045646/EU XXVIII.GP
Eingelangt am 19/11/25

EUROPEAN
COMMISSION

Brussels, 19.11.2025
COM(2025) 835 final

COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT AND THE COUNCIL

DATA UNION STRATEGY
UNLOCKING DATA FOR AI

www.parlament.gv.at

1. Introduction - Unlocking data for artificial intelligence

Artificial intelligence is transforming the global economy, and, the EU needs large volumes of high-quality data to compete and drive innovation. Without such data, the EU cannot build strong AI models, optimise healthcare or the energy system, or sustain industrial leadership. For small-and medium enterprises in particular, better data access will be decisive for scaling and remaining competitive.

The EU has laid strong foundations for the creation of a secure, interoperable single market for data through key legislations such as the Data Act[1], and investing in common European data spaces[2]. At the same time, the AI Continent Action Plan[3] and the Apply AI Strategy[4] have created the conditions for the EU to lead in AI development and uptake.

Yet, the EU is facing a scarcity of data for AI development, and growing geopolitical competition where data is increasingly seen as a strategic asset. Much valuable data remains siloed or underused, also due to a complex patchwork of data rules, while global competitors move faster to exploit it for technological and industrial advantage.

To facilitate compliance and improve predictability, the digital omnibus proposes to simplify the data regulatory landscape by **merging four legal instruments into one single, coherent data framework**. Moreover, to support companies and ease compliance, the strategy will be accompanied by a **comprehensive support package under the Data Act**. Model contractual terms standard cloud clauses and a dedicated helpdesk will help SMEs in particular navigate obligations, reduce legal complexity and focus on innovation. Model clauses will apply to both B2G and B2B relations, supporting data creation, sharing and simpler contracts.[5]

The Data Union strategy shifts the focus from rules to results, To achieve this, the EU will act in three priority areas:

- **scaling up access to data for AI**, with initiatives such as **data labs** that offer trusted pseudonymisation services and pool data resources across public and private actors to provide companies and researchers with high-quality datasets.
- **streamlining data rules** to make sharing data easier for businesses and researchers, including **reforming cookie consent** to reduce fatigue while protecting rights.
- **strengthening the EU's global position on international data flows**, by tackling unjustified trade barriers so that European companies can compete on a level playing field globally.

[1] Regulation (EU) 2023/2854 of the European Parliament and of the Council of 13 December 2023 on harmonised rules on fair access to and use of data and amending Regulation (EU) 2017/2394 and Directive (EU) 2020/1828
[2] European Commission, Commission Staff Working Document on Common European Data Spaces, SWD(2024) 21 final, 24 January 2024.
[3] European Commission (2025). AI Continent Action Plan. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. COM(2025) 165 final. Brussels: European Commission.
[4] European Commission, Apply AI Strategy, COM(2025) 723 final, Brussels, 8 October 2025
[5] Updated EU AI model contractual clauses | Public Buyers Community

www.parlament.gv.at

2. Building on the European strategy for data (2020–2025)

With the 2020 European strategy for data[6], the EU created the legal and institutional foundations for a secure and fair single market for data. The goal was to unlock the potential of data for innovation and growth while protecting rights. However, with generative AI and rising geopolitical competition, it is clear that the EU needs to go beyond the foundations it has built.

The European strategy for data was the driver of key legislation to build trust, promote data sharing, and clarify rules across the data value chain. The **Data Governance Act** created mechanisms for trustworthy data sharing, regulated intermediaries, introduced a framework for voluntary sharing of data by companies for purposes of general interest (voluntary data altruism), and opened up certain protected public-sector datasets. The **Data Act** unlocks data from connected products and services by clarifying access and usage rights. Lastly, under the **Open Data Directive** and its Implementing Act on high-value datasets (applicable since June 2024) certain public sector datasets had to be made freely and openly available in machine-readable formats. However, inconsistent national implementation and uncertainties around trade secrets are some of the remaining challenges of the existing legislative framework.

Supporting measures that were set up under the European strategy for data include working with the European Data Innovation Board to coordinate Member States' efforts and a standardisation request to lay the groundwork for a European trusted data framework.[7]

To make the European single market for data a reality, between 2021 and 2024 the Commission also invested €336 million in 14 strategic common European data spaces spanning key economic sectors and areas of public interest, complementing national and private-sector efforts. These spaces provide secure infrastructure and governance frameworks for voluntary data sharing under agreed conditions. The main challenge now is scaling up these efforts for EU-wide impact.

The European Cancer Image Data Space covers anonymized images and annotations. By 2027, ***it will include more than 60 million cancer images****.*

3. Three challenges the EU must address now

As AI technology and services reshape the global landscape, the EU must urgently confront three new, strategic challenges: data scarcity, regulatory complexity, and rising global competition.

[6] The European data strategy – Shaping Europe's digital future, Publications Office, 2020, *https://data.europa.eu/doi/10.2775/645928*

[7] European Commission, Commission Implementing Decision C(2025) 4135 of 1 July 2025 on a standardisation request to the European standardisation organisations as regards a European Trusted Data Framework in support of Regulation (EU) 2023/2854 of the European Parliament and of the Council, available at: https://ec.europa.eu/growth/tools-databases/enorm/mandate/614_en (accessed on 27 October 2025)

www.parlament.gv.at

Data Scarcity: A structural bottleneck for innovation

With the rise of generative AI, large language models (LLMs) and Agentic AI[8], access to high-volume, high-quality, unseen, and domain-specific datasets has become a defining factor of global competitiveness. According to Epoch AI, the size of datasets used to train LLMs doubles approximately every six months.[9]

LLMs and other kinds of foundation models demand massive, diverse sets of training data. Studies suggest that, at current trends, the volume of publicly available training data could be exhausted between 2026 and 2032.[10]

The EU's challenge is twofold: (i) to make high-quality datasets, including also sector-specific datasets, more widely available; and (ii), to ensure the computing infrastructure needed to process these datasets is accessible at scale. Many European firms, especially SMEs and start-ups, lack the data volume and diversity and the access to European compute capacities needed to develop competitive AI solutions. Without urgent action, the EU risks being left behind.

