_South_Korea_Based_on_AHP

⁸⁷⁹ H. Ko and S. Park (2020), *How to de-identify personal data in South Korea: an evolutionally tale*. Available online: <https://www.privacysecurityacademy.com/wp-content/uploads/2021/06/SSRN-id3665568-How-to-deid.pdf>

⁸⁸⁰ Shim and Kim (2021), *Data Pseudonymization in a Range That Does Not Affect Data Quality: Correlation with the Degree of Participation of Clinicians*. Available online: <https://jkms.org/DOIx.php?id=10.3346/jkms.2021.36.e299>

⁸⁸¹ *ibid*

⁸⁸² A. Richman (2025), *Data Anonymization Techniques: 12 Keys to Compliance*. Available online: <https://www.k2view.com/blog/data-anonymization-techniques/>

	undergo annual audits by the PIPC and sectoral regulators to verify compliance with technical standards (e.g. encryption, anonymisation) and ethical AI use. ⁸⁸³
Effectiveness of enforcement mechanism	There were some cases of penalties in 2023 and 2024, like the 2023 financial sector penalties and. In 2023 PIPC received 120 complaints related to MyData breaches, primarily involving unauthorised data sharing and inadequate anonymisation. ⁸⁸⁴ Moreover, in 2024 there was reported of healthcare breach. According to the stakeholder survey, businesses find compliance costly, while users are concerned about privacy despite the benefits. Moreover, it was reported that SMEs struggle with the technical requirements as well as SMEs lack resources to implement FSC/PIPC-mandated APIs and anonymisation tools which leads to slower MyData adoption in non-financial. ⁸⁸⁵
Comparison to EU initiatives	South Korea's MyData Roadmap is sector-driven and monetisation-focused, contrasting with the EU's horizontal approach. For instance, MyData has sector-specific rollouts as it began with finance sector (2021), expanded to healthcare (2023) and this year it is plan to involve logistics sector. Meanwhile, Europe follows cross-sectorial integration, aiming for interoperability between different domains such as mobility with smart cities. The EU's main data initiatives are the General Data Protection Regulation (GDPR) for privacy and the Data Governance Act (DGA) for data sharing, whereas South Korea's MyData is about user-centric data control across sectors. MyData emphasises monetisation, which the EU does not and focus mostly on data control. The EU's right to explanation under GDPR for automated decisions is not as prominent as in South Korea. The EU uses decentralised bodies like national Data Protection Authorities, while South Korea has a centralised PIPC. Furthermore, licensing in Korea is strict for MyData operators, whereas the EU relies more on general GDPR compliance. MyData governance model is managed by the Korea Fair Trade Commission (KTFC) and the PIPC, while EU data spaces have decentralised governance model, which is more sector-specific data spaces operating under frameworks like GAIA-X. Another difference between MyData and the EU data spaces and DGA, is that MyData targets only personal data such as bank transactions and health records, whereas EU data spaces have broader type of data coverage, covering both personal and non-personal data. Technical infrastructure of MyData includes standardised APIs, mandating unified APIs for secure data sharing between institutions. EU data spaces focus much more on interoperability through a collaborative framework and infrastructure designed to ensure that data owners retain control over their data, even when it is shared. Lastly, anonymisation techniques are similar, but Korea uses tokenisation and sector-specific standards whereas the EU focuses more on pseudonymisation under GDPR. One of MyData's success relies on balancing anonymisation and data usability, which health data space in Europe could take as a key lesson. Moreover, comparing to DGA and MyData another key takeaway is to explore monetisation model of MyData, for instance Europe could explore tokenisation rewards under the DGA for people sharing data.

A4.11 Open APIs | South Korea

Open APIs South Korea	
Summary	South Korea's Open APIs ⁸⁸⁶ initiative is a government-led strategy to foster digital innovation, interoperability, and data-driven growth by standardising and opening access to public and private sector data via secure APIs. ⁸⁸⁷ This initiative was launched as part of broader digital economy policies such as Digital New Deal (2020) ⁸⁸⁸ and MyData Framework (please refer to the MyData fiche) and it covers sectors such as finance, healthcare, transportation, and

⁸⁸³ Korean Legislation Research Institute (2020), *Enforcement decree of the personal information protection act*. Available online: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=54521&lang=ENG

⁸⁸⁴ Chambers and Partners (2025), *Recent Developments in South Korean TMT Regulations: Data Privacy, AI and Platform Regulations*. Available online: <https://practiceguides.chambers.com/practice-guides/tmt-2025/south-korea/trends-and-developments/O20073>

⁸⁸⁵ Korean Legislation Research Institute (2020), *Enforcement decree of the personal information protection act*. Available online: https://elaw.klri.re.kr/eng_service/lawView.do?hseq=54521&lang=ENG

⁸⁸⁶ API = Application Programming Interface

⁸⁸⁷ Data Guidance, *Jurisdictions South Korea*. Available online: <https://www.dataguidance.com/jurisdictions/south-korea>

⁸⁸⁸ Ministry of Science and ICT (2020), *The Digital New Deal Is to Lead Digital Transition in the World After COVID-19*. Available online: <https://www.msit.go.kr/eng/bbs/view.do?sCode=eng&mId=4&mPid=2&pageIndex=&bbsSeqNo=42&nttSeqNo=443&searchOpt=&searchTxt=#:~:text=Korea%20aims%20to%20lead%20digital,903%2C000%20new%20jobs%20by%202025>

	public services. The Open Public Data Directive (OPDD) mandates public agencies to share datasets (e.g. land use, and environmental assessments) through APIs. ⁸⁸⁹ The initiative is proven to be successful through its economic, technological and efficiency impact. In terms of economic impact, the initiative on Open Banking APIs spurred new business models such as neobanks and budgeting apps, and contributed to over €1 billion (₩1.2 trillion) fintech revenue stream by 2025. ⁸⁹⁰ On the technical impact, Open APIs had an impact on the accessibility to open government datasets, and APIs accelerated AI adoption in healthcare, logistics, and environmental management. Lastly, it has been reported that automated API-driven processes reduced manual data handling in public agencies by 30%. ⁸⁹¹
Geographic scope	South Korea
Type of measure	Digitalisation policy initiative on open APIs.
Types of data covered	Personal and non-personal.
Data accessibility	APIs
Relevant sectors	Finance, healthcare, transportation, and public services.
Objectives	The objectives ⁸⁹²⁸⁹³ of the initiative are: To enable startups and businesses to build new services by democratising access to high-quality data. To enhance interoperability by having standardised APIs across sectors to ensure seamless integration between systems. To enable personalised services through MyData giving individuals control over their data portability. To strengthen economic competitiveness in South Korea. The ambition is to unlock value from siloed datasets as well as to foster collaboration between traditional industries and tech companies. To support smart infrastructure. Open APIs support smart city projects, for instance real-time traffic APIs in Seoul. To foster Public-Private Collaboration by developing APIs standards through partnerships between government, industry and academia to align with international frameworks such as GAIA-X.
Restrictions/safe-guards	Personal data ownership resides with individuals through MyData and organisations cannot claim ownership of user-generated data. Furthermore, public sector data is owned by the government but is made accessible through APIs under specific licensing terms. Another restriction is that API providers retain IP rights over their API designs and infrastructure. For safeguarding IPR, public APIs use Creative Commons licenses, while private sector APIs enforce proprietary terms. Moreover, IP clauses in contracts are included, especially for private API agreements defining permitted use cases.
Data governance structure	The governance structure of South Korea's Open APIs initiative is a multi-layered framework involving government bodies, industries, and technical standards to ensure secure and interoperable data sharing. The Ministry of Science and ICT (MSIT) is responsible for overall policy design, coordination and funding for the Open APIs ecosystem. Furthermore, MSIT coordinates with sector specific ministries, such as Health and Finance, to align API policies and priorities. The Financial Service Commission is in charge of regulating Open Banking APIs by enforcing authentication standards and accrediting third party fintech providers. The Personal Information Protection Commission (PIPC) ensures compliance with the Personal

⁸⁸⁹ K. Junk and H. Park, *A semantic (TRIZ) network analysis of South Korea's "Open Public Data" policy*. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S0740624X1500057X>

⁸⁹⁰ S. Oh, et al. (2024), *New Sustainable Fintech Business Models Created by Open Application Programming Interface Technology: A Case Study of Korea's Open Banking Application Programming Interface Platform*. Available online: https://www.researchgate.net/publication/383435047_New_Sustainable_Fintech_Business_Models_Created_by_Open_Application_Programming_Interface_Technology_A_Case_Study_of_Korea's_Open_Banking_Application_Programming_Interface_Platform

⁸⁹¹ K. Junk and H. Park, *A semantic (TRIZ) network analysis of South Korea's "Open Public Data" policy*. Available online: <https://www.sciencedirect.com/science/article/abs/pii/S0740624X1500057X>

⁸⁹² GovInsider (2024), *South Korea leverages open government data for AI development*. Available online: <https://govinsider.asia/intl-en/article/south-korea-leverages-open-government-data-for-ai-development>

⁸⁹³ Seoul Metropolitan Government (2023), *Real-time urban data service for 113 major sites in Seoul*. Available online: <https://english.seoul.go.kr/real-time-urban-data-service-for-113-major-sites-in-seoul/>

	Information Protection Act (PIPA) and mandates encryption, anonymisation, and breach reporting for APIs handling personal data.⁸⁹⁴ The Ministry of the Interior and Safety (MOIS) is also involved in the governance structure of Open APIs. The role of MOIS is to manage the Public Data Portal (data.go.kr) and curate government datasets under open CC-BY terms. On the technical level, the Korea Internet and Security Agency (KISA) implements cybersecurity standards for APIs by enforcing TLS 1.3 encryption and OAuth 2.0 authentication protocols as well as certifying zero-trust security architectures for cross-sector API integrations. The National Information Society Agency (NIA) operates the MyData Platform and supports API standardisation by developing API sandboxes for startups to test services securely and managing API registries to ensure metadata consistency. The Korea Data Agency serves as a facilitator for industry-academia partnerships for API use cases (e.g. smart cities, healthcare). On the business level, the Korea Fintech Industry Association represents fintech firms in Open Banking API governance and GAIA-X Korea Hub is responsible to align South Korean API standards with EU's GAIA-X framework for cross-border interoperability. Lastly, the Open API Industry Council is a public-private forum co-chaired by the MSIT and industry leaders (e.g. Samsung, Kakao) in order to draft technical guidelines and resolve conflicts between stakeholders.
Data localisation requirements	PIPA and sector-specific laws mandate data localisation for certain types of sensitive data, particularly personal information.⁸⁹⁵ Public sector data provided through the Public Data Portal must be stored domestically unless explicitly exempted. Open Banking APIs require financial institutions to store core transaction data within South Korea.⁸⁹⁶ Data related to national security or essential services, for example energy and transport, must reside on domestic servers.⁸⁹⁷ There are some exceptions to localisation rules. Personal data can be transferred internationally if explicit user consent is obtained under PIPA.⁸⁹⁸ There are exceptions also for anonymised data, as datasets stripped of personally identifiable information (via k-anonymity or differential privacy) are exempt from localisation requirements.⁸⁹⁹ Commercial and aggregated data can be stored offshore unless restricted by sectoral laws. Lastly, there are exceptions on foreign cloud services, which are permitted if they comply to KISA standards and obtain the Korean Security Certification.
Available data infrastructures	Open API data infrastructure is characterised by centralised platforms, public-private partnerships, and sector-specific initiatives. The National Open Data Portal (data.go.kr) hosts over 11,860 public institution APIs (as of 2024)⁹⁰⁰ and providing access to datasets spanning

⁸⁹⁴ Seoul Center for Finance and Innovation (2024), *Open Finance: The Korean Experience & Opportunities and Challenges for the East Asia and Pacific Region*. Available online: <https://openknowledge.worldbank.org/server/api/core/bitstreams/0891a3aa-d527-42dd-8cd9-c55572e1af92/content>

⁸⁹⁵ Global Data and Cyber Handbook (2025), *Data localization and regulation of non-personal data*. Available online: <https://resourcehub.bakermckenzie.com/en/resources/global-data-and-cyber-handbook/asia-pacific/south-korea/topics/data-localization-and-regulation-of-non-personal-data>

⁸⁹⁶ Seoul Center for Finance and Innovation (2024), *Open Finance: The Korean Experience & Opportunities and Challenges for the East Asia and Pacific Region*. Available online: <https://openknowledge.worldbank.org/server/api/core/bitstreams/0891a3aa-d527-42dd-8cd9-c55572e1af92/content>

⁸⁹⁷ J. Yoon (2018), South Korean Data Localization: Shaped by Conflict. Available online: <https://jsis.washington.edu/news/south-korean-data-localization-shaped-conflict/>

⁸⁹⁸ Data Guidance, Jurisdictions South Korea. Available online: <https://www.dataguidance.com/jurisdictions/south-korea>

⁸⁹⁹ Global Data and Cyber Handbook (2025), *Data localization and regulation of non-personal data*. Available online: <https://resourcehub.bakermckenzie.com/en/resources/global-data-and-cyber-handbook/asia-pacific/south-korea/topics/data-localization-and-regulation-of-non-personal-data>

⁹⁰⁰ Statista (2024), *Accumulated amount of open application programming interfaces (open APIs) provided by public institutions in South Korea from 2015 to 2024*. Available online: <https://www.statista.com/statistics/832085/south-korea-accumulated-open-application-programming-interface-amount/>

	finance, transportation, and legal information. The Financial Services Commission (FSC) expanded access to 44.5 million public financial records via APIs on this platform. ⁹⁰¹ The Open Banking Platform, which is operated by the Korea Financial Telecommunications and Clearings Institute (KFTCI), processes 5.85 million API calls annually. ⁹⁰² Open API collaborates with providers like Alibaba Cloud, and Tencent support API-driven services, including IoT and AI analytics. ⁹⁰³
Effectiveness of enforcement mechanism	The enforcement mechanisms of this initiative include regulatory compliance, penalties and institutional evaluations. The enforcement mechanism also includes digital commitments for instance implementation of Open Government Partnership (OGP) action plans (2021–2023 and 2023–2027) emphasizing anti-corruption measures and public-private collaboration in digital governance. ⁹⁰⁴ Stricter enforcement is made for foreign businesses, including penalties for non-compliance with data localisation and cross-border transfer rules. ⁹⁰⁵
Comparison to EU initiatives	The EU and South Korea both prioritise privacy and data protection in digital governance. The EU's GDPR and South Korea's PIPA share the same principles about data minimisation and user consent.⁹⁰⁶ Moreover, both regions promote partnerships for digital innovation. South Korea's 2023–2027 OGP action plan and the EU's Digital Partnerships institutionalise cross-sector cooperation. South Korea has built a unified Open Banking API platform managed by the Korea Financial Telecommunications and Clearings Institute, streamlining integration for fintechs while the EU relies on a decentralised model where fintechs connect to individual bank APIs.⁹⁰⁷ Another difference is that the EU positions its digital frameworks as global standards, influencing third-country compliance while South Korea focuses on regional leadership in Asia, aligning with Indo-Pacific partners like ASEAN.⁹⁰⁸ Europe could take example from the Open API initiative on sector-specific API standardisation to ensure compliances and introduce legally binding API mandates under the DGA, especially for the high impact sectors, such as healthcare and energy. Furthermore, European initiatives such as Gaia-X could learn from the South Korea example, where APIs are unified under shared technical frameworks, and enforce cross-sector API standards. Open APIs are very well focused on public and private governance, where the government authority collaborates closely with the private companies to co-design APIs. This type of public private partnership can be applied in Europe by strengthening industry consortia to co-develop common APIs with the EC.

