INTERNATIONAL DATA SPACES ASSOCIATION

A Global Voice for the EU Data Union Strategy Consultation

International Data Spaces Association | July 2025

International Data Spaces Association (IDSA)

Since 2016, the International Data Spaces Association (IDSA) is an association of over 180 members that advocates for secure, trustworthy and sovereign data sharing, to enable innovation worldwide. Through the volunteer work of our members, we develop the best practices and technology, like the Reference Architecture Model for data spaces, the IDSA Rulebook, or the Dataspace Protocol, that are made available freely for anybody to use. The goal is to establish a global standard for International Data Spaces (IDS) while fostering related technologies and business models for open, federated data ecosystems and marketplaces that ensure data sovereignty for all participants. IDS-based solutions enable trusted data exchange between data providers and consumers.

In an increasingly interconnected world, data is a critical resource that holds immense potential to unlock new insights, drive innovation, and create value, especially in the context of AI. However, for data to generate this value, it cannot remain locked in silos or inaccessible vaults. Sharing, processing, combination, and analysis of data across organizational boundaries must be enabled. The Data Value Chain – from raw data to actionable insight – should be a central concept in the Data Union Strategy. Recognizing and supporting each stage of this chain will help unlock the full economic and societal value of data.

IDSA supports the Data Union Strategy's goals to improve data availability and governance for AI innovation. Flexible data governance, regulatory simplification, and legal clarity are essential to achieving these goals. We offer EU policymakers the following considerations to keep in mind when drafting the upcoming Data Union Strategy.

1. Increasing data availability for AI

As highlighted by Villalobos et al. (2024)[1], the availability of human-generated data for training large language models (LLMs) is increasingly constrained. Given the current trajectory of LLM development and the growing demand for high-quality training data, it is projected that model training will require datasets equivalent in size to the entire stock of human-created data between 2026 and 2032 – or potentially sooner in cases of overtraining.

To meet this demand, data from social media platforms and messaging applications has been proposed as a source, as seen in practices adopted by companies such as Meta since 2025[2]. Another valuable source is video content, where platforms like YouTube allow content creators to opt-in to third-party use of their videos for AI training through transcription[3]. Additional common strategies include the digitalization of analogue sources (such as the large-scale scanning of libraries, e.g. Google Books) or the creation of synthetic data (which is truly valuable only when of high quality, such as in physics simulations).

However, one largely untapped source is corporate data. Due to concerns about trust, confidentiality, and control, companies are generally reluctant to share their data, especially for LLM training. There is the hypothesis that private, industrial data is of significantly higher quality than publicly available data. This quality advantage can lead to more efficient and effective AI training and fine-tuning. Europe should leverage this by promoting a 'smart AI' approach – prioritizing quality over quantity – rather than following data-intensive strategies seen elsewhere.

This is where data spaces can play a transformative role. By enabling secure, governed, and sovereign data sharing, data spaces offer a framework through which companies can contribute to AI development – including LLMs – while maintaining control over their data assets.

Data spaces as key vehicle for AI development

Data spaces are the foundation for building trust in data sharing and can play a significant role in the context of AI. They offer a decentralized, neutral framework of protocols and frameworks that empowers participants to engage in trusted data sharing—allowing data to flow freely while ensuring autonomy for all parties involved.

Our principles of data spaces are simple, yet profound: "Your data, your choice" – participants must always retain full autonomy over how, when, and with whom they share their data and under which conditions. With great responsibility comes great power – agency comes with the responsibility to safeguard and manage that freedom. Equity, decentralization, and neutrality – no participant shall dominate or control the flow of data; all must share equal rights and

[1] https://arxiv.org/pdf/2211.04325
[2] https://about.fb.com/news/2025/04/making-ai-work-harder-for-europeans/
[3] https://support.google.com/youtube/answer/15509945?hl=en

responsibilities. These values form the cornerstone of a robust, interoperable, and adaptable data-sharing ecosystem.

Through free and open standards, infrastructure-agnostic design, and boundless business potential, data spaces serve not as rigid, centralized platforms, but as flexible building blocks for any data ecosystem and business models on top. Our commitment is to foster good faith in all data transactions, balanced by the ability to verify and uphold trust among participants. By adhering to these principles, we believe data spaces will catalyze the next wave of innovation in the context of AI, unlocking vast opportunities for organizations, communities, and society as a whole. We share additional considerations in our Data Spaces Manifesto from April 2025[4].

The creation and sharing of massive amounts of high-quality AI data products demand that all kinds of communities embark in intensive approaches to produce and make available such data products. The development of Data Products for AI ecosystems articulated around data spaces is therefore critical. The need to move from definition to adoption at scale of shared and common formats, schemes for data collection, curation and publication in sovereign data product catalogues is now of critical importance for the development of the AI and data economies in Europe and beyond. A strategy that considers specific actions and instruments for the development and coordination of nurturing and connecting regional data ecosystems under the European common data space framework is required. The need for speed and scale should be elements of critical importance for the implementation of the strategy.