Regulatory complexity: fragmentation hampers scale

Following the 2020 European strategy for data, the EU introduced landmark regulations building on pre-existing rules - the Data Governance Act[11], the Data Act, and various sectoral laws such as the European Health Data Space Regulation[12]. Each of these initiatives focused on specific issues, such as mechanisms for data sharing, fair distribution of value and tackling burdensome localisation requirements. However, complex interplay between the General Data Protection Regulation (GDPR)[13] and sectoral laws, and uneven implementation across Member States created a fragmented regulatory landscape, legal uncertainty, including for public authorities, and raise compliance costs, especially for start-ups and SMEs.

For example, providers of data intermediation services – still an emerging field – are subject to restrictive legal obligations that limit their ability to grow. There is a need to avoid burdening early-stage ecosystems with disproportionate requirements that impede the uptake of data-sharing models and the roll-out of data spaces. To unlock innovation, the EU must simplify the rules on data access and use.

---

[8] "Agentic AI are AI systems that can independently make decisions and take actions. This enables agents to understand language, reason about tasks, take actions autonomously to achieve predefined objectives, and interact with the world around them, orchestrating interactions including with humans

[9] Robi Rahman and David Owen (2024), "The size of datasets used to train language models doubles approximately every six months". *Published online at epoch.ai.* Retrieved from: 'https://epoch.ai/data-insights/dataset-size-trend' [online resource]

[10] Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024). *Position: Will we run out of data? Limits of LLM scaling based on human-generated data*. In K. Chaudhuri, S. Jegelka, L. Song, D. L. Silver, & Y. Ermon (Eds.), *Proceedings of the 41st International Conference on Machine Learning* (Vol. 235, pp. 42085–42101). PMLR. https://proceedings.mlr.press/v235/villalobos24a.html

[11] Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation (EU) 2018/1724 (OJ L 152, 3.6.2022, p. 1)

[12] Regulation (EU) 2025/327 of the European Parliament and of the Council of 11 February 2025 on the European Health Data Space and amending Directive 2011/24/EU and Regulation (EU) 2024/2847

[13] 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

www.parlament.gv.at

Global competition: data as a strategic asset

In the AI race, access to high-value data is a key strategic advantage. Globally, data has become a geopolitical asset, with data access, localisation, and control increasingly used as instruments of power. While the EU promotes open, secure, fair and trusted data flows, other jurisdictions follow assertive or protectionist strategies. Localisation and restrictive access regimes abroad limit the EU's access to global resources and expose EU firms to economic and security risks. To unlock the full potential of European AI, the Union must treat data as a core strategic resource and invest in secure, high-quality, and interoperable datasets that reflect European values and standards. Strengthening Europe's ability to collect, curate, and use its own data is both an economic and security imperative. The EU must secure beneficial flows, safeguard sensitive non-personal data within the EU, and support digital sovereignty amid intensifying technological rivalry.

4. The three pillars of the Data Union Strategy

Pillar I: Scaling up access to quality data for AI and innovation

The EU's competitiveness in AI and digital innovation depends on access to high-quality data and the infrastructure to share and use data securely at scale. The EU has already laid strong foundations with common European data spaces, governance frameworks, and major investments in cloud technology and computing. The challenge now is to move from pilot projects and fragmented initiatives to a seamless, interoperable, and sustainable data ecosystem, encouraging breakthrough innovation and strengthening the EU's digital sovereignty.

To achieve this, the Commission will act along two complementary tracks. First, it will launch flagship initiatives that address the EU's most immediate bottlenecks: limited access to critical datasets, insufficient infrastructure for large-scale AI development, and the need for trusted environments, including data labs that connect data spaces with AI developers. These data labs will serve as specialised service facilities providing secure environments, practical tools, and expert support for data pooling, curation, pseudonymisation, and

Data Spaces and Data Labs: The Building Blocks of Europe's AI Ecosystem

Common European Data Spaces are data-sharing ecosystems built on cloud infrastructure and clear governance rules defining who can access, use, and share data. They connect public and private actors around trusted mechanisms for data exchange within and across sectors.

Data Labs are data service providers that link these data spaces with the AI ecosystem. They give companies and researchers secure, practical access to high-quality datasets, the support they need to ensure compliance with EU rules, and offer tools, guidance, and trusted environments for data pooling, curation, labelling and pseudonymisation.

Data spaces provide the structured sources of trustworthy data, while data labs turn this data into usable resources for innovation and AI development, ensuring a seamless flow from availability to application.

www.parlament.gv.at

anonymisation. They will help companies, especially SMEs, turn data into usable resources for AI training while preserving data control. These efforts will work hand in hand with the Apply AI Strategy, ensuring that data availability directly supports AI deployment and innovation across industries and public sectors. Second, it will reinforce these efforts with horizontal enablers: legal clarity for data pooling, standards for data quality, and investment in synthetic data[14] capacities, ensuring scale, trust, and long-term sustainability across all sectors.

i. Scaling up the common European data spaces

The **common European data spaces (CEDS)** are central to building a single market for data. The next phase will scale them up and link them to AI infrastructure through data labs and AI factories, turning the EU's data assets into fuel for trustworthy AI. In close synergy with the Apply AI Strategy, these efforts will ensure that data spaces directly enable AI development and deployment across sectors.

Simpl cloud middleware[15] will enable interoperability across initiatives through an open-source, modular, and secure set of components. This lowers barriers for SMEs and creates faster links between ecosystems. The data spaces support centre will reinforce uptake, especially among SMEs, by raising awareness and practical guidance.

Future EU funding for CEDS will prioritise sectors of public interest, such as health, mobility, energy, public administrations and the environment, while mature domains like manufacturing and finance transition to market-driven models. The Commission will support this transition by promoting standardisation, interoperability, and co-investment frameworks. End-user integration, AI readiness, and financial sustainability will remain key objectives.