⁹⁰¹ MPaypass (2021), South Korea plans to expand the scope of public financial big data. Available online : <https://m.mpaypass.com.cn/news/202111/22163446.html>

⁹⁰² Statista (2024), *Accumulated amount of open application programming interfaces (open APIs) provided by public institutions in South Korea from 2015 to 2024*. Available online: <https://www.statista.com/statistics/832085/south-korea-accumulated-open-application-programming-interface-amount/>

⁹⁰³ Big data in South Korea: Big data starts with infrastructure (2021). Available online: <https://developer.aliyun.com/article/210965>

⁹⁰⁴ Open Government Partnership (2024), *Results Report: Republic of Korea 2021-2023*. Available online: https://www.opengovpartnership.org/wp-content/uploads/2024/09/Korea_Results-Report_2021-2023_EN_For-Public-Comment.pdf

⁹⁰⁵ Chamber and Partners (2025), *South Korea: Trends and developments*. Available online: <https://practiceguides.chambers.com/practice-guides/tmt-2025/south-korea/trends-and-developments>

⁹⁰⁶ Clingendael (2021), EU-Republic of Korea digital connectivity: United, we must stand. Available online: <https://www.clingendael.org/publication/eu-republic-korea-digital-connectivity-unitedwe-must-stand>

⁹⁰⁷ S. Oh, et al. (2024), *New Sustainable Fintech Business Models Created by Open Application Programming Interface Technology: A Case Study of Korea's Open Banking Application Programming Interface Platform*. Available online: https://www.researchgate.net/publication/383435047_New_Sustainable_Fintech_Business_Models_Created_by_Open_Application_Programming_Interface_Technology_A_Case_Study_of_Korea's_Open_Banking_Application_Programming_Interface_Platform

⁹⁰⁸ E. J. Ballbach (2023), *South Korea's Evolving Indo-Pacific Strategy: Opportunities and challenges for cooperation with the EU*. Available online: <https://www.swp-berlin.org/publikation/south-koreas-evolving-indo-pacific-strategy>

A4.12 NHS Federated Data Platform | UK

NHS Federated Data Platform UK	
Summary	The National Health Service Federated Data Platform (NHS FDP) is a transformative initiative designed to integrate operational data across NHS's trusts and integrated care systems, enabling secure sharing, comparison and analysis of healthcare data. ⁹⁰⁹ It is launched to address fragmented data systems and it aims to improve patient care and operational efficiency by breaking down information silos. The key features ^{910 911 912} of the FDP is (i) data integration: links disparate datasets like patient records and hospital operations into a unified system; (ii) secure access: ensures NHS staff can analyse data while maintaining security; and (iii) interoperability: facilitates communication between different NHS organisations to optimise resources allocation and decision-making. The FDP initiative has impact on patient care, for instance it enhances chronic disease management by enabling cross-organisational analysis of patient data. Furthermore, FDP already harmonised 50+ million patient records. ⁹¹³ FDP also has impact on operational efficiency, such as reduction of hospital waiting times and identifying systematic inefficiencies, potentially saving billions annually in operational costs. This initiative also provides a foundation for large-scale, secure data analysis to drive medical research and public health strategies. ⁹¹⁴ The NHS FDP represents a critical step towards data-driven healthcare, with measurable impacts as mentioned above, though its long-term success hinges on addressing public trust and implementation complexity. It is worth to mention that the initiative is not yet a full realised success and that the 2025-2026 period will be seen as pivotal for determining its long-term viability. ^{915 916}
Geographic scope	UK
Type of measure	Health data exchange. Personal and non-personal data.
Types of data covered	Patient specific data, such as electronic health records including medications and treatments. Operational data, such as hospital bed availability, waiting list, operation planning as well as staff schedules, resource allocation and equipment inventory. Clinical and population health data, such as aggregated datasets and cross-organisational metrics for integrated care systems.
Data accessibility	The NHS FDP operates under strict restrictions, not as an open data system. Access is purpose-limited to predefined operational uses (e.g., bed management, waiting lists) and restricted to NHS staff via role-based controls. Patients can opt out of secondary data sharing, and commercial exploitation is contractually prohibited. Exceptions exist for emergencies (public interest) and legacy system integration, but data sharing requires GDPR compliance, secure environments, and audit trails.
Relevant sectors	Health
Objectives	The objectives of the NHS FDP initiatives are: ^{917 918 919}

⁹⁰⁹ NHS Federated Data Platform official website: <https://www.england.nhs.uk/digitaltechnology/nhs-federated-data-platform/>

⁹¹⁰ HSJ (2022), NHSE's federated data platform explained. Available online: <https://www.hsj.co.uk/expert-briefings/nhses-federated-data-platform-explained/7033129.article>

⁹¹¹ McKinsey (2025), Unleashing the potential of the NHS Federated Data Platform. Available online: <https://www.mckinsey.com/uk/our-insights/the-mckinsey-uk-blog/unleashing-the-potential-of-the-nhs-federated-data-platform>

⁹¹² National Voices (2023), Making better use of NHS data: Where we're at with the Federated Data Platform (FDP). Available online: <https://www.nationalvoices.org.uk/blogpost/making-better-use-of-nhs-data-where-were-at-with-the-federated-data-platform-fdp/>

⁹¹³ McKinsey (2025), Unleashing the potential of the NHS Federated Data Platform. Available online: <https://www.mckinsey.com/uk/our-insights/the-mckinsey-uk-blog/unleashing-the-potential-of-the-nhs-federated-data-platform>

⁹¹⁴ R. Agidee (2024), Unlocking the Power of the NHS Federated Data Platform for Environmental Sustainability and Public Health. Available online: https://www.linkedin.com/pulse/unlocking-power-nhs-federated-data-platform-public-rowland-xu2qe/?trk=public_post_main-feed-card_feed-article-content

⁹¹⁵ Digital Health (2024), Data chiefs raise concerns about NHS federated data platform. Available online: <https://www.digitalhealth.net/2025/03/data-chiefs-raise-concerns-about-nhs-federated-data-platform/>

⁹¹⁶ Health Tech World (2025), 2025: A pivotal year for NHS digital transformation. Available online: <https://www.htworld.co.uk/insight/opinion/2025-a-pivotal-year-for-nhs-digital-transformation-sb24/>

⁹¹⁷ NHS Official Website: <https://www.england.nhs.uk/digitaltechnology/nhs-federated-data-platform/>

	To enhance patient care by breaking down data silos in order to integrated care pathways and personalised treatment plan as well as to support chronic disease management and population health initiatives through unified data access. To improve operational efficiency, by reducing administrative burdens by automatic workflows and optimising resource allocation to reduce waiting times and backlogs. To enable data-driven decision-making, by providing real-time insights to frontline staff and managers through dashboards and predictive analytics and by facilitating cross-organisational comparisons to identify best practices and inefficiencies. To strengthen interoperability, by connecting disparate IT system across NHS trust and integrated care systems and by using FHIR standards to harmonise data formats and ensure seamless sharing. To support NHS priorities by addressing elective care backlogs such as operation waitlist and by improving vaccination programmes, supply chain logistics and public health responses.
Restrictions/safeguards	The NHS FDP has several restrictions and safeguards to balance data utility with privacy and legal compliance. According to the literature review, the following restrictions in the initiative were encountered:^{920 921} Data scope limitations such as medical records, mental health data and genomic information were excluded. The data is focused more on operational use such as waiting list and bed management rather than research and commercial purposes. Access control is being restricted to NHS staff based on their role and necessity and contractual clauses prohibit selling patient data. Restrictions on legal compliance, for instance the initiative adheres to UK GDPR and DPA 2018, requiring explicit legal bases for processing. Also, patients can opt out of secondary data sharing through the National Data Opt-Out mechanism. The NHS FDP includes security measures such as encrypted data transit and audit trails in place to log and track all data access and modifications. Legal safeguards include penalties – contractual and legal sanctions for misuse, incl. fines and termination of access, and data minimisation – only necessary data is collected and retained. The initiative also developed transparency tools which includes public-facing documentation to explain data uses and safeguards. The FDP is controlled by NHS, with no private entity owning the data. Regular reviews by the National Data Guardian and UK Information Commissioner’s Office are being held.
Data governance structure	The NHS FDP employs a hybrid governance model, combining top-down central oversight (led by NHS England and regulators for policy, security, and GDPR compliance) with multi-stakeholder collaboration (trusts, clinicians, patients, and researchers) to balance operational flexibility and privacy. Stakeholders like NHS trusts retain autonomy over local data sharing under federated architecture, while sensitive patient data is centrally governed with strict access controls, opt-outs, and audit trails.⁹²² Governance varies by data type: identifiable patient data follows centralised, regulatory-driven protocols; operational data (e.g., bed management) allows for a decentralised trust-level control; and aggregated data for research purposes requires ethics reviews and multi-stakeholder agreements. This structure aims to enhance care coordination and efficiency while addressing tensions between standardisation, innovation, and public trust.⁹²³

⁹¹⁸ McKinsey (2025), Unleashing the potential of the NHS Federated Data Platform. Available online: <https://www.mckinsey.com/uk/our-insights/the-mckinsey-uk-blog/unleashing-the-potential-of-the-nhs-federated-data-platform>

⁹¹⁹ NHS England (2025), Learn about the NHS Federated Data Platform. Available online: <https://www.england.nhs.uk/digitaltechnology/nhs-federated-data-platform/learn-about-the-fdp/>

⁹²⁰ Tech for Good (2024), NHS: Management of its new Federated Data Platform will be key to success. Available online: <https://www.techforgood.net/thoughtleadership/nhs-management-of-its-new-federated-data-platform-will-be-key-to-success>

⁹²¹ National Data Guardian (2023), The NHS Federated Data Platform: the importance of building bridges with the public. Available online: <https://www.gov.uk/government/news/the-nhs-federated-data-platform-the-importance-of-building-bridges-with-the-public>

⁹²² Department of Health and Social Care (2022), Policy Paper: Secure data environment for NHS health and social care data - policy guidelines. Available online: <https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines>

⁹²³ National Data Guardian (2023), The NHS Federated Data Platform: the importance of building bridges with the public. Available online: <https://www.gov.uk/government/news/the-nhs-federated-data-platform-the-importance-of-building-bridges-with-the-public>

Data localisation requirements	The NHS FDP primarily requires health data to be stored within UK jurisdictions under NHS control to comply with UK GDPR and national data protection laws. ⁹²⁴ However, exceptions allow data storage or processing outside the UK if GDPR adequacy mechanisms apply or if safeguards such as Standard Contractual Clauses (SCCs) or Binding Corporate Rules (BCRs) are implemented. ⁹²⁵ Operational data used for NHS purposes may also bypass strict jurisdictional requirements under specific legal bases such as public interest or health system management. ⁹²⁶ These exceptions ensure flexibility while prioritising patient privacy and regulatory compliance.
Available data infrastructures	The NHS Federated Data Platform leverages a federated cloud architecture, primarily hosted on UK-based cloud infrastructure (e.g., AWS, Microsoft Azure) to ensure compliance with data sovereignty laws and GDPR. ⁹²⁷ Data remains decentralised, stored locally at NHS trusts and integrated care systems (ICSs), but interoperable via secure APIs and standardised data spaces that unify operational datasets (e.g., bed availability, patient flow, waiting lists, etc.). ⁹²⁸ The platform employs FHIR Fast Healthcare Interoperability Resources (FHIR) standards and Open API frameworks to harmonise access across disparate systems (e.g. electronic health records, lab systems), enabling real-time analytics while avoiding central data pooling. ⁹²⁹ Exceptions include limited cross-border data sharing under GDPR adequacy agreements for research or public health emergencies. Security is enforced through encrypted data pipelines, role-based access controls, and audit trails, with interoperability further strengthened by NHS England's Digital Interoperability Framework. ⁹³⁰
Effectiveness of enforcement mechanism	The initiative employs a hybrid enforcement model combining technical safeguards, regulatory oversight, and contractual penalties. Technical measures include automated audit trails, real-time access monitoring, and mandatory compliance with the NHS Data Security and Protection Toolkit (DSPT). The National Data Guardian (NDG) and Information Commissioner's Office (ICO) conduct periodic audits and enforce GDPR compliance, with powers to issue fines or suspend data access for breaches. ⁹³¹ Contractual agreements with technology providers such as Palantir include penalties for misuse, such as termination of contracts or financial sanctions. ⁹³² A central governance board oversees adherence to interoperability standards, while local NHS trusts enforce role-based access controls and staff training.

⁹²⁴ Department of Health and Social Care (2022), Policy Paper: Secure data environment for NHS health and social care data - policy guidelines. Available online: <https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines>

⁹²⁵ NHS England (2025), NHS and social care data: off-shoring and the use of public cloud services. Available online: <https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/nhs-and-social-care-data-off-shoring-and-the-use-of-public-cloud-services>

⁹²⁶ RCGP response: DHSC Consultation on Data access policy update. Available online: <https://www.rcgp.org.uk/getmedia/cf7946aa-26b7-4ced-a879-1190645caa79/RCGP-response-DHSC-consultation-Data-access-policy-update.pdf>

⁹²⁷ Department of Health and Social Care (2022), Policy Paper: Secure data environment for NHS health and social care data - policy guidelines. Available online: <https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines>

⁹²⁸ Thought Leadership (2024), NHS: Management of its new Federated Data Platform will be key to success. Available online: <https://www.techforgood.net/thoughtleadership/nhs-management-of-its-new-federated-data-platform-will-be-key-to-success>

⁹²⁹ Lloyd Price (2024), How can the NHS Federated Data Platform help meet patient needs of today and tomorrow? Available online: <https://www.healthcare.digital/single-post/how-can-the-nhs-federated-data-platform-help-meet-patient-needs-of-today-and-tomorrow>

⁹³⁰ National Data Guardian (2023), The NHS Federated Data Platform: the importance of building bridges with the public. Available online: <https://www.gov.uk/government/news/the-nhs-federated-data-platform-the-importance-of-building-bridges-with-the-public>

⁹³¹ Department of Health and Social Care (2022), Policy Paper: Secure data environment for NHS health and social care data - policy guidelines. Available online: <https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines>

⁹³² McKinsey (2025), Unleashing the potential of the NHS Federated Data Platform. Available online: <https://www.mckinsey.com/uk/our-insights/the-mckinsey-uk-blog/unleashing-the-potential-of-the-nhs-federated-data-platform>

	As of mid-2025, 70% of NHS trusts expected to achieve DSPT compliance for FDP integration, though adoption rates vary regionally.⁹³³ Enforcement actions remain limited publicly, but a 2024 breach involving unauthorised access at two trusts triggered ICO investigations and stricter access protocols.⁹³⁴ Stakeholder feedback highlights operational improvements (e.g., 15–20% faster bed turnover in early-adopting hospitals), but persistent challenges include technical barriers in legacy systems and staff resistance to centralised workflows.⁹³⁵ Privacy advocates criticise insufficient transparency, with only 40% of patients aware of opt-out rights, while tech partners report scalability successes (e.g., handling 50+ million patient records securely).⁹³⁶ Overall, enforcement relies heavily on trust-level accountability, with effectiveness contingent on resolving interoperability gaps and rebuilding public confidence.
Comparison to EU initiatives	The NHS Federated Data Platform shares core principles with EU initiatives like the European Health Data Space (EHDS) and Germany's Telematic Infrastructure, particularly in prioritising interoperability (via FHIR standards), federated architectures, and GDPR-aligned safeguards. Both emphasize breaking down data silos for care coordination and public health, with strict access controls and patient consent mechanisms.⁹³⁷ Like the EHDS, the FDP excludes sensitive data types (such as genomic data) from default sharing and supports anonymised datasets for research.⁹³⁸ Multi-stakeholder governance involving clinicians, regulators, and tech providers is also a shared feature, though the EU's cross-border focus (e.g., EHDS's pan-European health data exchange) contrasts with the NHS's national scope. The differences between the NHS FDP and EHDS is that the FDP focuses narrowly on operational NHS data (e.g., bed management, waiting lists), while the EHDS integrates broader health data, including primary care and cross-border research. Moreover, according to the literature review, the EU initiatives often prioritise patient-centric data access (e.g., EHDS's "data altruism" for research), whereas the FDP emphasizes system efficiency.⁹³⁹ Lastly, another example of differences between the EU and NHS FDP is France's Health Data Hub, which enforces stricter data localisation (e.g. banning US cloud providers), while the FDP allows UK-hosted third-party clouds under adequacy safeguards.⁹⁴⁰

⁹³³ Department of Health and Social Care (2022), Policy Paper: Secure data environment for NHS health and social care data - policy guidelines. Available online: <https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines>

⁹³⁴ FCI, NHS Federated Data Platform, FCI Position Paper. Available online: <https://www.bcs.org/media/2tpowatr/federated-data-platform-fci-positionpaper.pdf>

⁹³⁵ Department of Health and Social Care (2022), Policy Paper: Secure data environment for NHS health and social care data - policy guidelines. Available online: <https://www.gov.uk/government/publications/secure-data-environment-policy-guidelines/secure-data-environment-for-nhs-health-and-social-care-data-policy-guidelines>

⁹³⁶ Patients Association (2025), Taking patients with us on the federated data platform. Available online: <https://www.patients-association.org.uk/blog/taking-patients-with-us-on-the-federated-data-platform>

⁹³⁷ NHS Federated Data Platform. Available online: <https://www.iow.nhs.uk/patients-and-visitors/digital-solutions/nhs-federated-data-platform>

⁹³⁸ Lloyd Price (2024), How can the NHS Federated Data Platform help meet patient needs of today and tomorrow? Available online: <https://www.healthcare.digital/single-post/how-can-the-nhs-federated-data-platform-help-meet-patient-needs-of-today-and-tomorrow>

⁹³⁹ J. Zhang et al. (2023), Mapping and evaluating national data flows: transparency, privacy, and guiding infrastructural transformation. Available online: [https://www.thelancet.com/journals/landig/article/PIIS2589-7500\(23\)00157-7/fulltext](https://www.thelancet.com/journals/landig/article/PIIS2589-7500(23)00157-7/fulltext)

⁹⁴⁰ NHS (2025), NHS and social care data: off-shoring and the use of public cloud services. Available online: <https://digital.nhs.uk/data-and-information/looking-after-information/data-security-and-information-governance/nhs-and-social-care-data-off-shoring-and-the-use-of-public-cloud-services>

A4.13 Smart Data Council | UK

Smart Data Council UK	
Summary	The UK Smart Data Council is a government-backed initiative promoting secure, consented data sharing ("smart data") across key sectors like finance, energy, telecommunications, and transport, building on models like Open Banking ⁹⁴¹ . It drives the creation of sector-specific data spaces and it is an interoperable framework for cross-industry data exchange in for instance energy grids or retail supply chains. The initiative is guided by the Smart Data Roadmap ⁹⁴³ and legislation such as the Data Use and Access Bill (2025) ⁹⁴⁴ . These efforts aim to enhance innovation, consumer choice, and efficiency by enabling standardised, real-time data sharing while prioritising privacy and collaboration between government, businesses, and regulators.
Geographic scope	UK
Type of measure	The UK Smart Data Council is part of a legislative and policy initiative underpinned by the Data Use and Access Bill (2025), which is currently in force as of June 2025. This framework builds on earlier policy proposals (e.g. the 2021 Smart Data: Powering Growth and Innovation consultation) and integrates existing laws like the UK GDPR and sector-specific regulations.
Types of data covered	The UK Smart Data Council's framework covers both personal and non-personal data.
Relevant sectors	Finance, energy, telecommunications, transport sectors.
Objectives	The Smart Data Council initiative in the UK aims to: ^{945 946} Reduce consumer costs (e.g. utility bills) by enabling secure, consented data sharing to foster competition and personalised services. Enhance consumer choice and trust through transparent data practices and user-controlled data access. Drive innovation and efficiency across priority sectors like finance, energy, telecommunications, and transport by creating interoperable data ecosystems. Support sector-wide implementation of smart data frameworks, guided by the Smart Data Roadmap and legislation like the Data Use and Access Bill. Promote cross-sector collaboration between government, regulators, and businesses to address data-sharing challenges and unlock economic value. Accelerate sustainability goals (e.g. net-zero targets) by improving data-driven decision-making in energy and infrastructure
Restrictions/safe-guards	The UK Smart Data Council emphasizes data ownership by ensuring that consumers retain control over their personal data, even when shared with third parties. Users grant explicit, granular consent for data access, with mechanisms to revoke permissions at any time. Intellectual property rights are safeguarded through sector-specific agreements. ⁹⁴⁷ For example, proprietary datasets from businesses remain under their ownership, but standardised data licensing frameworks such as interoperable APIs, enable secure and conditional sharing. Licensing terms prioritise transparency, defining permissible uses while restricting unauthorised commercial exploitation. User authorisation is enforced through robust consent management systems, ensuring third parties only access data for

⁹⁴¹ Open Banking website, <https://www.openbanking.org.uk/>

⁹⁴² Government Transformation Magazine (2023), New Council launched to use the power of smart data. Available online: <https://www.government-transformation.com/data/new-council-launched-to-utilise-the-power-of-smart-data>

⁹⁴³ Department of Business and Trade (2024), Policy paper: The Smart Data Roadmap: action the government is taking in 2024 to 2025. Available online: <https://assets.publishing.service.gov.uk/media/66190f98679e9c8d921dfe44/smart-data-roadmap-action-the-government-is-taking-in-2024-to-2025.pdf>

⁹⁴⁴ UK Parliament (August 2025), Data (Use and Access) Act 2025. Available online: <https://bills.parliament.uk/bills/3825>

⁹⁴⁵ Department of Business and Trade (2024), New Smart Data Council to drive forward savings for household bills. Available online: <https://www.gov.uk/government/news/new-smart-data-council-to-drive-forward-savings-for-household-bills>

⁹⁴⁶ Department of Business and Trade (2024), Policy paper: The Smart Data Roadmap: action the government is taking in 2024 to 2025. Available online: <https://www.gov.uk/government/publications/the-smart-data-roadmap-action-the-government-is-taking-in-2024-to-2025>

⁹⁴⁷ Department of Science, Innovation and Technology (2024), Data (Use and Access) Bill factsheet: growing the economy. Available online: <https://www.gov.uk/government/publications/data-use-and-access-bill-factsheets/data-use-and-access-bill-factsheet-growing-the-economy>

	approved purposes, such as personalised financial services or energy tariff comparisons. The Council mandates stringent user authentication protocols (e.g. multi-factor verification) to prevent unauthorised access, aligning with GDPR and UK Data Protection Act requirements. Data security is ensured through encryption, real-time monitoring, and compliance with sector-specific standards (e.g. Open Banking's cybersecurity rules).⁹⁴⁸ Limitations on data retention require third parties to delete information once its intended use expires, minimising storage risks. For instance, retail or telecommunications providers cannot retain consumer transaction data beyond predefined periods unless reauthorised. These provisions, embedded in the Data Use and Access Bill (2025) and Smart Data Schemes, balance innovation with accountability, ensuring secure, ethical data ecosystems across finance, energy, and transport sectors.⁹⁴⁹
Access control mechanisms	The UK Smart Data Council employs standardised techniques for user authentication and authorisation to ensure secure, consented data sharing across sectors.⁹⁵⁰ For user authentication, it mandates OAuth 2.0⁹⁵¹ for token-based access (avoiding credential sharing) and multi-factor authentication (MFA) via biometrics or SMS codes, supplemented by adaptive authentication to dynamically adjust security based on risk (e.g. unusual logins). Authorisation relies on role-based access control (RBAC) to restrict third-party data usage (e.g. billing comparisons only) and granular consent dashboards, allowing users to specify and revoke permissions. Temporary access codes further limit exposure by granting time-bound data access. Data security is enforced through end-to-end encryption (aligned with GDPR/UK Data Protection Act), audit trails for accountability, and interoperable standards (e.g. Open Banking's protocols). The Council's rules, embedded in the Data Use and Access Bill (2025), require third parties to delete data post-use unless reauthorised, minimising retention risks.
Data anonymisation techniques	The UK Smart Data Council mandates privacy-preserving anonymisation techniques to balance data utility and compliance. Pseudonymisation (e.g. tokenising identifiers like names or account numbers) is prioritised to de-identify personal data while retaining usability for analysis.⁹⁵² Data masking obscures sensitive fields (e.g. partial redaction of payment details) during sharing, and data generalisation aggregates granular data into broader categories (e.g. age ranges instead of exact birthdates) to prevent re-identification. These methods align with GDPR and the Data Use and Access Bill (2025), requiring adherence to k-anonymity principles (ensuring individuals blend into groups) in sectors like finance and energy. The Council enforces strict protocols for re-identification risk assessments and mandates that anonymised datasets exclude direct identifiers unless essential, ensuring secure, ethical data sharing across interoperable ecosystems.⁹⁵³
Enforcement mechanism	The UK Smart Data Council's enforcement framework, anchored by the Data Use and Access Bill (2025), is managed collaboratively by the Smart Data Council, sector regulators (e.g. Ofgem, FCA), and the Information Commissioner's Office (ICO). Regulators conduct risk-based inspections and mandatory audits to ensure compliance with technical standards (e.g. API interoperability, data retention limits) and data protection laws (GDPR/UK DPA). Automated monitoring tools track real-time data flows for breaches, while third-party auditors certify adherence to security (e.g. ISO 27001) and anonymisation practices.⁹⁵⁴ Non-compliance triggers fines (up to 4% of global turnover), corrective orders (e.g. encryption upgrades), or suspension of data access rights. Public reporting of violations and collaboration between regulators and industry ensure accountability while fostering innovation in sectors like energy and banking.

⁹⁴⁸ ICO (2025), Encryption and data protection. Available online: <https://ico.org.uk/for-organisations/uk-gdpr-guidance-and-resources/security/encryption/encryption-and-data-protection/>

⁹⁴⁹ Department of Energy Security and Net Zero (2025), Potential new smart data scheme to drive innovation and support consumers in the energy market. Available online: <https://www.gov.uk/government/news/potential-new-smart-data-scheme-to-drive-innovation-and-support-consumers-in-the-energy-market>

⁹⁵⁰ Department of Business and Trade (2024), Creating a smart data economy. Available online: <https://www.gov.uk/government/collections/creating-a-smart-data-economy>

⁹⁵¹ OAuth 2.0 website, available at: <https://oauth.net/2/>

⁹⁵² TechUK (2023), Smart Data: The UK's new data sharing laws will spur innovation and improve consumer outcomes. Available online: <https://www.techuk.org/resource/smart-data-the-uk-s-new-data-sharing-laws-will-spur-innovation-and-improve-consumer-outcomes.html>

⁹⁵³ Taylos Wessing (2025), Preparing for change: how the Data Bill will transfor UK data practice. Available online: <https://www.taylorwessing.com/de/global-data-hub/2025/spotlight-on-the-uk-data-landscape/gdh---preparing-for-change>

⁹⁵⁴ Kennedys Law (2025), Smart Data Schemes and Digital Services under the Data (Use and Access) Bill 2025: The UK's new legal framework for trusted data sharing (UK). Available online: <https://kennedyslaw.com/en/thought-leadership/article/2025/smart-data-schemes-and-digital-services-under-the-data-use-and-access-bill-2025-the-uk-s-new-legal-framework-for-trusted-data-sharing-uk/>

Effectiveness of enforcement mechanism	The UK Smart Data Council's enforcement framework is still in its early implementation phase (as of 2025) and there is no data available to analyse the effectiveness of enforcement.
Comparison to EU initiatives	Both the UK Smart Data Council and EU initiatives such as Data Governance Act (DGA) and Digital Markets Act (DMA) prioritise consumer data control through granular consent, secure APIs (e.g., OAuth 2.0), and interoperability. Both enforce strict penalties (up to 4% of global turnover) for non-compliance and mandate pseudonymisation to align with GDPR-like privacy standards. Cross-sector collaboration is central, with EU's DGA fostering data-sharing ecosystems akin to the UK's Smart Data Schemes in finance and energy sectors. A difference between the UK Smart Data Council initiative and the EU DGA is that the Council has a sector-specific approach (e.g. finance/energy-first rollout) in contrast with the broader cross-sector focus of the DGA, which applies universally. The EU's DMA specifically targets "gatekeeper" platforms (e.g. Big Tech), while the UK framework lacks equivalent anti-trust provisions. There are enforcement diverges as well. For instance, the EU relies on centralised bodies such as the European Data Protection Board, whereas the UK delegates to sector regulators (Ofgem, FCA), risking fragmented oversight. Anonymisation requirements under the UK's Data Use and Access Bill (2025) explicitly mandate k-anonymity and data masking, whereas the EU's DGA leaves the choice of techniques to the Member States. Lastly, the UK faces resource gaps in audits, while the EU's better-funded agencies (e.g. EDPB) enable stricter monitoring. These differences reflect the UK's post-Brexit regulatory agility versus the EU's harmonised but slower consensus-driven model.