Among the flagship actions under the Apply AI Strategy, the EU will leverage **Common European Data Spaces** to accelerate AI deployment across key sectors and support the development of Frontier AI models through the Frontier AI Initiative. These actions are closely linked to other Apply AI flagships, such as Foundational Models for Industry, AI-powered Pharma Discovery, Autonomous Drive Ambition Cities, each drawing on sectoral

Next Steps for European Health Data Space:
The EHDS will serve as a key bridge between health data ecosystems and AI development, allowing data labs and AI Factories to leverage anonymised and synthetic datasets within trusted processing environments.
From March 2029, patient summaries and ePrescriptions will be exchanged across all Member States, alongside secondary use of most health data. By March 2031, this will extend to medical images, lab results and discharge reports, with genomic and other data added for secondary use.

[14] Synthetic data is artificially generated data that is not collected from real-world events but is engineered to statistically mimic the properties, patterns, and relationships of a real dataset.
[15] Simpl is an open source, secure middleware that supports data access and interoperability in European data initiatives. It provides multiple compatible components, free to use, that adhere to a common standard of data quality and data sharing; https://simpl-programme.ec.europa.eu/

www.parlament.gv.at

data made available through the Common European Data Spaces. This approach translates into concrete applications: AI-powered screening centres in healthcare that validate diagnostic tools using the European Health Data Space[16]; trusted data pooling in manufacturing through the Data Space for Manufacturing to train specialised and frontier AI models; and an Agri-Food AI Platform that supports the uptake of AI-enabled farming tools using the Common European Agricultural Data Space.

As of 2026, the roll-out of data spaces across priority sectors will continue, supported by ongoing EU investment of around EUR 100 million enabling trusted and large-scale data use for AI applications. The **European Health Data Space** will support AI-based diagnostics and personalised medicine and serve as a key bridge between health data ecosystems and AI development, allowing Data Labs and AI Factories to leverage anonymised and synthetic datasets within trusted processing environments; the common European **Mobility Data Space** will enable the connection vehicles, infrastructure, and logistics for safer, greener transport; the Energy Data Space will facilitate smart and flexible energy services; and the **Media Data Space** will boost creative industries through AI-driven cultural innovation. Data labs will act as practical entry points to these data spaces, helping organisations access, prepare and use the data effectively for AI. Within this framework, the **European legal data space** will expand access to legal and judicial data through common identifiers and metadata for case law and legislation, enabling LegalTech to use this data. The need for a contract terms data pool for automated contracting will be explored in this context.

The Commission will fast-track environmental digitalisation through the **Green Deal Data Space**, enabling the DigitalGreenTech community to scale cross-sector solutions using reusable components and high-quality datasets. Priority actions include data-driven services for the European Water Resilience Strategy, digitisation of permitting processes, pilots on textile traceability and nature credits, and advanced forest monitoring with machine learning on open and confidential data.

A **European Defence Data Space** will create a trusted environment for pooling operational, industrial, and research data to develop next-generation defence systems, boost industrial capabilities, and strengthen EU technological sovereignty by reducing reliance on third-country providers. Drawing on Ukraine's experience in data-driven defence, the Commission will explore cooperation and knowledge exchange. The initiative will be developed with Member States and relevant stakeholders, including businesses. [17]

ii. Data labs

As outlined in the AI Continent Action Plan, data labs will be specialised facilities, linking data holders, common European data spaces, domain-specific data ecosystems, and the EU AI

[16]It will also build on Europe's Beating Cancer Plan, the Life Sciences Strategy, and the EU Cardiovascular Health Plan
[17] This initiative will be guided by the European Defence Agency's feasibility study due by end-2025

www.parlament.gv.at

ecosystem. Data labs[18] will provide hands-on services – such as data pooling[19], curation[20], labelling and pseudonymisation[21] – to help organisations, in particular start-ups and scale-ups, share and use data safely, facilitate cooperative AI training and support the development of AI models in key sectors and covering different governance and licensing models. In line with the Apply AI Strategy, data labs will translate the availability of high-quality data into concrete AI deployment, serving as practical enablers that accelerate experimentation, adoption and scaling.They can also be used to carry out tasks that require advanced AI resources on behalf of Data Spaces and other Data infrastructures, for example producing synthetic data, or carrying out advance privacy and business secrecy preserving, to help organisations to share and use data safely.

By pooling public and private resources, data labs will help overcome a key market failure: limited availability of diverse, high-quality data and reluctance to share privately held data for AI training. They will operate through existing access channels and frameworks without requiring direct data transfer. In doing so, data spaces remain the trusted infrastructures where data is governed and made available, while data labs can act as the operational interface that enables its safe, value-adding use for AI.

Participation will be voluntary, and data holders decide how, when, and by whom data can be used. No data will be transferred without explicit consent. All activities will be protected by strict confidentiality safeguards and supported by privacy-preserving and decentralised techniques such as federated learning, homomorphic encryption, and secure multi-party computation. Data can be processed locally or across nodes without being merged into a single repository, ensuring that it remains under the control of the original holder. This model – particularly beneficial for SMEs – supports compliance with EU data protection rules, safeguards confidentiality, and builds trust while expanding data use for AI.

The EU's compute capacity has evolved from science-oriented high-performance computing (HPC) under EuroHPC to AI Factories, which expand this concept to support AI development, linking compute infrastructure with data access and experimentation. The upcoming AI Gigafactories will further scale AI compute facilities.

Within this framework, the first data labs will be established under the AI Factories initiative through EuroHPC, providing secure environments and data services to connect AI developers with common European data spaces in areas such as healthcare, manufacturing, energy and climate, and expanded to languages, cybersecurity, and cultural heritage. To ensure their services reach companies and public administrations, data labs will work in close coordination

[18] In some contexts, the term "data containers" is used to refer to similar facilities that enable structured, secure, and trusted data use across different settings. Together with the broader concept of "data containerisation," they reflect a complementary approach to organising and governing data exchange, promoting interoperability and consistency across the EU AI ecosystem.
[19] Combining and sharing of data from multiple sources into a single, centralized repository or shared environment.
[20] Organising, integrating, validating, and maintaining data including its labelling for improving access and use.
[21] Art 4(5) of the Regulation (EU) 2016/679: "The processing of personal data in such a manner that the personal data can no longer be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organisational measures to ensure that the personal data are not attributed to an identified or identifiable natural person."

www.parlament.gv.at

with the European Digital Innovation Hubs (EDIHs), which act as user-facing contact points and help match data needs with concrete applications.