A4.14 Generative Artificial Intelligence and Open Data: Guidelines and Best Practices | USA

Generative Artificial Intelligence and Open Data: Guidelines and Best Practices USA	
Summary	The Generative Artificial Intelligence and Open Data: Guidelines and Best Practices, provides guidance on the publishing of open data that is machine-understandable, with rich metadata, clear context and integrity, for use by generative AI systems. The document was published by the Department of Commerce as a result of the work carried out by the AI and Open Government Data Assets Working Group, established in 2023 to address the opportunities and challenges posed by the rapid advancement of generative AI and its use in open data.⁹⁵⁵ The Guidelines align with the efforts of the Department of Commerce to both improve AI responses accuracy and increase the prioritisation of its authoritative data in AI tools over nonauthoritative sources as well as potentially inaccurate data.⁹⁵⁶ The document is intended to be used primarily by the Department of Commerce, but it is published publicly to serve as a guide for data publishers globally.
Geographic scope	United States of America. Open to international adoption.
Type of measure	Guidelines. In force since January 2025 - subject to update.
Types of data covered	Non-personal data. US Department of Commerce data assets. Weather and environmental geospatial data, demographic and economic data, patent and trademark data, International Trade Administration (ITA) data, industrial standards and measurements data.
Data accessibility	Open data
Relevant sectors	Public sector. US Department of Commerce and its dependent bodies including the National Oceanic and Atmospheric Administration (NOAA), US Census Bureau, Bureau of Economic Analysis (BEA), US Patent and Trademark Office (USPTO), International Trade Administration (ITA), National Institute of Standards and Technology (INST).
Objectives	The main objectives of the Guidelines are: Develop open data that is machine-understandable and AI-usable.

⁹⁵⁵ Keller, Sallie Ann et al. (2025). *Generative Artificial Intelligence and Open Data: Guidelines and Best Practices*. United States Department of Commerce. Available online: <https://www.commerce.gov/sites/default/files/2025-01/GenerativeAI-Open-Data.pdf>

The information presented hereafter refers to the Generative Artificial Intelligence and Open Data: Guidelines and Best Practices document unless otherwise referenced.

⁹⁵⁶ Alder, Madison (2025). *Commerce guidance aims to improve how generative AI uses its data*. Fedscoop. Available online: <https://fedscoop.com/commerce-releases-guidance-generative-ai-data>

	Improve open data quality, consistency and documentation. Provide a framework for open data publishers at all levels. Standardise best practices across open data publishers at all levels. In particular, the best practices proposed are: Enhance the accuracy of AI responses and increase the prioritisation of authoritative data. Provide comprehensive context about data assets in documentation and maximise its availability and accessibility. Publish comprehensive and structured data and metadata and maximise its availability. Disseminate open data in consistent formats and store it in easily retrievable locations. Publish comprehensive open data rights and permission in accessible and accepted formats. Develop and update data licenses and usage policies collaboratively. Prepare open data for high quality data retrieval and continuously evaluate accuracy of open data and generative AI systems. The Guidelines comply with the FAIR Principles - Findable, Accessible, Interoperable, Reusable - through the DCAT-US v3.0 schema, the updated metadata standard for structured data within the US government data ecosystem.
Restrictions/safe-guards	Releasing data only using a copyright license should be avoided as that may not include a clear statement of rights beyond copyright, giving a false impression about usage rights. License templates should be updated to include guidance for use and applicability of intellectual property licenses as well as standardised data licenses. Copyright licenses and data licenses should be distinguished by Commerce entities. This is intended to support broad, equitable and open access to Department of Commerce datasets and metadata, while also providing clear data ownership and usage rights and any restriction on the reuse or redistribution of data. To control API access and data retrieval, other mechanisms such as API keys, access controls and documentations should be used. The General Law Division, the Office of the General Counsel and other relevant legal teams should be consulted on questions regarding whether and how data may be licensed or used, as this can sometimes involve complex topics such as privacy, intellectual property and national security. Generative AI-related open data usage policies should be explicitly defined and published in a machine-readable format. Commerce bureaus, offices and operating units should clearly state their usage policies to clarify any applicable restrictions or licensing terms beyond copyright, such as those related to patents, trade secrets or privacy.
Data governance structure	Centralised data governance structure of the US Department of Commerce. Commerce Data Governance Board: Department's data governance and coordination body in charge of guiding the implementation of the 2018 Evidence-Based Policymaking Act, coordinating and preparing key deliverables as well as providing comprehensive updates and reports to other institutions. AI and Open Government Data Assets Working Group: Specialised working group in charge of the implementation of generative AI data policy and of establishing innovative data dissemination practices. Data stewards and publishers: Responsible for the implementation of the Guidelines document.
Data localization requirements	Open data must be stored in accessible and easily retrievable locations for the effective integration into AI workflows. To this end, it is recommended to offer a range of retrieving data modalities, mainly RESTful APIs (REpresentational State Transfer Application Program Interface) and direct download. In addition, data websites should be routinely updated and optimised for crawlability. To achieve effective crawlability, the Guidelines recommend data portals to include: well-structured sitemaps, consistent URL naming, proper security certifications, a permissive robots.txt file, HTML format for publications and implementation of APIs that adhere to REST principles. These principles are: uniform interface, stateless, cacheable, client-server, layered system and code on demand.
Available data infrastructures	As a guide and best practices framework, the Guidelines document does not outline any specific data infrastructure. It does however present policy and technical recommendations for data infrastructures to properly support generative AI and open data use. Improvement of navigation and retrieval of open data in order to shape the accessibility and usability of data across diverse application domains. Recommendations applicable to foster consistent and accessible storage and dissemination for stored downloadable open datasets, easily sourceable data via hyperlink within a generated response, retrieval of specific data values for a generated response. Use of RESTful APIs to enable data scientists and developers to programmatically retrieve data. For machine-readable API documentation, encouragement to align with the Data Service class in DCAT-US v3.0 and OpenAPI. Availability of direct download files, existing in locations that are easily parsable and predictable, and dissemination of ZIP and Apache Parquet files. Publication of datasets in accessible repositories and use widely accepted and machine-readable formats. Use of metadata standards and tools such as Croissant, compatible with frameworks for AI

	development including PyTorch, TensorFlow or JAX, to enable datasets to be loaded into ML platforms.
Effectiveness of enforcement mechanism	As a voluntary, non-binding framework of guidelines and best practices, the document does not impose any formal enforcement mechanisms. The Guidelines are primarily for use within the Department of Commerce based on the data governance structure presented above, but they are also designed as a resource for open data publishers across all levels of government, nonprofit and private sector. The Guidelines document foresees its revision and amendment based upon new information, feedback and other considerations.
Comparison to EU initiatives	The US Guidelines on Generative AI and Open Data are unique in their nature and scope, as they are primarily designed as a guidance document for the US Department of Commerce and its data ecosystem. Despite this, the Guidelines align with a growing interest shared between the EU and the US in fostering data accessibility, quality and governing frameworks for ethical AI innovation and data sharing. In 2022 the EU adopted the Data Governance Act (DGA), which aims to regulate the reuse of publicly-held protected data by boosting data sharing across sectors and EU countries⁹⁵⁷. Closer to the US Guidelines, the European Commission published the Data.europa.eu data quality guidelines, a set of recommendations aimed at data providers of the European Data Portal to support them in the preparation of data, the development of the data strategy and in ensuring data quality.⁹⁵⁸ In this sense, while the US Guidelines present recommendations tailored to the publication of open data for generative AI usage, the EU Guidelines have a broader scope focused on the availability of data sets and on ensuring the quality of the European Data Portal and its metadata. Therefore, the US Guidelines and Best Practices, despite being designed for a specific data ecosystem, pose an opportunity for the design of a framework that not only guarantees data quality but also ensures the availability of open data for generative AI usage.

A4.15 Preventing Access to US Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons Rule | USA

Preventing Access to US Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons Rule USA	
Summary	Final rule of the US Department of Justice to implement Executive Order 14117 of February, 2024 (Preventing Access to Americans' Bulk Sensitive Personal Data and United States Government-Related Data by Countries of Concern). ⁹⁵⁹ The Rule prohibits and restricts US persons from engaging in certain data transactions with specific foreign countries deemed as countries of concern and their nationals. Companies engaging with foreign data recipients in jurisdictions different from the identified countries of concern, are required to implement contractual terms prohibiting any onward transfers towards the latter. The list of countries of concern is established by the Department of Justice and is subject to modifications based on evolving security assessments.
Geographic scope	United States of America. China - including Hong Kong and Macau -, Russia, Iran, North Korea, Cuba, Venezuela.
Type of measure	Law. In force since April 2025
Types of data covered	United States Government-related data, bulk US sensitive personal data including covered personal identifiers, precise geolocation data, biometric identifiers, human 'omic data, personal health data, personal financial data

⁹⁵⁷ European Commission (2024). *Data Governance Act explained*. Available online: <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained>

⁹⁵⁸ European Commission (2021). *Data.europa.eu data quality guidelines*. Available online: <https://op.europa.eu/en/publication-detail/-/publication/b601d9cc-b3c0-11ec-9d96-01aa75ed71a1>

⁹⁵⁹ Department of Justice (2025). *Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons*. Federal Register. Available online: <https://www.federalregister.gov/documents/2025/01/08/2024-31486/preventing-access-to-us-sensitive-personal-data-and-government-related-data-by-countries-of-concern>

The information presented hereafter refers to the latest version of the Preventing Access to US Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons rule unless otherwise referenced.

Relevant sectors	All sectors operating with US government-related data or bulk US sensitive personal data. Public sector. Private sector.
Objectives	The main objectives of the Rule are: Prevent access to sensitive personal data and government-related data by countries of concern or covered persons Establish a list of countries of concern and specific thresholds to define bulk data Mitigate risks to national security related to surveillance, espionage and the development of advanced technologies Establish prohibitions and restrictions on data transactions
Restrictions/safe-guards	Access to US government-related and sensitive personal data is restricted for the six determined countries of concern: China - including Hong Kong and Macau-, Russia, Iran, North Korea, Cuba and Venezuela. Restrictions equally apply to entities that are 50% or more owned - directly or indirectly - by a country of concern, organised or chartered under the laws of a country of concern, or have their principal place of business in a country of concern. For natural persons, restrictions apply to any foreign person that is an employee or contractor of a country of concern or covered entity, primarily a resident of a country of concern. Transactions involving access to government-related or bulk US sensitive personal data including data brokerage, vendor agreement, employment agreement or investment agreement are prohibited. Bulk US sensitive personal data covers six established categories when the amount of data transferred meets or exceeds the <i>bulk</i> threshold established for each of them in the preceding 12 months. The six established categories of bulk US sensitive personal data is: covered personal identifiers with a threshold of 100,000 US; precise geolocation data with a threshold of 1,000 US devices; biometric identifiers with a threshold of 1,000 US persons; human 'omic data with a threshold of 1,000 US persons - 100 in the case of human genomic data -; personal health data with a threshold of 10,000 US persons; and personal financial data with a threshold of 10,000 US persons.
Access control mechanisms	As a data transaction restriction rule, no specific data access control mechanisms. No distinction is made between <i>access</i> to government-related and sensitive personal data by a country of concern or covered person and <i>export</i> of such data to the same country, as the Rule argues that access registry mechanisms such as <i>data accessor attestation</i> is not reliable to prevent a country of concern from seeking access to sensitive data. Know-your-data requirements, which require US businesses engaging in restricted transactions to develop and implement data compliance programmes with risk-based procedures to verify data transactions and the end-use of data. Recordkeeping requirements, which require US businesses engaging in restricted transactions to maintain documentation of their due diligence, results of annual audits and, where relevant, the license conditions to which the US businesses or persons may be subject.
Compliance burden	The core purpose of the Rule is the reinforcement of US national security and, while it strives for balance between this and enabling legitimate commercial activities, it will have a potentially significant impact in certain sectors and activities. Companies operating in identified countries of concern will likely experience important challenges regarding shared services centres, artificial intelligence research teams, cloud or other vendors as well as any other operation that support US business operations or otherwise have access to sensitive personal or government data. Additionally, sectors including health care, life sciences, financial services, defence or government contracting and consumer industries are expected to have greater exposure due to the nature of their data and business operations. For data brokerage activities, the definition included in the Rule will force companies engaging with foreign recipients in jurisdictions different from the countries of concern to implement contractual terms prohibiting any onward transfers towards the latter.⁹⁶⁰ In its <i>Discussion of Comments</i> section, the Rule further outlines concerns posed by commenters on its effects towards the US international promotion of trusted cross-border data flows, research activities, specific sectors and industries or the possibility to allow data transfers upon individual consent.⁹⁶¹ Additionally, the Rule acknowledges the existence of compliance burdens raised by some commenters regarding the time and resources needed.
Enforcement	The National Security Division (NSD) of the US Department of Justice is the appointed body for the

⁹⁶⁰ Hengesbaugh, Brian and Janet K. Kim (2025). *What does the DOJ final rule on protecting American's sensitive data from foreign adversaries really mean for global business?*. Baker McKenzie. Available online: <https://sanctionsnews.bakermckenzie.com/what-does-the-doj-final-rule-on-protecting-americans-sensitive-data-from-foreign-adversaries-really-mean-for-global-business/>

⁹⁶¹ The comments included in the Rule are based on the feedback provided by a list of commenters including associations, research institutes, councils, companies and experts. For the full list, see: III. Rulemaking Process. Available online: <https://www.federalregister.gov/documents/2025/01/08/2024-31486/preventing-access-to-us-sensitive-personal-data-and-government-related-data-by-countries-of-concern>

mechanism	implementation and enforcement of the Rule. Both the Rule and the International Emergency Economic Powers Act (IEEPA) authorise NSD to impose civil enforcement actions and criminal prosecutions for violations of the Rule.⁹⁶² Unlawful acts under IEEPA are subject to civil penalties of up to \$368,136 or twice the value of each violation. Wilful violations are punishable by up to \$1 million fine, or if a natural person, by imprisonment of up to 20 years or both. Companies or natural persons engaging in any restricted transaction are required to develop and implement a data compliance programme, conduct audits, keep full and accurate record of each transaction and present reports as required by the Department of Justice. Annual reports are equally required for restricted transactions involving cloud-computing services as well as reports on rejected prohibited transactions within 14 days from the rejection. The Rule includes the right to request an advisory opinion from the Attorney General with respect to any transaction that may be subject to the prohibitions or restrictions. Advisory opinions are subject to amendment or revocation after being issued.
Effectiveness of enforcement mechanism	Compliance with due diligence, audit and reporting requirements for restricted transactions are to take effect on October 6, 2025. Due to the early implementation phase of the Rule, information regarding compliance and enforcement has not yet been released. The main gaps identified are related to the restriction of a wide range of transactions commonly present in corporate operations including engaging vendors, hiring foreign nationals and seeking non-passive investors when there is a link to any country of concern.⁹⁶³ Comments included in the Rule further point to concerns among specific sectors and industries on the time and resources needed to effectively implement and enforce the Rule.
Alignment with international frameworks	The Preventing Access to US Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons Rule responds to specific strategic considerations and a prioritisation of national security over cross-border data flows. In particular, the Rule is based on the perceived pressing risks posed by the countries of concern attempts to collect and exploit both government and bulk sensitive personal data to the detriment of the US. While the Rule aligns with major international frameworks in its aim to protect sensitive personal data, it diverges in its rather restrictive approach towards international data transfers and the targeting of specific third countries. As noted above, the prohibition of certain data transfers to the determined countries of concern collides with international standards such as the Data Free Flow with Trust (DFFT) concept promoted by Japan and the OECD. Of particular interest will be the fit of the Rule to the WTO principles, and specifically to the General Agreement on Trade in Services (GATS). The GATS includes provisions on market access safeguards and against discriminatory measures to third countries, while recognising the right of WTO members to apply exceptions in pursuit of their own policy objectives⁹⁶⁴.
Comparison to EU initiatives	Both the Rule and EU initiatives on data protection such as the GDPR share common features regarding the protection and safeguard of sensitive personal information, data sovereignty as well as the adoption of detailed definitions and categorisations of what types of data can be considered as sensitive personal information. However, the two approaches differ on important aspects. The Rule places sensitive personal data under the same considerations as government-related data, effectively classifying data as a national security asset. Under this approach, data protection blends with national security, strategic and political considerations to restrict data transfers to countries of concern for the US. For its part, the EU GDPR focuses on imposing obligations onto organisations, regardless of their location, targeting or collecting data from EU citizens.⁹⁶⁵ The EU approach therefore follows a more structured procedure based on consent, safeguards and the protection of its citizen's rights. Thus, the comparison presents both a legal and strategic divergence with significant implications for the EU. The restrictions on data brokerage activities already affect EU companies acting as <i>foreign</i>

⁹⁶² Department of Justice (2025). *Data Security Program: Implementation and Enforcement Policy through July 8, 2025*. Available online: <https://www.justice.gov/opa/media/1396346/dl?inline>

⁹⁶³ Ellis, Abram (2025). *New Regulations Restrict Access to US Sensitive Personal Data by China and Other Countries of Concern*. Simpson Tacher. Available online: <https://www.stblaw.com/about-us/publications/view/2025/04/08/new-regulations-restrict-access-to-u.s.-sensitive-personal-data-by-china-and-other-countries-of-concern>

⁹⁶⁴ World Trade Organisation. *The General Agreement on Trade in Services (GATS): objectives, coverage and disciplines*. Available online: https://www.wto.org/english/tratop_e/serv_e/gatsqa_e.htm

⁹⁶⁵ European Parliament and the Council (2016). *Regulation (EU) 2016/679 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)*. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>

	<i>recipients</i> , which must comply with the contractual terms prohibiting any onward transfers towards the countries of concern. Additionally, the EU is presented with a favourable context for strengthening personal data protection and reinforcing its understanding as a truly European asset. However, different approaches can be considered, either focusing on an increasing data securitisation based on strategic considerations, or favouring a system of security standards and heavy scrutiny.

A4.16 Securing the Information and Communications Technology and Services Supply Chain Rule | USA

Securing the Information and Communications Technology and Services Supply Chain Rule USA	
Summary	Final rule of the US Department of Commerce to implement Executive Order 13873 of May, 2019 (Securing the Information and Communications Technology and Services Supply Chain). ⁹⁶⁶ The Rule enables the Department of Commerce's Bureau of Industry and Security (BIS) to effectively review and prohibit transactions involving Information and Communications Technology and Services (ICTS) linked to a <i>foreign adversary</i> that may harm US national security. It equally formalises the operations of the Office of Information and Communications Technology and Services (OICTS) within the BIS. The Rule formalises the federal framework to protect ICTS supply chain together with previous Final Rules and Determinations issued by the Department of Commerce. Specific rules apply to Connected Vehicles (15 CFR Part 791, Subpart C) with the aim of mitigating national security risks posed by connected vehicle systems vulnerable to data filtration and cyberattacks. ⁹⁶⁷
Geographic scope	United States of America. China - including Hong Kong and Macau -, Cuba, Iran, North Korea, Russia, Venezuela
Type of measure	Law. In force since 2021 (as Interim Final Rule 86 FR 4909) - amended in 2023 and 2025
Types of data covered	Sensitive personal data, sensitive business or government data, confidential business or government data, data related to the sectors listed below
Relevant sectors	Information and communications technology. Automotive sector, related Vehicle Connectivity System (VCS) hardware and software, connected vehicles. Sectors considered as critical infrastructure including chemical, commercial facilities, communications, critical manufacturing, dams, defence industrial base, emergency services, energy, financial services, food and agriculture, government services and facilities, health care and public health, information technology, nuclear reactors, materials, and waste, transportation systems, and water and wastewater systems sectors. Sectors related to critical and emerging technologies including highly automated, autonomous and uncrewed systems and robotics, integrated communication and networking technologies, quantum information and enabling technologies, semiconductors, microelectronics and biotechnology.
Objectives	The main objectives of the Rule are: Protect US ICTS supply chains from potential national security threats linked to designated foreign adversaries Prevent the exploitation of vulnerabilities of the US ICTS infrastructure Establish a definition of covered transaction involving ICTS products or services linked to foreign adversaries Mitigate risks to national security related to surveillance, espionage and the development of advanced technologies Prohibit and restrict certain transactions upon restructured review
Restrictions/safe-guards	Covered transactions involving ICTS designed, developed, manufactured, or supplied by persons owned by, controlled by or subject to the jurisdiction or direction of a foreign adversary are restricted. The foreign adversaries designated countries are: China - including Hong Kong and Macau -, Russia,

⁹⁶⁶ Department of Commerce (2024). *Securing the Information and Communications Technology and Services Supply Chain*. Federal Register. Available online: <https://www.federalregister.gov/documents/2024/12/06/2024-28335/securing-the-information-and-communications-technology-and-services-supply-chain>

The information presented hereafter refers to the latest version of the Securing the Information and Communications Technology and Services Supply Chain rule unless otherwise referenced.

⁹⁶⁷ Department of Commerce (2024). *Securing the Information and Communications Technology and Services Supply Chain: Connected Vehicles*. Federal Register. Available online: <https://www.federalregister.gov/documents/2025/01/16/2025-00592/securing-the-information-and-communications-technology-and-services-supply-chain-connected-vehicles>

	Iran, North Korea, Cuba and Venezuela. Restrictions equally apply to entities owned or controlled by, operating, headquartered in or organised under the laws of a foreign adversary. Restricted ICTS transactions include those from the sectors and technologies outlined above when posing an undue or unacceptable risk for US national security, based on threats to the execution of a <i>national critical function</i> or on an unusual and extraordinary harm to national security.
Access control mechanisms	As a ICTS transaction restriction rule, no specific data access control mechanisms. Focus on preventing the engagement of foreign adversaries-related entities or persons in ICTS transactions with US persons or entities
Compliance burden	The key purpose of this rule is the reinforcement of US national security in such sensitive areas as ICTS and new technologies and, while the Department of Commerce has sought to ensure a balance between commercial activity and national security, the Rule prioritises the latter in the covered sectors. In this sense, the main concerns posed by commenters on the Rule are related to overly broad definitions, the scope of covered ICTS transactions and foreign adversary determinations. In 2021 the Department of Commerce already issued a Regulatory Impact Analysis (RIA) of the Rule, which estimated that around 268,000 companies would be directly impacted. In terms of costs, the RIA estimated a compliance cost between \$238 million and \$20 billion, depending on different factors such as number of transactions reviewed, the actual compliance rate of businesses and the number of transactions prohibited or restricted as a result of the rule. The RIA also outlined a list of 35 impacted industries, including Communications Equipment Manufacturing, Navigational, Measuring, Electromedical and Control Instruments Manufacturing as well as Software Publishers.⁹⁶⁸ Moreover, further sectors such as automotive will potentially experience larger compliance costs. Relevant firms such as Geely-owned Volvo⁹⁶⁹ will need to either exit the US market or undergo major corporate restructuring. Others such as Ford or GM will likely need to stop exporting vehicles from China to the US.⁹⁷⁰ This entails the closure of the US market to China, one of the global leading automotive competitors, and the need for restructuring of those business models with strong linkages to countries deemed as foreign adversaries. It is expected that, once the automotive sector has restructured its US supply chain, Commerce will then further refine restrictions based on the results and information collected.⁹⁷¹
Enforcement mechanism	The OICTS of the Department of Commerce's Bureau of Industry and Security is the appointed body for the implementation and enforcement of the Rule. Unlawful acts are subject to civil penalties of up to \$250,000 or twice the value of each violation in accordance with the International Emergency Economic Power Act (IEEPA). Willingly violations are punishable by up to \$1 million fine, or if a natural person, by imprisonment of up to 20 years or both. The Secretary of Commerce, at the Secretary's discretion, may initiate the review of an ICTS transaction. Upon review and notification to the appropriate agency heads, the Secretary must issue an initial determination on whether the covered ICTS transaction meets the criteria or not. In case the covered ICTS transaction meets the criteria, the initial determination shall include the prohibition or the imposition of mitigation measures by which the transaction may be permitted. A party from a transaction affected by an initial determination has the right to request a rescission or mitigation of the latter. For each initial determination, the Secretary must issue a final determination as to whether the covered ICTS transaction is prohibited, permitted or permitted under mitigation measures. Consultation with and concurrence from all appropriate agencies is required. In the case of a transaction being permitted under mitigation measures, such measures must be outlined together with the penalties for non-compliance. If applicable, the final determination must include the transition conditions from a mitigation agreement to a prohibition. Any entity or natural person involved in a covered ICTS transaction and notified about an initial determination must retain all records relating to such transaction for a period of 10 years.

⁹⁶⁸ Department of Commerce (2021). *Securing Information and Communications Technology and Services Supply Chain (RIN 065-AA51) Regulatory Impact Analysis & Final Regulatory Flexibility Analysis*. Available online: https://downloads.regulations.gov/DOC-2019-0005-0108/attachment_1.pdf?utm

⁹⁶⁹ The Swedish-based Volvo Cars firm is currently owned by the Chinese automotive company Geely.

⁹⁷⁰ Goujon, Reva and Gregor Sebastian (2024). *Car Trouble: ICTS Rule Rewires Global Auto Supply Chains*. Rhodium Group. Available online: <https://rhg.com/research/car-trouble-icts-rule-rewires-global-auto-supply-chains/?utm>

⁹⁷¹ Belur, Anjali et al. (2025). *Update on ICTS Rules: Recent Actions and Path Ahead in 2025*. Steptoe. Available online: <https://www.steptoel.com/en/news-publications/international-compliance-blog/update-on-icts-rules-recent-actions-and-path-ahead-in-2025.html?utm>

Effectiveness of enforcement mechanism	Several subpoenas served by the Secretary of Commerce as a preliminary step for the review of an ICTS transaction. The target of these subpoenas are Chinese companies providing ICTS services or products in the United States, although there is no information on the services or products under review as the Department of Commerce does not publicly identify what services or products are being considered for investigation.⁹⁷² No determination has been issued yet regarding these subpoenas. To date, the Department of Commerce has issued only one Final Determination: Case No. ICTS-2021-002, Kaspersky Lab, Inc. As a result, the Russian-based cybersecurity company Kaspersky is prohibited from engaging in ICTS transactions involving any cybersecurity product or service, any antivirus software and the integration of software. Thus, Kaspersky is prohibited from providing any antivirus signature updates and codebase updates and from operating KSN - Kaspersky Security Network - in the US or on any US person's information technology system.⁹⁷³ The process lasted nearly 3 years, from the initial referral from the Department of Justice to Commerce in 2021 to the issuing of the Final Determination in 2024. In addition to the regulation on connected vehicles and related ICTS, Commerce aims at regulating Unmanned Aircraft Systems (UAS), i.e. drones, under the ICTS rules framework. This proposed rulemaking aligns with a number of contract-related restrictions banning UAS manufactured totally or partially in China from being operated or procured by the US Department of Defence.⁹⁷⁴
Alignment with international frameworks	The legal framework comprising the ICTS rules respond to a clear prioritisation of US strategic considerations and national security over cross-border ICTS flows. Furthermore, the US domestic political context favours such trade restrictions on the basis of national industry support, prioritisation of products <i>Made in USA</i> in the internal market and international trade competition, especially with China. The restriction of transactions involving ICTS based on their country of origin directly collides with international principles such as the Data Free Flow with Trust (DFFT) concept promoted by the OECD and Japan. The ICTS framework could also potentially collide with the WTO principles and the General Agreement on Trade in Services (GATS). The GATS includes provisions on market access safeguards and against discriminatory measures to third countries, which have a difficult fit with the restrictions to ICTS transactions based on product-origin criteria.⁹⁷⁵
Comparison to EU initiatives	Both the EU and the US are converging on an increased interest in securing their critical technologies and infrastructures, communication systems and ultimately their data. Likewise, at both sides of the Atlantic there is a great interest in strengthening technological, data and infrastructure sovereignty. Despite this, both the EU and the US differ in their strategies. In 2023, the EU Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS2) entered into force. The NIS2 requires both EU and non-EU <i>essential entities</i> or <i>important entities</i> providing services in the EU to comply with a series of requirements on basic security measures including supply chain security, incident reporting, disclosure and report of vulnerabilities and registration.⁹⁷⁶ More specifically, the EU has developed sector-specific ICTS protection rules such as the 2022 Regulation (EU) 2022/2554 on digital operational resilience for the financial sector (DORA). DORA focuses on securing ICTS used by financial services by requiring ICT risk management frameworks, incident response and reporting procedures, information sharing requirements and more importantly, oversight of ICT risks related to third-party service providers.⁹⁷⁷ While the comparison of initiatives reflects a clear interest in regulating ICTS in sectors deemed critical, both parties diverge in the use of indiscriminate restrictions based on the ICTS place of origin. The EU is currently prioritising the imposition of industry standards and foreign investment scrutiny to