Further Data Labs will be set up independently in other domains to address specific sectoral or research needs, such as the energy sector. The upcoming AI Gigafactories will further scale AI compute facilities and prepare the Data Lab model for commercial rollout across the EU, turning it into a self-sustaining service ecosystem that connects compute, data, and AI innovation.

Data labs will provide services specifically across nine key areas:

- **Bridge between data spaces and AI ecosystems:** practical linkage that enables companies to access high-quality, interoperable data by connecting common European data spaces with AI developers, infrastructures and sectoral ecosystems.
- **Technical infrastructure and tools:** data containers will enable efficient storage and organisation of data complemented by secure environments for the on-site processing of sensitive data, along with ready-to-use tools for data preparation and privacy-preserving techniques to achieve anonymisation and synthetic data generation. A high standard of usability, speed, and scalability will be ensured so that tools are simple, reliable, and easy to adopt.
- **Data pooling:** support for companies in aggregating data from public and restricted sources - particularly data used for innovative purposes -, using the trusted data sharing mechanisms of the common European data spaces. Data Labs will support businesses to comply with EU competition law when exchanging or pooling data. Building on and complementing the Horizontal Guidelines, which provide companies with practical guidance on collaboration and shared resources, the Commission will further support Data Labs in this role with dedicated guidance on best practices in data exchange and pooling. In addition, tailored guidance for individual data labs following a request under the Informal Guidance Notice will be available.
- **Pseudonymisation and anonymisation services:** provision of advanced tools and expertise to remove or mask personal identifiers. These services will include techniques such as pseudonymisation, anonymisation, and differential privacy, enabling safe data reuse while maintaining analytical utility.
- **Synthetic data generation:** support for creating high-quality synthetic datasets that replicate the statistical properties of real data without exposing sensitive or confidential information. Data labs will provide tools and expertise to generate, validate, and benchmark synthetic data for AI model training and testing, complementing anonymisation efforts and improving data availability in sensitive domains.
- **Data curation, labelling, and vectorisation:** comprehensive support for cleaning, labelling, annotating, enriching, and vectorising datasets to make them reliable, representative, and usable for AI training. This includes quality assurance processes, transparent documentation, and collaboration with expert communities for domain-specific labelling.

www.parlament.gv.at

- **Regulatory guidance and training:** tailored advice to help businesses comply with EU law combined with training for AI developers on data use and legal obligations, such as AI regulations, copyright, trade secrets and competition law, combined with training for AI developers on data use and legal obligations.
- **Bridge between data spaces and AI ecosystems:** practical linkage that enables companies to access high-quality, interoperable data by connecting common European data spaces with AI developers, infrastructures and sectoral ecosystems.
  Data access facilitation: a demand-driven service where start-ups and SMEs can signal their data needs, with data labs helping them find relevant datasets and overcome market, legal or administrative barriers.

How would a data lab work in practice?

A company in Member State X develops AI-based predictive maintenance systems for electric vehicles but struggles to access enough high-quality sensor data from different car models and charging infrastructures. Individual manufacturers are hesitant to share this data due to trade secrets, privacy, and competition concerns. AI Factories will provide the computing resources and, through their integrated data labs, data management services needed to overcome these barriers.

Through the data lab, the company would access trusted, anonymised, and aggregated datasets coming from different sources, such as public charging operators, participating Original Equipment Manufacturers (OEMs) and other data discovered via the European mobility data space.

As a component of the AI Factory, the data lab would offer:

- *Secure environments to analyse real-time sensor data through federated learning without the data leaving OEM systems.*
- *Anonymisation services ensuring privacy-compliant use of driver and vehicle data.*
- *Regulatory guidance on applying the Data Act's data access provisions and managing trade secret protection.*
- *Data curation tools that harmonise different sensor formats and quality standards.*

The lab thus would act as a bridge between the mobility data space and the AI ecosystem, allowing the company to train robust AI models while safeguarding manufacturers' confidentiality.

iii. The Cloud and AI Development Act

Sustainable data centre capacity and sovereign cloud and AI services are a prerequisite for the EU to achieve the objectives laid down in this strategy. As increasing amounts of data are generated, there is a growing need to collect, store, combine and process this data. To minimise

www.parlament.gv.at

latency[22] and decrease reliance on infrastructure located in other parts of the world, the EU needs to house sufficient data centre capacity.

To ensure the availability of sustainable data centre infrastructure and sovereign cloud and AI services for EU businesses and public administrations, the Commission will propose a **Cloud and AI Development Act** in Q1 2026. This initiative will support innovation across the entire cloud and AI value chain, from the integration of cutting-edge processors to sustainable cooling technologies and AI hardware and software. It will also accelerate the roll-out of sustainable data centre capacity, ensuring that the EU has the infrastructure needed for secure and sovereign cloud and AI services.

iv. Strategic data assets: public sector, scientific, cultural and linguistic resources

The EU's competitiveness in AI depends on access to high-quality, structured, and trustworthy data. Scientific, cultural, and linguistic datasets are critical enablers for robust AI models, research breakthroughs, and technological sovereignty.

Public-sector reference datasets under the Open Data Directive will be scaled up. The high-value datasets[23] have to be made available free of charge, through application programming interfaces (APIs), in a machine-readable format and, where relevant, provided as a bulk download. In 2026, the Commission will propose to expand the list of high-value datasets to cover legal, judicial, administrative and other data. This will be beneficial for start-ups and SMEs. The Commission will also monitor whether further datasets should be added.