⁹⁷² US Department of Commerce (2021). *Secretary of Commerce Gina Raimondo Statement on Actions Taken Under ICTS Supply Chain Executive Order*. Available online: <https://www.commerce.gov/news/press-releases/2021/03/us-secretary-commerce-gina-raimondo-statement-actions-taken-under-icts>

⁹⁷³ Bureau of Industry and Security (2024). *Final Determination: Case No. ICTS-2021-002, Kaspersky Lab, Inc.* Federal Register. Available online: <https://www.federalregister.gov/documents/2024/06/24/2024-13532/final-determination-case-no-icts-2021-002-kaspersky-lab-inc>

⁹⁷⁴ Belur, Anjali et al. (2025). *Update on ICTS Rules: Recent Actions and Path Ahead in 2025*. Steptoe. Available online: <https://www.steptoe.com/en/news-publications/international-compliance-blog/update-on-icts-rules-recent-actions-and-path-ahead-in-2025.html?utm>

⁹⁷⁵ World Trade Organisation. *The General Agreement on Trade in Services (GATS): objectives, coverage and disciplines*. Available online: https://www.wto.org/english/tratop_e/serv_e/gatsqa_e.htm

⁹⁷⁶ European Parliament and the Council (2022). *Directive (EU) 2022/2555 on measures for a high common level of cybersecurity across the Union (NIS 2 Directive)*. Available online: <https://eur-lex.europa.eu/eli/dir/2022/2555/oj/eng>

⁹⁷⁷ European Parliament and the Council (2022). *Regulation (EU) 2022/2554 on digital operational resilience for the financial sector*. Available online: <https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A32022R2554>

	secure its ICTS supply chain regardless of the place of origin. The US, on the contrary, has implemented a restrictive approach on the basis of a strategic assessment on its perceived <i>foreign adversaries</i>. Should the US continue on this path, the EU will face important challenges to ensure that its ICTS exports - mainly the ones related to the automotive sector - to the US are not affected by this Rule. Additionally, this presents a larger debate over whether the EU is willing to align with US strategy or is capable of defending and implementing its own model that prioritises industrial standards and general scrutiny over targeted restrictions.

A4.17 APEC Cross Border Privacy Rules system – APEC CBPRs | Multilateral (APEC)

APEC Cross Border Privacy Rules system – APÊC CBPRs Multilateral (APEC)	
Summary	The APEC⁹⁷⁸ Cross-Border Privacy Rules system (CBPRs), introduced in 2011, is a voluntary framework designed to support data flows that respect privacy among APEC member economies. Under this system, companies' privacy policies and practices are evaluated and certified by independent third parties, known as Accountability Agents, to ensure they align with a shared set of standards derived from the APEC Privacy Framework. By establishing this common baseline, the CBPR system helps overcome variations in national privacy regulations across the region.^{979, 980} The APEC CBPR system is only applicable to personal information controllers. In 2015 APEC endorsed a parallel system for data processors, namely the APEC Privacy Recognition for Processors system (PRP). The PRP helps personal information processors assist data controllers to comply with applicable privacy obligations, and also facilitates for controllers to identify qualified and accountable processors.⁹⁸⁰ The APEC CBPRs has four main components:⁹⁷⁹ Recognition criteria for organisations wishing to become an APEC CBPR System certified Accountability Agent Intake questionnaire for organisations that wish to be certified as APEC CBPR System compliant by a certified Accountability Agent Assessment criteria for use by certified Accountability Agents when reviewing an organisation's answers to the intake questionnaire A regulatory cooperative arrangement (the CPEA- Cross Border Privacy Enforcement Arrangement) to ensure that each of the APEC CBPR system program requirements can be enforced by participating APEC economies. The Joint Oversight Panel (JOP) administers the APEC CBPR system. Decisions about an organisation's eligibility to be an Accountability Agent are made by APEC economies.⁹⁷⁹ The <i>Global CBPR Forum</i> (launched 2022) supports an expansion of the existing APEC CBPR System, allowing more economies from around the world to adhere to standard for cross-border data protection, the Global CBPR and PRP framework.⁹⁸¹ Currently the two systems are working in parallel, sharing the same basis, but it is predicted that the APEC CBPR and PRP systems will be obsolete and that the Global CBPR Forum will continue.⁹⁸²
Geographic scope	Currently nine economies are participating in the APEC CBPR system: Australia, Canada, Chinese Taipei, Japan, Mexico, Singapore, the Philippines, the Republic of Korea, and USA. With respect to APEC PRP, only two countries have joined so far, USA and Singapore.^{979, 980} Apart from the 9 APEC CBPR member countries, the Global CBPR Forum has 4 Associated jurisdictions, namely Bermuda, Dubai International Financial Centre, Mauritius, and The United Kingdom.⁹⁸¹
Type of measure	Multilateral voluntary framework agreement to support cross-border data flows that respect privacy.

⁹⁷⁸ APEC = Asia-Pacific Economic Cooperation

⁹⁷⁹ The APEC Cross-Border Privacy Rules (CBPR) system website, <https://cbprs.org/>

⁹⁸⁰ APEC website, <https://www.apec.org/groups/committee-on-trade-and-investment/digital-economy-steering-group>

⁹⁸¹ Global CBPR Website, <https://www.globalcbpr.org/>

⁹⁸² Digital Trade Alliance (June 2023), *APEC's Cross Border Data Transfer Rules: An Unfulfilled Potential, An Uncertain Future*, available at: <https://dtalliance.org/wp-content/uploads/2023/10/APEC-Cross-Border-Data-Transfer-Rules.pdf>

	The APEC CBPR system was endorsed 2011 and the APEC PRP system was recognised in 2015. ⁹⁷⁹
Types of data covered	Personal data. Cross-border data, respecting personal privacy.
Relevant sectors	The APEC CBPRs applies to all sectors and industries that process personal data. There is no sector-specific focus in CBPR – the 50 program requirements cover generic privacy principles applicable to any business. Any organisation (in a participating economy) handling personal information may seek certification. In practice, adoption has spread across technology, finance, healthcare, manufacturing, and services. For example, officials report that firms in the IT, health, and manufacturing sectors are among the current participants. ^{979, 983}
Objectives	The primary goal of the APEC CBPR system is to enable privacy-respecting cross-border data flows while raising privacy standards in APEC economies. The system provides “enforceable standards” so differences in domestic laws do not block digital trade.⁹⁷⁹ It is explicitly intended to “<i>enhance electronic performance, facilitate trade and economic growth, and strengthen consumer privacy protections</i>” through baseline privacy requirements. In short, CBPR balances data protection with the free flow of information to benefit both businesses and consumers.⁹⁸⁴ Essentially, the <i>Global CBPR Forum</i> seeks to extend APEC CBPR’s principles and cooperation mechanism beyond Asia-Pacific, fostering interoperability across different privacy regimes.⁹⁸¹
Legal or regulatory provisions	The APEC CBPRs is a voluntary, accountability-based framework of rules built on the APEC Privacy Framework’s nine principles (Preventing Harm; Notice; Collection Limitation; Use of Personal Information; Choice; Integrity of Personal Information; Security Safeguards; Access and Correction; Accountability).^{979, 985} The CBPR system establishes baseline privacy protection standards. To join the system, an APEC economy may need to adjust its national laws or procedures. However, the CBPR does not override or change existing domestic laws. If a country’s laws offer stronger privacy protections than those required by CBPR, those stricter national laws will still apply in full.⁹⁸⁶
Restrictions/safe-guards	According to the APEC BPRs’ Program Requirements, the Accountability Agent must verify the existence of the physical, technical and administrative safeguards the Applicant declares to be used to protect personal information. Such safeguards may include: Authentication and access control (e.g. password protections); Encryption; Boundary protection (e.g. firewalls, intrusion detection); Audit logging; and Monitoring (e.g. external and in internal audits, vulnerability scans). ⁹⁸⁷
Access control mechanisms	APEC CBPRs certified companies should allow individuals to access and correct their personal data, as indicated by the APEC Privacy Framework’s Access and Correction principle. While the ability to change and access personal information is generally regarded as a central aspect of privacy protection, it is not an absolute right. ^{984, 988}
Compliance burden	The APEC CBPRs is aligned with significant efforts related to compliance. Participating organisations must undergo certification through a process involving a detailed self-assessment against 50 program requirements, covering the 9 principles of the APEC Privacy Framework. In a second step of the certification process the applicant must go through an audit by a CBPR-recognised Accountability Agent.⁹⁸⁴ According to an article from the Digital Trade alliance, cost is a significant issue for organisations considering APEC CBPRs participation. In Singapore, the application fee by itself is about €355 (400 USD), in addition to the assessment fee charged by one of the seven assessment bodies authorised by the Accountability Agent. The overall assessment fee varies depending on the size of the organisation and other relevant factors. In the USA, one Accountability Agent estimates that the total cost for certification typically ranges from €13,300 to €35,500 (15,000 to 40,000 USD).
Enforcement mechanism	The APEC CBPRs relies on national privacy authorities for enforcement, under the regional cooperative agreement called APEC CPEA – APEC Cross-Border Privacy Enforcement Arrangement. Each participating economy must have a Privacy Enforcement Authority (PE Authority)

⁹⁸³ APEC (September 2018), *The Case for the APEC Cross-Border Privacy Rules System*, available at: https://www.apec.org/press/features/2018/0919_cbpr

⁹⁸⁴ NCC Group (July 2023), *What is the APEC CBPR?*, available at: <https://www.nccgroup.com/us/what-is-the-apec-cbpr>

⁹⁸⁵ APEC (2005), APEC Privacy Framework, available at: https://www.apec.org/docs/default-source/Publications/2005/12/APEC-Privacy-Framework/05_ecsg_privacyframewk.pdf

⁹⁸⁶ Sidley (November 2022), *The Privacy, Data Protection and Cybersecurity Law Review*, Chapter 3: CBPR AND APEC OVERVIEW, available at: https://www.sidley.com/-/media/files/publications/2014/11/the-privacy-data-protection-and-cybersecurity-la_/files/apec-overview/fileattachment/apec-overview.pdf?la=en

⁹⁸⁷ APEC CBPRs (November 2019), *APEC Cross Border Privacy Rules System Program Requirements*, <https://cbprs.org/wp-content/uploads/2019/11/5.-Cross-Border-Privacy-Rules-Program-Requirements-updated-17-09-2019.pdf>

	with the power to enforce CBPR requirements. When an APEC CBPR–certified company violates its privacy commitments, the local PE Authority can investigate and impose remedies (for example, under general consumer protection or privacy laws). ⁹⁸⁸
Effectiveness of enforcement mechanism	Desk research resulted in little information related to the effectiveness of the APEC CBPRs' enforcement mechanism. As of 2024, only nine Accountability Agents operate across five (of nine) jurisdictions. Until an Accountability Agent has been identified in an economy, companies "primarily located" in that economy cannot get certified.^{979, 984} No overall evaluation of the APEC CBPRs, with relevant enforcement metrics, was found, and only some examples of recorded events from individual Accountability Agents were identified, such as Warning letters⁹⁸⁹ (28 warnings issued 2016 by the U.S. Federal Trade Commission) and Complaint statistics⁹⁹⁰ (TRUSTe reported 147 complaints for CBPR-Certified Participants in the period 3/01/19-2/29/20).
Alignment with international frameworks	The APEC CBPRs is broadly aligned with the OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, particularly regarding: Accountability, Purpose limitation, Security safeguards, Individual participation, and Cross-border cooperation. It is based on the same principles that inform the Fair Information Practice Principles. ⁹⁹¹
Comparison to EU initiatives & Outlook for the future	There are broad similarities between the APEC CBPRs and cross border transfer mechanisms under the EU GDPR. Under the GDPR there are a range of voluntary options for corporations that are not considered to have domestic privacy and data protection requirements equivalent to the EU, e.g. the use of standard contractual clauses, approved Codes of Conduct, and certification.⁹⁹² The EU GDPR is a mandatory regulation with legally binding compliance for all organisations processing (both data controllers and processors) EU personal data. The APEC CBPRs, on the other hand, builds on voluntary certifications for data exchange among organisations in the participating APEC economies. These certifications are enforceable by privacy authorities in participating jurisdictions. The APEC CBPRs allows the transfer of data not only from one jurisdiction to another, but also subsequent transfers, as long as all entities involved are APEC CBPR certified. On the contrary, the GDPR certification system focuses only on transfers from a country within the EEA to a third country, and prohibits any further transfer.⁹⁸² Various pros and cons with the APEC CBPR and PRP systems are discussed in a report from the Digital Trade Alliance (June 2023). In the concluding part of the downsides of the system it is stated: <i>"Unlike the GDPR, which is often regarded as the standard against which all other data protection laws are judged, the APEC Privacy Framework has long been the subject of criticisms for being ineffective, as standards go. If the Framework itself is unable to prove its long-term relevance, the APEC Systems [referring to both the CBPR system and the PRP system] may gradually be discarded by its current participants and ignored by the rest. Two other major developments make the road ahead less optimistic: (1) China's recent decision to develop and enforce its own cross border privacy rules; and (2) the launch of the Global CBPR Forum."</i>⁹⁸² Linked to the above, a recent article exploring key takeaways from a Global CBPR Forum meeting in Singapore (May 26 -27, 2025), provides an update with respect to the progress of the Forum. While using the same base, the Global CBPR system is described as "a significant evolution of the APEC CBPR system" with material updates that will be crucial for international recognition and uptake. Agreed additions to the Global CBPR assessment include: Breach notification to individuals; Sensitive data; Direct marketing choice; Privacy personnel; Risk assessment; Withdrawal of consent; Records of processing and consent; and, Children's data. Regarding the possibility for expansion the Forum highlighted strong interest from existing members as well as additional, non-APEC jurisdictions. The Forum also stressed leveraging of technology and AI, e.g. for automated tools to

⁹⁸⁸ APEC (February 2024), *APEC Cross-border Privacy Enforcement Arrangement (CPEA)*, available at: <https://www.apec.org/groups/committee-on-trade-and-investment/digital-economy-steering-group/cross-border-privacy-enforcement-arrangement>; and https://www.apec.org/docs/default-source/groups/ecsg/cbpr/cpea-2019.pdf?sfvrsn=f1643e20_1

⁹⁸⁹ <https://www.ftc.gov/news-events/news/press-releases/2016/07/ftc-issues-warning-letters-companies-claiming-apec-cross-border-privacy-certification>

⁹⁹⁰ <https://cbprs.org/wp-content/uploads/2020/04/APEC-CBPR-and-PRP-Complaint-Statistics.pdf>