Scientific data has already proven transformative, as seen with AlphaFold.[24] Well-structured databases reduce research and development (R&D) costs, accelerate innovation, and open up new frontiers in materials, pharmaceuticals, energy, and biotech. To build on this, the Commission will continue to map existing databases, to set priorities with experts, secure usage rights, and fund new digital infrastructures according to the European strategy on research and technology infrastructure. In this regard, the **European Open Science Cloud** (EOSC), the common European data space for R&D, is developing a federation of data repositories with a trusted platform for sharing and reusing high-quality, findable, accessible, interoperable and reusable (FAIR) research data, tools and services across disciplines and borders in Europe. This will support the scientific activities with AI in RAISE.[25] In parallel, the forthcoming proposal

[22] Latency is the time it takes for data to pass from one point of a network to another.
[23] In line with Annex I of the Open Data Directive these high value datasets come from the following categories: geospatial, earth observation and environment; meteorological; statistics, companies and company ownership, mobility. New categories can be added.
[24] AlphaFold is an Artificial Intelligence system developed by Deep Mind, which uses deep learning and large amounts of data to predict protein structures. This helps accelerating breakthrough research in many fields of biology.
[25] European Commission (2025). Communication from the Commission to the European Parliament and the Council – A European Strategy for Artificial Intelligence in Science: Paving the way for the Resource for AI Science in Europe (RAISE). Brussels, 8 October 2025, COM(2025) 724 final

www.parlament.gv.at

for a European Research Area (ERA) Act[26] will strengthen legal conditions to share, access and reuse publicly funded research results, publications and data for scientific purposes.

The EU's cultural and linguistic resources will also be scaled up. More than 30 million digitised works from Europe's cultural institutions will be made available for AI development, building on the Europeana initiative.[27] The Commission will explore how to strengthen cooperation and encourage licensing between public broadcasters and AI providers, in order to make their audiovisual archives accessible for AI training, taking into account the remuneration of rightsholders.

Pilot projects under the **European common language data space and the Alliance for Language Technologies (ALT-EDIC)** will crowdsource domain-specific datasets including from smaller languages, adding to the 477 billion tokens already available - comparable to leading AI training datasets. This will also help to ensure that rare languages are included in AI Large Language Models (LLM) development, which will have an impact on the quality of the results of AI systems in these languages.

v. Horizontal enablers: synthetic data, data pooling, and standards

Alongside flagship initiatives, the EU also needs horizontal measures that cut across sectors and give scale to the entire data economy.

Synthetic data as a driver of AI leadership

Synthetic data[28] can unlock AI training in areas where data is scarce or sensitive, from rare disease research through to robotics or autonomous driving edge cases. It enables AI model development without exposing personal or proprietary information, strengthening both competitiveness and privacy-preserving innovation.

To harness this potential, the Commission will develop guidance and standards for trusted synthetic data use, examine the related legal questions, consult on a voluntary European certification scheme, and explore the possibility of setting-up a 'synthetic data factory' to provide access to high-performance computing for large-scale dataset generation. Horizon Europe will also fund cutting-edge R&D in synthetic data generation techniques.

[26] European Commission, Forthcoming proposal for a European Research Area (ERA) Act, announced in the Commission Work Programme 2025, Brussels, 11 February 2025, available at: https://commission.europa.eu/strategy-and-policy/strategy-documents/commission-work-programme/commission-work-programme-2025

[27] Europeana, *The European digital platform for cultural heritage*, available at: https://www.europeana.eu/en (accessed on 27 October 2025)

[28] See definition above

www.parlament.gv.at

Clearing the path for strategic data pooling

Many companies for example in health, mobility, energy, agriculture, and manufacturing lack the large, diverse datasets needed to train advanced AI models. Pooling of data related to early stages of the production cycle of products and services could unlock shared benefits, but legal uncertainty and fear of breaching competition law hold back collaboration.

> ***Draghi report:*** *"In particular, to overcome the EU's lack of large data sets, model training should be fed with data freely contributed by multiple EU companies within a certain sector. It should be supported within open-source frameworks, safeguarded from antitrust enforcement by competition authorities."*

The Commission will continue to act to provide legal clarity for companies, in line with the call to turn rules into results in the report on the future of European competitiveness by Mario Draghi. The 2023 Horizontal Guidelines on cooperation agreements between competitors already explain when data pooling is compatible with EU competition law, with practical examples and safeguards.

To further facilitate lawful and effective data collaboration through Data Labs, the Commission will issue dedicated guidance on best practices in data exchange and pooling.

In addition, competition law guidance can be provided by the Commission upon request under the Informal Guidance Notice for specific data-related multi-country projects and initiatives that foster cross-border innovation, industrial resilience, and AI development. By making data pooling a trusted and legally secure option, the EU can unlock efficiencies and accelerate breakthroughs in key sectors.

Raising the bar on data quality and data capturing

Without reliable standards, even the most ambitious data-sharing efforts risk fragmentation and low uptake. The European trusted data framework[29] already sets rules for sharing, metadata, and governance, but further work is needed to address emerging issues.

The Commission will launch a standardisation request for a **European data quality standard** covering completeness, consistency, provenance, semantic clarity, and governance, giving businesses, regulators, and researchers shared benchmarks for reliable datasets. This work will complement the ongoing standardisation efforts on data quality and documentation under the AI Act, ensuring coherence between data management and AI development requirements.

A dedicated initiative will aim to standardise **annotation and labelling practices**, making data easier to find, combine, and reuse while ensuring trust in its origin and conditions of use, which is critical for scaling AI training and cross-sector reuse. A multi-stakeholder workshop will also investigate standards for data capture from connected products, sensors, and cameras -

[29] See also European Commission, Implementing Decision C(2025) 4135 on the European Trusted Data Framework

www.parlament.gv.at

including sampling, metadata, timestamping, calibration, and integrity - addressing a key barrier to effective data pooling and reuse.

Flagship actions

- Launch first **data labs** to scale data availability and link to AI ecosystems (Q4 2025). They will also offer trusted pseudonymisation services.
- **Launching the quality data for AI initiative:** expanding high-value datasets under the Open Data Directive (Q4 2026); setting up a stakeholder forum with public broadcasters and AI developers (Q2 2026); making 30 million digitised cultural objects available for AI training (Q4 2026); and launching a crowdsourcing initiative for domain-specific data and language data in smaller European languages (Q2 2026).