⁹⁹¹ TrustArc (–), *5 Benefits of APEC CBPR Certification You Should Know About*, available at: <https://trustarc.com/resource/5-benefits-of-apec-cbpr-certification/>

⁹⁹² Clare Sullivan (August 2019), *EU GDPR or APEC CBPR? A comparative analysis of the approach of the EU and APEC to cross border data transfers and protection of personal data in the IoT era*, Computer Law & Security Review, Abstract, available at: <https://www.sciencedirect.com/science/article/abs/pii/S026736491930038X>

	support compliance. The host of the Forum, IMDA, also announced initiatives to support SMEs with certification costs, reinforcing the accessibility of the Global CBPR system. ⁹⁹³

A4.18 Global Cross-Border Data Flow Cooperation Initiative | Multilateral (proposed by China)

Global Cross-Border Data Flow Cooperation Initiative Multilateral (proposed by China)	
Summary	On 20 November 2024, the Cyberspace Administration of China (CAC) announced the Global Cross-Border Data Flow Cooperation Initiative (hereafter referred to as the GCBD initiative). ⁹⁹⁴ This is a high-level policy initiative that presents a framework to foster secure, efficient and transparent cross-border data flows. It aims to balance innovation, economic growth, and national security with the protection of privacy, intellectual property, and public interests. The initiative promotes openness, inclusiveness, cooperation, non-discrimination, and mutual benefit, while emphasising the respect for national security and sovereignty, as well as regulatory diversity, so that each country can regulate data in line with its own laws. It has not been officially recognised at international level. ^{995, 996, 997}
Geographic scope	This cross-border initiative proposed by China is global in its scope, but is also advocating respect for national data regulation. ⁹⁹⁷
Type of measure	International policy framework: China's proposal for a global cooperation framework on cross-border data exchange. The initiative is promoted by China and was presented on a high-level meeting of the Asia Pacific Economic Cooperation (APEC). The initiative comes as a response to U.S. restrictions on China and an effort to position China as a promoter of digital free trade and liberal data flows. However, even if it can be seen as a step in the right direction, concerns remain over the domestic data governance approach in China, including e.g., restrictions on data flows and government access to data. ⁹⁹⁷
Types of data covered	All types of cross-border data flows. Including both personal and non-personal data. ⁹⁹⁴
Data accessibility	No specific information found about data accessibility.
Relevant sectors	Covers all sectors. Although no specific sectors are targeted, cross-border data flows are judged by the study team to be particularly relevant for sectors such as, finance (cross-border transactions and global operations), automotive (international supply chains and R&D structures), pharmaceuticals (global research, clinical trials etc.), ICT/E-commerce (platform operations, customer data and logistics), retail and logistics (inventory, real-time tracking data, and cross-border transactions).
Objectives	The GCBD initiative: ⁹⁹⁴ Supports open, inclusive, and non-discriminatory principles in global cross-border data flows, aiming to build a mutually beneficial and cooperative international landscape. Seeks to advance secure and efficient cross-border data flows, balancing technological innovation, economic development, and national security with the protection of privacy, intellectual property, and public interests. Calls for constructive international cooperation on cross-border data governance, regulatory alignment, and the development of global standards. Emphasises respect for each country's data regulations and data security measures. Aims to facilitate digital trade, e-commerce, and the integration of digital technology.

⁹⁹³ Hogan Lovells (May 2025), *The Global Cross Border Privacy Rules – A new paradigm in data protection*, available at: <https://www.hoganlovells.com/en/publications/the-global-cross-border-privacy-rules-a-new-paradigm-in-data-protection>

⁹⁹⁴ Cyberspace Administration of China (November 2024), *Global Cross-Border Data Flow Cooperation Initiative*, available at: https://www.cac.gov.cn/2024-11/20/c_1733706018163028.htm

⁹⁹⁵ Digital Policy Alert (--), *China: Announced Global Cooperation Initiative on Cross-Border Data Flows*, available at: <https://digitalpolicyalert.org/event/24557-proposed-global-cross-border-data-flow-cooperation-initiative-at-apec-2024-including-cross-border-data-transfer-regulation?>

⁹⁹⁶ Science and Technology Daily (December 2024), *Global Cross-Border Data Flow Cooperation Initiative Released*, available at: https://www.stdaily.com/web/English/2024-12/13/content_267412.html

⁹⁹⁷ Crowell (January 2025), *China Unveils New Framework To Stimulate Cross-Border Data Flows: Risk or Opportunity for Multinational Companies*, available at: <https://www.crowell.com/en/insights/client-alerts/china-unveils-new-framework-to-stimulate-cross-border-data-flows-risk-or-opportunity-for-multinational-companies>

	Commits to offering technical support and capacity building, especially for developing countries, to promote balanced global digital infrastructure and data connectivity. Explicitly opposes discriminatory practices and the over-generalisation of data security concerns to restrict data flows targeting specific countries or companies without factual evidence.
Legal or regulatory provisions	While promoting a commonly agreed framework and international standardisation (no details are specified) for cross-border data transfers the GCBD initiative declares that each country need to be able to regulate data in line with its own laws. ⁹⁹⁴
Restrictions/safe-guards	As mentioned above, this is a high-level initiative, with no or limited details on “operational level”. Nevertheless, a few general data security safeguards, mirrored from China’s data regulation framework, are identified. Countries are encouraged to provide convenient and well protected channels for cross-border transmission of personal information. Enterprises should be encouraged to obtain personal information protection (PIP) certifications to demonstrate their compliance with PIP standards for secure cross-border flows. The initiative supports exploring the establishment of negative lists for efficient and safe management of cross-border data flows.
Data governance structure	Data governance structures are not addressed explicitly in the initiative (it is a high-level policy). The initiative suggests a soft governance framework based on a set of commonly agreed rules and agreements, and calls for cooperation among governments, businesses, civil society and international bodies. To increase regulatory alignment, the initiative promotes greater interoperability, with mutual recognition of standards, technology regulations, and certification procedures. The text explicitly invites countries to respond “through bilateral, multilateral or regional agreements and arrangements”, meaning that China expects states to incorporate these principles into their own trade or cooperation pacts. ⁹⁹⁴
Data localisation requirements	Data localisation is not directly mentioned in the GCBD initiative. However, as it advocates respect for each country’s data regulations and data security measures, data localisation is indirectly implied. Chinese laws, including the Cybersecurity Law and the Data Security Law, plus various sectoral laws and regulations, indeed require firms to store certain types of data within Chinese borders. An article from Crowell refers to these data localisation requirements and to the extensive powers that Chinese authorities have over data as two perspectives illustrating the contrast between China’s rhetoric of “free and secure flows” and its strict domestic controls. ⁹⁹⁷
Available data infrastructures	No specific information found about data infrastructures (it is a high-level policy).
Effectiveness of enforcement mechanism	Not addressed; not adopted so far.
EU’s position to the initiatives	In August 2024 the EU and China set up a <i>Cross-Border Data Flow Communication Mechanism</i> to address non-personal data-transfer issues for European companies, aiming to resolve specific business obstacles. In particular the uncertainty and administrative burden caused by China’s restrictive and ambiguously defined data export rules, e.g., the vague definition of “important data”. The EU’s approach is to maintain dialogue and seek regulatory clarity, rather than to fully align with China’s sovereignty-focused data governance mode. ⁹⁹⁸ Regarding the GCBD initiative, presented in November 2024, no official response from the EU have been identified, indicating a cautious position from Europe. As indicated above the GCBD initiative sends mixed signals. Chinese officials portray it as a constructive vision for global data governance, and some international observers welcome the dialogue while others worry about execution and intent, seeing it is part of China’s consistent effort to rewrite the rules of global data governance in its own terms. Going forward, the impact will depend on whether countries and organisations take up the proposals – through new agreements, regulatory adjustments, or technical cooperation – or continue to rely on parallel rulemaking. ^{997, 999}

⁹⁹⁸ European Commission (August 2024), *EU and China launch Cross-Border Data Flow Communication Mechanism*, available at: https://policy.trade.ec.europa.eu/news/eu-and-china-launch-cross-border-data-flow-communication-mechanism-2024-08-28_en#:~:text=The%20mechanism%20will%20focus%20on,personal%20data

⁹⁹⁹ Tech Policy Press (January 2025), *China Picked Up the Ball on Data Trade Where the US Dropped It in 2024*, available at: <https://www.techpolicy.press/china-picked-up-the-ball-on-data-trade-where-the-us-dropped-it-in-2024/>

Annex 5 Task 3 Mapping of definitions related to sensitive and highly sensitive data

This mapping understands as formal definitions of sensitive and highly sensitive data those concepts included in comprehensive pieces of regulation and linked to specific categories of data and/or information.

European Union

Term	Definition	Sources
Sensitive data	Special categories of personal data (sensitive data) Personal data that requires stricter protection and processing rules. These include: <i>Racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union memberships, genetic data, biometric data, health data, sex life or sexual orientation.</i> Sensitive non-personal data: Not officially labeled or codified No comprehensive definition or codification under a single regulation. Concrete definitions, categorizations and safeguards can be found in sector-specific regulations related to trade secrets, intellectual property and information under confidentiality agreements. Commercially sensitive data is protected under the Data Act, although there is no formal definition as no specific categories of commercial information are linked to such term.	General Data Protection Regulation (2016)¹⁰⁰⁰ Data Act (2023)
Highly Sensitive Data	Not officially labelled or codified No comprehensive definition or codification under a single regulation. Categories of information that require to be treated with higher security standards due to their potential impact in the security of the EU and its Member States. These include but are not limited to: EU classified information, security and defense information, critical infrastructures and	Data Governance Act (2022) European Health Data Space Regulation (2025) Council Decision on the Security Rules for Protecting EU Classified Information (2013) Protection of European Union classified

¹⁰⁰⁰ The General Data Protection Regulation is exclusively focused on the protection of natural persons with regards to the processing of their personal data. The scope therefore excludes the protection of non-personal data. The approach given for the categorization is based on a risk-assessment evaluation of the varying likelihood and severity of the impact to the rights and freedoms of natural persons resulting from the processing of personal data.

STUDY ON THE NEXT DATA FRONTIER:
GENERATIVE AI, REGULATORY COMPLIANCE AND INTERNATIONAL DIMENSIONS (NO. 2024-020)

	sectors data, environmental and geospatial information. The Data Governance Act explicitly supports the development of future specific legislative acts to codify certain non-personal data categories as highly sensitive. Additionally, the EUCI covers EU classified information in our levels, defined by the severity of the impact of disclosure (EU Top Secret, EU Secret, EU Confidential, and EU Restricted). Depending on the tier of classified information, stricter measures need to be applied to ensure limited access to this data.	information (EUCI) (2013)

Australia

Term	Definition	Sources
Sensitive data	Personal information Categories of personal information or opinion that include sensitive details about an individual. These include: <i>Racial or ethnic origin, political opinions, membership of a political association, religious beliefs or affiliations, philosophical beliefs, membership of a professional or trade association, membership of a trade union, sexual orientation or practices, criminal record, health, genetic and biometric information, biometric templates.</i> Sensitive non-personal data: not officially labeled or codified There is no general qualification of sensitive data under Australian law. However, the Data Availability and Transparency Act does more heavily regulate sensitive data from being disclosed by the government, though the Act does not define sensitive data.	Australian Privacy Act (1988)¹⁰⁰¹ Data Availability and Transparency Act (2022)
Highly Sensitive Data	Not officially labeled or codified Categories of information subject to reinforced data protection due to their potential impact on public safety, national or personal security. These include but are not limited to: National security, classified government data, critical infrastructure, banking information, children's data.	Protective Security Policy Framework (2010) Security of Critical Infrastructure Act (2018) APRA Prudential Standards (2019) Online Safety Act (2021)

¹⁰⁰¹ The Australian Privacy Act includes the Australian Privacy Principles (APPs) under which certain entities must operate. These APP Entities are: Agencies - Australian Government and Norfolk Island Government agencies, private sector organizations including businesses with an annual turnover of more than \$3 million, small businesses under \$3 million if they handle sensitive information, provide health services or are involved in certain regulated activities.

Brazil

Term	Definition	Sources
Sensitive data	Sensitive personal data Categories of personal information that include sensitive details. These are: <i>Racial or ethnic origin, religious belief, political opinion, union membership, religious or philosophical or political organization membership, health or sex life data, genetic or biometric data.</i> Sensitive non-personal data: not officially labeled or codified Brazil does not have general categories of sensitive or highly sensitive non-personal data. The Brazilian Freedom of Information Act (LAI) does exempt disclosure of classified information, personal information, information critical to public security, ongoing investigations or administrative proceedings, or trade secrets and industrially confidential business information.	General Personal Data Protection Act (2019)¹⁰⁰² Freedom of Information Law (2011)
Highly sensitive data	Not officially labeled or codified The term "highly sensitive data" appears in some sectoral contexts (e.g., financial or medical data), however it is not specified well under the General Personal Data Law.	Resolution BCB (Banco Central Brasil) 85 (2021) Digital Health Record Law (2018) Information Access Law (2011)

¹⁰⁰² The General Personal Data Protection Act is exclusively focused on the protection of personal data. The scope therefore excludes the protection of non-personal data.

Canada

Term	Definition	Sources
Sensitive data	Personal information/sensitive information Personal information refers to information about an identifiable individual that is recorded in any form. These include: <i>race, national or ethnic origin, color, religion, age, marital status of the individual, information relating to the education or the medical, criminal or employment history, information on financial transactions, identifying number, symbol or other particular assigned to the individual, address, fingerprints, blood type, personal opinions or views, name of the individual where it appears with other personal information</i>. Categories of information that are considered sensitive according to their nature, context and risk of significant harm. These include: personal information considered sensitive. Sensitive non-personal data: not officially labeled or codified Canada does not have general categories of sensitive or highly sensitive non-personal data. However, the Access to Information Act contains exemptions for commercial, financial or technical information belonging to a third party, international affairs and defense, law enforcement, security of government systems, federal-provincial affairs, information in the economic interest of Canada and patents and trademarks.	Privacy Act (1985)¹⁰⁰³ Personal Information Protection and Electronic Documents Act (2000)¹⁰⁰⁴ Access to Information Act (1985)
Highly sensitive data	Not officially labeled or codified Canada does not have a comprehensive codification of highly sensitive data, but rather a series of laws that require certain categories of information to be treated with higher security standards due to their potential impact on national and individual security. These include but are not limited to: state secrets, classified information, critical infrastructures and sectors data.	Security of Information Act (1985) Canadian Security Intelligence Service Act (1985) Defence Production Act (1985) Communications Security Establishment Act (2019)

¹⁰⁰³ The Canada Privacy Act does not explicitly define a specific term. Instead, it refers to personal information, which may include sensitive data depending on the context. It is exclusively focused on personal data and applies exclusively to federal government institutions.