Pillar II: Streamlining data rules

The EU's data framework must remain clear, practical, and innovation friendly. To reduce burdens and boost competitiveness, the Commission is presenting a legislative proposal, known as the **Digital Omnibus,** aiming, amongst other things, to modernise and consolidate the EU's horizontal data acquis. In addition, the Commission will also announce work on one-click compliance to enable automated regulatory reporting, and a support package for the Data Act, including model contracts, standard clauses, guidance on compensation and trade secrets, and a legal helpdesk for SMEs.

i. Simplifying the EU's data acquis

The EU's regulatory data framework has grown rapidly, creating new rights but also increasing complexity and fragmentation. Simplification is needed to reduce compliance costs, make rules easier to apply, and better support innovation.

To this end, the Commission is presenting the above-mentioned Digital Omnibus. It will update the acquis, removing unnecessary burdens while safeguarding the core principles of the EU's data economy. The Omnibus will focus on the following priority reforms:

- **Deleting outdated rules.** The Omnibus will repeal the Free Flow of Non-Personal Data Regulation,[30] as its functions are already covered by the Data Act, while explicitly preserving the principle of free movement of non-personal data and the ban on unjustified localisation.
- **Streamlining data-sharing rules.** The Omnibus will repeal the Data Governance Act (DGA) and migrate its essential provisions into the Data Act. Obligations for data intermediaries will be clearer, lighter and voluntary to enable viable models and wider uptake.
- **Consolidating public-sector data-sharing.** Rules now split between the DGA and the Open Data Directive will be kept and merged into one Data Act chapter. This simplifies obligations while preserving openness, transparency, and fair access. The new

[30] Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union

www.parlament.gv.at

framework will furthermore tackle power imbalances in data sharing, ensuring fair conditions and tangible benefits for SMEs. Data labs will flag promising new public sector datasets not yet covered.

- **Modernising rules for cookies and similar technologies**. The Omnibus will reform the rules on cookies currently in the ePrivacy Directive and bring them into the GDPR framework. It will propose practical solutions: cookies and similar technologies for certain low-risk purposes should be considered lawful, while other purposes, the operators should rely on one of the legal bases under the GDPR. It will also simplify banners with one-click options. It will oblige websites to respect users' preferences, also through their browsers. Beyond the Digital Omnibus, the ePrivacy framework will be reformed to ensure that current rules meet today's needs and allow for effective protection of people and business,without compromising fundamental rights and preserving independent journalism. The relevant provisions will be integrated into other legal instruments, allowing the Directive to be ultimately withdrawn.

> ***Participant in the Data Union public consultation:*** *"ePrivacy rules need to be urgently updated. The rules set in place were designed against a totally different technological background and do not reflect the current market needs."*

- **Developing an innovation-friendly privacy framework.** Targeted **GDPR** amendments will, in particular, clarify the notion of personal data, harmonise at EU level when data protection impact assessments should be conducted, simplify data breach notifications to supervisory authorities, streamline breach notifications via a single EU entry-point, simplify information obligations where there are reasonable grounds to expect that individuals already have the information and the risk to the data subject is low; clarify that legitimate interest can be a legal basis for training AI, including the incidental processing of special categories of data; clarify the provisions on automated individual decision-making.

  One key change concerns liberating data for AI through trusted anonymisation. Today, uncertainty about sufficient anonymisation of personal data is a core concern, often discouraging data sharing. Businesses struggle in particular to determine when pseudonymised data no longer constitutes personal data for certain entities. This uncertainty makes data sharing more complex where the GDPR requirements are complied with out of precaution. The Commission will support businesses by specifying the means and criteria to determine whether data resulting from pseudonymisation constitutes personal data for certain entities.

  This will include an assessment of the state of the art of available techniques and the development of criteria to assess the risk of re-identification. While businesses remain fully responsible for compliance with the GDPR, they can use the implementation of those means and criteria to demonstrate that data cannot lead to the reidentification. The amendments will also facilitate AI model training, with the appropriate safeguards.

www.parlament.gv.at

The goal of these changes is to provide legal clarity for AI development, including cases of incidental processing of sensitive data where developers have made genuine efforts to remove such data. while protecting individuals' rights and competitiveness of businesses.

- **Refining the Data Act for practical implementation.** The essential features of the **Data Act** will remain unchanged. At the same time, business-to-government data sharing will be limited to emergencies, easing burdens while safeguarding crisis response. Targeted additional adjustments will prevent data 'leakage' to outside the EU, introduce tailored regimes for custom-made cloud services and remove the provisions on smart contracts.
- **Reducing burdens for scaling companies.** A new category of small mid-caps (250–749 employees) will extend SME-type provisions under the Data Act, the Open Data Directive, and integrated DGA rules.

ii. Building a future-proof data framework

As part of the Digital Fitness Check, the Commission will continue reviewing the EU's data acquis to keep it coherent, proportionate, and innovation friendly. With particular attention to SMEs, it will identify overlaps, gaps, and unclear interactions, including with sectoral data laws, to create a more predictable cross-sectoral framework.

In addition, we will modernise digital legislation and data protection.[31] Targeted adjustments may ease compliance and strenghten enforcement, supporting the development of robust and trustworthy innovations.

Data brokerage has become a growing concern, with certain companies collecting, aggregating, and trading personal data without individuals' awareness, meaningful consent, or control. Such opaque practices undermine core principles of data protection law, privacy, distort competition, and erode public trust in digital markets. A strengthened enforcement of the existing rules is needed. The Commission will assess whether additional safeguards are needed to curb these practices, enhance transparency in data trading, and ensure that individuals and businesses can trust how data is accessed and exchanged across the Union.

iii. One-click compliance

Today, companies spend significant time and money on compliance. Even data already in digital form must often be reformatted and resubmitted to multiple authorities, where it is checked manually. This duplication creates fragmented oversight and diverts resources from innovation.

Beyond simplifying rules, the EU is investing in technologies to automate compliance. Through Horizon Europe and the Digital Europe Programme, it supports common data models, interoperability frameworks, and automated analysis. Pilot projects already show how real-

[31] Commission work programme EUR-Lex - 52025DC0870 - EN - EUR-Lex.

www.parlament.gv.at

time, automated compliance checks can work in practice. The Digital Product Passport (DPP) is an early example of this approach in product legislation.

Building on these experiences, "one-click compliance" would make regulatory requirements machine-verifiable, turning company data into standardised digital compliance certificates - much like the DPP enables automatic product compliance.

One-click compliance could be particularly valuable in areas like cybersecurity, where companies face requirements under NIS2[32], the Cyber Resilience Act[33], and other frameworks.

The **European Business Wallet** Regulation will be a key enabler of this approach. It will provide a trusted and interoperable digital environment for storing, managing, and sharing verifiable credentials, including compliance certificates. Companies could use **European business wallets** to digitally identify themselves, identify and validate users of the ecosystem and demonstrate conformity with multiple EU rules through the submission of compliance certificates, while public sector bodies regulators are provided with secure, immediate access to validated information. Over time, the **European business wallet** will become a common infrastructure supporting administrative processes such as licensing, public procurement, and access to funding, enabling seamless digital interactions between businesses and authorities across the Single Market.

Determining who is accountable in case of errors, misuse, or system failures - whether the company, the certifier, or the regulator - will be essential to ensure trust and legal certainty. The Commission will therefore explore these issues in an upcoming public consultation, assessing both the opportunities and the safeguards needed to build a reliable and accountable automated compliance ecosystem.

Beyond cutting costs for SMEs and mid-caps, such a system would also give policymakers insights into how rules work in practice, strengthening evidence-based regulation. One-click compliance could become a cornerstone of the EU's digital simplification agenda, aligning competitiveness with trust and accountability.

iv. Helping businesses comply with the Data Act

The **Data Act** constitutes the key set of rules for using and sharing data. To ensure that companies, especially SMEs and small mid-caps, can fully use its potential and focus on innovation rather than red tape, the Commission has already issued a FAQ document[34] and

[32] Directive (EU) 2022/2555 of the European Parliament and of the Council of 14 December 2022 on measures for a high common level of cybersecurity across the Union, amending Regulation (EU) No 910/2014 and Directive (EU) 2018/1972, and repealing Directive (EU) 2016/1148 (NIS 2 Directive), OJ L 333, 27.12.2022, p. 80–152

[33] Regulation (EU) 2024/2847 of the European Parliament and of the Council of 23 October 2024 on horizontal cybersecurity requirements for products with digital elements and amending Regulations (EU) No 168/2013 and (EU) No 2019/1020 and Directive (EU) 2020/1828 (Cyber Resilience Act), OJ L [2847], 20 November 2024

[34] European Commission, Frequently Asked Questions – Data Act, version 1.3, Brussels, 12 September 2025, available at: https://digital-strategy.ec.europa.eu/en/library/commission-publishes-frequently-asked-questions-about-data-act (accessed on 27 October 2025)

www.parlament.gv.at

guidance on in-vehicle data[35], and will further complement these with a broader package of support measures.

Immediate measures include:

- model contractual terms for data sharing to reduce legal complexity, cut transaction costs, and give businesses confidence when entering into new partnerships;
- standard contractual clauses for cloud services to make switching easier and contracts fairer, supporting competition and innovation in the European cloud market.

Further measures, to be phased in, will include:

- Guidelines on reasonable compensation to clarify what can be charged for data sharing, providing legal certainty to both data holders and data recipients (Q1 2026);
- New guidance on selected definitions in the Data Act (Q1 2026);
- A Data Act legal helpdesk to provide direct assistance for companies with concrete questions on how to apply the new rules, giving priority to SMEs to ensure their queries are addressed swiftly and with dedicated attention (Q4 2025).

Together, these measures will make the Data Act easier to navigate, reduce unnecessary costs, and give companies the clarity and confidence they need to seize new opportunities in the EU's data economy. The Commission will closely monitor the uptake of the contractual tools, in particular the model contractual terms and standard contractual clauses, and will review, complement or adapt them as needed in line with international developments in data sharing.

The Commission will seek synergies between the public buyers community and the European Data Spaces to enhance public-sector efficiency, drawing on the blueprint established between the European Health Data Space and the Big Buyers Working Group on Healthcare Efficiency.[36]

Flagship actions

- Proposal to **consolidate data legislation** (Q4 2025)
- Proposal to **update ePrivacy rules** on cookies and similar technologies (Q4 2025)
- Proposal for **targeted GDPR adjustments** (Q4 2025)
- Launching a **one-click compliance initiative** (from Q4 2025 onwards)
- Rolling out support measures for the implementation of the Data Act (from Q4 2025 onwards)

[35] European Commission, Guidance on vehicle data, accompanying Regulation (EU) 2023/2854 (Data Act), C(2025) 6119 final, Brussels, 12 September 2025

[36] Can the European Health Data Space enable better procurement? – Big Buyers are investigating | Public Buyers Community

www.parlament.gv.at

Pillar III: Safeguarding the EU's data sovereignty through a strategic international data policy

Data sovereignty is at the core of the EU's digital future. It means that the EU must retain control over how data is accessed, used, and protected – both within its territory and abroad. Sovereignty requires openness to trusted partners, including exchange of data across borders, but on terms that are fair, secure, and consistent with EU values and interests. A situation in which foreign actors enjoy unfettered access to the EU market while European companies face unjustified barriers abroad cannot be sustained.

Safeguarding sovereignty also means protecting the EU's resilience. Cyberattacks, technology leakage, surveillance, and coercive dependencies put critical data at risk. The EU must ensure the availability, integrity, and security of sensitive datasets, preventing their misuse or exploitation, in particular by actors outside the EU.

To this end, the Commission will pursue a strategy that combines openness with strength: making fair conditions for data access and cross-border transfer a pillar of digital trade, protecting sensitive EU non-personal data through clear safeguards, and deepening cooperation with trusted partners. It will also work to shape global governance models that reflect EU interests and values and prevent fragmentation into rival spheres. This strategy will complement the long-lasting EU approach to safe personal data flows developed through the EU data protection acquis.