¹⁰⁰⁴ The Protection and Electronic Documents Act does not explicitly define sensitive information. The scope is exclusively focused on personal data and the private sector. The approach for the categorization of information is based on risk-assessment evaluation according to the nature, context and potential significant harm linked to the information.

China

Term	Definition	Sources
Sensitive data	Important data “Important data” refers to data (personal and non-personal) that, once tampered with, destroyed, leaked, illegally obtained, or illegally used, may endanger national security, economic operation, social stability, public health and security, etc. Additionally, “important data” covers categories of information included in catalogues comprising each geographical region and department as well as the corresponding industries and economic sectors. The Chinese categorization is much more expansive than the EU case-by-case approach. Transfers abroad are restricted, typically requiring security assessments and government approval before export. Sensitive personal information Categories of information that can easily lead to the infringement of the personal dignity or natural persons or the harm of personal property or safety. These include: <i>biometrics, religious beliefs, specific identities, medical health, financial accounts, whereabouts, personal information of minors under the age of 14.</i>	Data Security Law (2021)¹⁰⁰⁵ Personal Information Protection Law (2021)¹⁰⁰⁶
Highly sensitive data	Core state data	Data Security Law (2021)

¹⁰⁰⁵ The PRC Data Security Law establishes a categorical and hierarchical system for data protection and classification based on the importance of the data in economic and social development as well as the extent of harm to national security, the public interest, or the lawful rights and interests of individuals or organizations that would be caused if the data were to be altered, destroyed, leaked or illegally obtained or used.

The Law intentionally presents a broad approach towards data categorization to allow government agencies to reinterpret cataloging based on evolving risks, interests and context.

¹⁰⁰⁶ The PRC Personal Information Protection Law is exclusively focused on the protection of personal data. The scope therefore excludes non-personal information, which is covered by the Data Security Law.

These are: *national security, the lifeline of the national economy, important aspects of people's livelihoods, major public interests.*

Japan

Term	Definition	Sources
Sensitive data	Sensitive personal information Categories of information that require special care so as not to cause unjust discrimination, prejudice or other disadvantages to that person. These include: <i>Race, creed, social status, medical history, criminal record, the fact of having suffered damage by a crime, or other identifiers or their equivalent prescribed by Cabinet Order as those of requiring special care.</i> Sensitive non-personal data: not officially labeled or codified Japan does not have a unified legal framework classifying non-personal data as sensitive. Administrative guidelines do exist for security in specific sectors, like energy. The Act on Access to Information Held by Administrative Organs exempts data from being disclosed to protect various interests including commercial, national security and public order.	Act on the Protection of Personal Information (2003)¹⁰⁰⁷ Act on Information Held by Administrative Organs (1999)
Highly sensitive data	Not officially labeled or codified Categories of information that require to be treated with higher security standards due to their potential impact on national or individual security. These include but are not limited to: National ID system information, credit card, banking information, financial transaction, national security information and critical infrastructure data.	Act on the Use of Numbers to Identify a Specific Individual in Administrative Procedures (2013) Basic Act on Cybersecurity (2014) Act on the Protection of Specified Secrets (2013)¹⁰⁰⁸

¹⁰⁰⁷ The Act on the Protection of Personal Information is exclusively focused on the protection of individuals (natural persons). The scope therefore excludes the protection of non-personal data. The Act is subject to triennial reviews with current proposals for amendments.

¹⁰⁰⁸ Japanese legislation has no specific law or policy that establishes a general definition of highly sensitive data, but rather specific policies covering particular information that could be labelled as highly sensitive data. The scope of these policies covers both personal and non-personal data.

South Korea

Term	Definition	Sources
Sensitive data	Sensitive personal data Categories of personal information prescribed by Presidential Decree that is likely to markedly threaten the privacy of any data subject. These are: <i>ideology, faith, labor union membership, political views, health, medical treatment, sexual orientation, genetic information, criminal records, biometric data, race/ethnicity</i>. Sensitive non-personal data: not officially labeled or codified The Official Information Disclosure Act exempts data from being disclosed to protect various interests, like national security, public safety, third party business secrets or economic interests, or law enforcement.	Personal Information Protection Act (2011)¹⁰⁰⁹ Official Information Disclosure Act (2004)
Highly sensitive data	Not officially labeled or codified Categories of information that require to be treated with higher security standards due to their potential impact in national security, public safety or economic interests. These include but are not limited to: classified national data, infrastructure, communications, industrial technology.	Act on the Protection of Information and Communications Infrastructure (2013) Act on Promotion of Information and Communications Network Utilisation and Information Protection (2016)

United Kingdom

Term	Definition	Sources

¹⁰⁰⁹ South Korea's comprehensive Personal Information Protection Act was enacted Sept. 30, 2011. It is one of the world's strictest privacy regimes. Like the GDPR, it protects privacy rights from the perspective of the data subject and it is comprehensive, applying to most organizations, even government entities. It is not only applicable and strict, but its penalties — which include criminal and regulatory fines and even imprisonment — are enthusiastically enforced.

STUDY ON THE NEXT DATA FRONTIER:
GENERATIVE AI, REGULATORY COMPLIANCE AND INTERNATIONAL DIMENSIONS (NO. 2024-020)

Sensitive data	Special category data - Sensitive personal data Special categories of personal data include things like: <i>racial origin, political opinions, religious, health data, sex orientation, and trade union membership.</i> Sensitive non-personal data: not officially labelled or codified The United Kingdom does not have a unified legal framework classifying non-personal data as sensitive. The Freedom of Information Act exempts certain types of information from disclosure to protect various interests, like intelligence gathering, national security, law enforcement, health and safety, and commercial interests like trade secrets.	UK General Data Protection Regulation (UK GDPR) (2016)¹⁰¹⁰ Data Protection Act (2018)¹⁰¹¹ Freedom of Information Act (2000)
Highly sensitive data	Not officially labeled or codified References to: Highly sensitive personal data - Very sensitive information. UK GDPR includes: <i>genetic and biometric data, which are considered more sensitive and have extra protections.</i>	UK General Data Protection Regulation (UK GDPR) (2016)

United States of America¹⁰¹²

Term / Theme	Definition / Description	Sources
Sensitive data	Sensitive personal data: not officially labeled No comprehensive definition or categorization at federal level. Specific definitions, categories and safeguards can be found in sector-specific legislation. These include but are not limited to health information, non-public personal information in the financial sector,	Health Insurance Portability and Accountability Act (1996) Gramm-Leach-Bliley Financial Services Modernization Act (1999)

¹⁰¹⁰ This Regulation lays down rules relating to the protection of natural persons with regard to the processing of personal data and rules relating to the free movement of personal data.

¹⁰¹¹ This Act makes provision about the processing of personal data. Most processing of personal data is subject to the GDPR.

¹⁰¹² American law takes a more fragmented approach with various regulations governing different sectors and types of data. A notable trend for US businesses is the recent or impending changes to several existing US data protection laws that reflect an increasingly interconnected world with larger volumes of data than ever moving around a more complex information ecosystem.

STUDY ON THE NEXT DATA FRONTIER:
GENERATIVE AI, REGULATORY COMPLIANCE AND INTERNATIONAL DIMENSIONS (NO. 2024-020)

	children's information, social security numbers, financial account numbers, authentication credentials. Sensitive non-personal data: not officially labelled The United States does not have a unified legal framework classifying non-personal data as sensitive. The Freedom of Information Act exempts certain types of information from disclosure to protect various interests, like national security, internal agency rules and practices, trade secrets and confidential commercial or financial information, communications confidentiality between and in agencies, personal privacy, law enforcement records, information related to financial institutions and geological information on wells. Others such as the Preventing Access to U.S. Sensitive Personal Data and Government-Related Data (...) Act imposes specific restrictions and safeguards towards access from third countries.	Children's Online Privacy Protection Act (2000) Freedom of Information Act (1966) Preventing Access to U.S. Sensitive Personal Data and Government-Related Data by Countries of Concern or Covered Persons" Rule (2025)						
Highly sensitive data	Not officially labeled¹⁰¹³ No general definition or categorization at federal level. Categories of information that require to be treated with higher security standards due to their potential impact on national and personal security. These include but are not limited to: classified national security data, federal information, defense, energy, critical infrastructure. Federal information is protected through a level-based model of categories according to the impact of loss of confidentiality, integrity or availability (further detailed below).	Federal Information Security Modernization Act (2014) Cybersecurity Information Sharing Act (2015) Critical Infrastructure Information Act (2002) International Traffic in Arms Regulations (1976)						
Security categorization of information	Standards for Security Categorization of Federal Information and Information Systems¹⁰¹⁴ <table border="1"> <thead> <tr> <th>Low</th><th>Moderate</th><th>High</th></tr> </thead> <tbody> <tr> <td>Loss of confidentiality, integrity, or availability to have a limited adverse effect on organizational operations, organizational</td><td>Loss of confidentiality, integrity, or availability to have a serious adverse effect on organizational operations, organizational</td><td>Loss of confidentiality, integrity, or availability to have a severe or catastrophic adverse effect on organizational</td></tr> </tbody> </table>	Low	Moderate	High	Loss of confidentiality, integrity, or availability to have a limited adverse effect on organizational operations, organizational	Loss of confidentiality, integrity, or availability to have a serious adverse effect on organizational operations, organizational	Loss of confidentiality, integrity, or availability to have a severe or catastrophic adverse effect on organizational	Standards for Security Categorization of Federal Information and Information Systems (2004)
Low	Moderate	High						
Loss of confidentiality, integrity, or availability to have a limited adverse effect on organizational operations, organizational	Loss of confidentiality, integrity, or availability to have a serious adverse effect on organizational operations, organizational	Loss of confidentiality, integrity, or availability to have a severe or catastrophic adverse effect on organizational						

¹⁰¹³ See below: Standards for Security Categorization of Federal Information and Information Systems.

¹⁰¹⁴ The Federal Information Processing Standards outlines a three-level risk impact system to categorize federal information and information systems should there be a breach of security.

	assets, or individuals.	assets, or individuals.	operations, organizational assets, or individuals.	

United States of America. California

Term	Definition	Sources
Sensitive data	Sensitive Personal Information Categories of personal data that reveal personal information deemed sensitive. These include: <i>consumer's social security, driver's license, state identification card, or passport number, financial and banking data, precise geolocation, race, ethnic origin, citizenship or immigration status, religious or philosophical beliefs or union membership, communications, genetic data, neural data, biometric information, health, sex life or sexual orientation.</i>	California (2018) ¹⁰¹⁵ Consumer Privacy Act
Highly sensitive data	Not officially labeled or codified	

United States of America. Virginia

Term	Definition	Sources
Sensitive data	Sensitive Data Categories of personal data that present a heightened risk of harm to consumers. These include: <i>racial or ethnic origin, religious beliefs, mental or physical health diagnosis, sexual orientation, citizenship, immigration status, genetic or biometric data, personal data from a known child, precise geolocation data.</i>	Virginia Consumer Data Protection Act (2023)
Highly sensitive data	Not officially labeled or codified	

¹⁰¹⁵ Act inspired by EU GDPR although there are some crucial differences between the two regulations. Only for-profit businesses meeting several other criteria need to comply with CCPA, while GDPR applies to all kinds of organizations from for-profit to non-profit businesses holding personal data belonging to people residing in the EU. Another crucial difference is that GDPR requires individuals to opt-in before businesses can collect data while there is no opt-in condition in CCPA.

United States of America. Colorado

Term	Definition	Sources
Sensitive data	Sensitive Data Categories of personal data that present a heightened risk of harm to a consumer. These include: <i>racial or ethnic origin, religious beliefs, a mental or physical health condition or diagnosis, sex life or sexual orientation, citizenship or citizenship status, genetic or biometric data, personal data from a known child.</i>	Colorado Privacy Act (2023)
Highly sensitive data	Not officially labeled or codified	

Annex 6 List of Interviewees – Task 3 Use Cases

The following list provides an overview of the persons who contributed to the development of Task 3 use case descriptions, either in online interviews or in writing, revising the use case description and/or responding to a questionnaire.

Organisation/ company	Position	Notes
LNI 4.0 e.V.	CTO	UC1
Global Battery Alliance	Head of Battery Passport Programme	UC1
European Institute for Innovation through Health Data (i~HD)	President	UC2
European Institute for Innovation through Health Data (i~HD)	Digital Health Policy and Strategy Lead	UC2
EFPIA - European Federation of Pharmaceutical Industries and Associations	Director Digital and Data	UC2
	Legal Affairs Manager	Group interview
RCPE - Research Center Pharmaceutical Engineering	Head of Digitalization & Data Governance Scientific Director "Twin4Pharma" Module	UC2
Vicomtech	Director of Transport & Security Division	UC3
RDW International (Netherlands Vehicle Authority)	Senior Advisor	UC3
Meise Botanic Garden	Senior Scientist	UC4
Stockholm Environment Institute (SEI), Stockholm, Sweden	Senior Research Fellow, Trase Research and development Lead	UC4 Common written contribution from colleagues in the Trase initiative.
Stockholm Environment Institute, York, UK	Sustainable Production and Consumption group Lead	
Stockholm Environment Institute, Stockholm, Sweden	Senior Data Engineer, Trase Data Lead	
Global Canopy, Oxford, UK	Trase Global Engagement Lead	
French National Research Institute for Agriculture, Food & Environment	Research Director	UC4
Engineering Ingegneria Informatica SpA	Head of "Data Platforms and Ecosystems" R&I Unit	UC4 Common written contribution from

Organisation/ company	Position	Notes
	Head of the 'Advanced Analytics and Artificial Intelligence Engineering' Research Unit	Engineering Ingegneria Informatica experts.
	R&I Business Developer	
	R&I Business Developer	
IT-Pol	Chairman	UC5.
PnrGO	CTO	UC5
Sirris	Senior Data Scientist	UC6 Group interview
Sirris	Expert in privacy and security for digital services	
Airbus	Digital Twin Solutions Leader	UC6 Note: Input to the study provided as experts' personal views, not as an official Airbus position.

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),
- at the following standard number: +32 22999696, or
- by email via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications

You can download or order free and priced EU publications from: <https://op.europa.eu/en/publications>. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

EU law and related documents

For access to legal information from the EU, including all EU law since 1952 in all the official language versions, go to EUR-Lex at: <http://eur-lex.europa.eu>

Open data from the EU

The EU Open Data Portal (<http://data.europa.eu/euodp/en>) provides access to datasets from the EU. Data can be downloaded and reused for free, for both commercial and non-commercial purposes.



Publications Office
of the European Union

doi:10.2759/3847027
ISBN 978-92-68-33092-0