In a stakeholder poll, 75% of participants supported a more assertive EU approach to international non-personal data flows.

While the EU has built a robust legal framework and promoted "data free flow with trust" internationally, new unjustified data localisation requirements, export controls, and discriminatory rules abroad threaten to undermine sovereignty. The Commission will therefore act more assertively to defend EU interests and regulatory autonomy, with proportionate measures where openness is abused or vulnerabilities weaponised.

i. Fair cross-border data flows and safeguards for EU sensitive non-personal data

The Commission will embed fair conditions and effective control of cross-border data flows into international digital trade. Structured exchanges, e.g. in the framework of the EU's digital partnerships and dialogues will address existing imbalances where EU data flows abroad without adequate safeguards.

If gaps persist, and on the basis of objective criteria, the Commission will take proportionate action in full respect of the Union's international commitments. It will issue guidelines in Q2-2026 to assess the treatment of EU entities by third countries and develop an anti-data-leakage toolbox in Q1 2026 to address localisation demands, market exclusion, or insufficient safeguards or any other unjustified treatment. This toolbox may draw on or be inspired by

www.parlament.gv.at

instruments such as the Trade Enforcement Regulation[37], the Anti-Coercion Instrument[38], and economic security considerations, as applicable, and will focus on technologies and best practices to strengthen the EU's resilience. Should structural distortions or persistent discriminatory practices remain unaddressed, the Commission will, where necessary, consider additional measures to ensure fair conditions for data access and use.

In parallel, the Commission will better protect EU sensitive non-personal data, complementing the protection of personal data guaranteed through the GDPR and adequacy decisions. Working with stakeholders, and following the results of in-depth risk assessments, it will adopt a first package of targeted measures by Q3 2026.

ii. Linking EU data-sharing ecosystems with those of like-minded third countries

The EU's legal framework for data protection, cybersecurity, enforcement, and judicial redress is a reliable basis for foreign data holders. The Commission will foster secure, convergent and interoperable links between EU data ecosystems and those of like-minded partners to attract more data flows to the EU.

Planned measures include (i) supporting services and infrastructure such as the CEDS to enable seamless cross-border sharing; (ii) providing tools like standard contractual clauses to help businesses ensure lawful exchanges; (iii) and embedding commitments on cross-border data sharing in bilateral and plurilateral international agreements.

To strengthen convergence and interoperability, the Commission will promote the European Trusted Data Framework in international dialogues and the Digital Partnership Network. It will also explore creating a trust label, potentially linked to the data spaces maturity model – a standardised framework designed to assess the capabilities of data space initiatives - to support cooperation with governments and businesses abroad.

iii. Boosting the EU's voice in global data governance

Competing models of data governance are fragmenting the global landscape. The Commission will intensify the promotion of EU approaches internationally, in particular in emerging frameworks, and strengthen coalitions with like-minded partners.

By 2026, in line with the International Digital Strategy[39], the Commission and the European External Action Service (EEAS) will deepen and connect digital partnerships on data governance, aligning with partners that share common objectives and further develop digital trade agreements and digital chapters within traditional trade agreements. It will continue to

[37] Regulation (EU) No 654/2014 of the European Parliament and of the Council of 15 May 2014 concerning the exercise of the Union's rights for the application and enforcement of international trade rules and amending Council Regulation (EC) No 3286/94, OJ L 189, 27 June 2014, p. 50–58

[38] Regulation (EU) 2023/2675 of the European Parliament and of the Council of 22 November 2023 on the protection of the Union and its Member States from economic coercion by third countries (Anti-Coercion Instrument), OJ L 322, 27 November 2023

[39] European Commission and High Representative of the Union for Foreign Affairs and Security Policy, *Joint Communication to the European Parliament and the Council — An International Digital Strategy for the European Union*, JOIN(2025) 140 final, Brussels, 5 June 2025

www.parlament.gv.at

engage actively in fora such as the G7, the G20, the OECD and the UN, using instruments like the OECD 'Declaration on Government Access to Personal Data.'

Particular attention will be paid to promoting EU approaches and mutually beneficial collaboration with candidate countries, potential candidate and closest neighbours. The EU will also work with partners to explore setting up a shared platform for selected high-value public data (e.g. cultural heritage) and pursue trusted arrangements on sensitive data flows, government access, and sector-specific rules. a shared platform for selected high-value public data (e.g. cultural heritage) and pursue trusted arrangements on sensitive data flows, government access, and sector-specific rules.

Flagship actions

- Issuing guidelines to assess fair treatment of EU data abroad (Q2 2026)
- Creating a toolbox to counter unjustified localisation, exclusion, weak safeguards, and data leakage (Q2 2026) and adopting measures to protect sensitive non-personal data (Q3 2026)

5. The Data Union strategy: unlocking data for AI

To ensure competitiveness in the age of AI, the Data Union strategy shifts gears from setting rules to delivering results. Building on the foundations in place since 2020, it tackles data scarcity, regulatory complexity, and global competition.

The European Data Innovation Board will remain the central governance forum, reformed for deeper technical debates and strategic dialogue with Member States and industry. In parallel, the Apply AI Alliance will become the main channel for sectoral feedback, ensuring that companies, researchers, and public actors shape implementation. The AI Observatory will track emerging trends and translate them into policy insights.

Targeted actions will scale up high-quality data, simplify the regulatory landscape, and strengthen the EU's role in global data flows. For SMEs and innovators, this means cheaper compliance, easier access to data, and a more conducive international environment.

Only what gets measured, gets done. This is why the Commission has announced a single market roadmap to increase the pace and speed up the processes. The Data Union strategy can contribute as appropriate to the roadmap to help guide policymakers and industry, in particular SMEs, in removing barriers and completing the single market for data.

Working hand in hand with the Apply AI Strategy, the Data Union Strategy ensures that the EU's data foundations directly power the development, deployment, and uptake of AI across all sectors.

The long-term vision is clear: a sovereign European data economy where data flows securely and responsibly, powering AI, fuelling innovation, and reinforcing competitiveness.

www.parlament.gv.at