Overcoming barriers to data access

One of the primary challenges is the lack of incentives for data sharing. Many companies are reluctant to share data because they fear losing their competitive advantage, facing legal liability, or surrendering control over their data. Stakeholder engagement is often low due to several factors. First, there can be a misalignment of goals: the objectives of stakeholders may differ from those driving a data-sharing project, especially in complex, multi-stakeholder environments like smart cities. Second, a lack of transparency and involvement can exclude external stakeholders from decision-making processes, leading to distrust and a lack of understanding regarding the motivations behind certain decisions. Third, evolutionary challenges are common: stakeholders' needs change significantly over time, but both technical and organizational data-sharing infrastructures may not adapt accordingly. As a result, stakeholders may stop perceiving ongoing financial or social value in continuing to share their data. All these factors create additional barriers to effective data sharing.

Legal and regulatory complexity adds to the challenge. Companies remain uncertain about data protection, liabilities, and intellectual property issues, which indicates a need for further simplification and harmonization. Technical fragmentation is another hurdle, as the lack of common standards calls for open, modular, and interoperable architectures that can be used across sectors and borders.

[4] https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/The-Data-Space-Manifesto-Version-1.0-April-2025.pdf

Skills and infrastructure gaps also persist in the data ecosystem, highlighting the need for public investment in data infrastructure, tooling, and capacity building. Finally, there is a risk of unintentional sharing of insights that Big Data and AI technologies can infer from shared data when combined with other contextual information. To address these risks, template contracts are recommended as a possible solution.

Top-down approaches should be aligned and complemented by bottom-up initiatives initiated at regional and/or member state levels. Data ecosystems rooted in Data Spaces are fundamental to accelerate and ensure a sovereign and trusted data exchange for AI frontier model adaptation and/or foundational AI model development; particularly in the industrial domain, where fragmentation and heterogeneity are the norm and the highest barrier to any form of general AI development. Specialized Data Regions and Territories across the EU could and should be connected and join forces under the European Data Union to unlock smart specialization and sustainable strategies.

2. Bringing legal clarity, regulatory simplification and trust to stimulate data sharing for AI

Prioritize consistent implementation of existing data rules to stimulate innovation

In the short term, the focus of European data policy should be on the effective and harmonized implementation of recently adopted legislation, including the Data Governance Act, Data Act, AI Act, Interoperable Europe Act, European Health Data Space (EHDS), and other key instruments. With a complex and fast-evolving regulatory environment, ensuring coherence and legal certainty for businesses, public authorities, and citizens is essential. Alignment across these legislative frameworks is critical, particularly where definitions, obligations, and data-sharing mechanisms intersect. To that end, the European Commission should keep helping industry by providing clear, consistent compliance guidance, practical implementation toolkits, and sector-specific FAQs to clarify regulatory overlaps and grey areas.

This need for regulatory consistency and administrative simplicity is also underscored in the Draghi report on the future of the EU single market (2024)[5], which highlights regulatory fragmentation and overlapping compliance requirements as significant barriers to innovation, investment, and competitiveness in the EU. It specifically calls for better coordination of implementation practices and improved regulatory interoperability across digital and data markets. Delivering on this priority would not only reduce compliance burdens for businesses but also help build trust and legal clarity for cross-border data use a precondition for scaling European data spaces and realizing the EU's ambition to become a global leader in the data-driven economy.

The General Data Protection Regulation (GDPR) serves as a positive example of how EU regulation can inspire policy development in other regions. However, with emerging

[5] https://commission.europa.eu/topics/eu-competitiveness/draghi-report_en

frameworks such as the AI Act, it is not certain that EU regulation will be adopted globally. One reason might be the complexity of EU regulation, which can be even more difficult to translate at a global level. Instead, mutual development, alignment and recognition of policies and rules may be more likely.

It is also essential to promote the participation of small and medium-sized enterprises (SMEs) to prevent their exclusion from future data-driven markets and to secure their innovative strength. This involves providing targeted support initiatives, guidance, and tooling solutions – especially regarding open source – to help SMEs navigate regulatory complexities.

Europe requires dynamic, fast-growing companies, yet significant financial resources are often needed to achieve legal compliance, which poses a considerable challenge. While legal sandboxes offer valuable assistance, many businesses still choose to expand in other countries due to these barriers. Though public funding is necessary to foster the co-development of innovative solutions such as open-source projects, it should be complemented by additional measures to support regulatory compliance and further stimulate sustainable growth and market participation.

Address legal uncertainties to unlock data sharing opportunities

The current European data policy framework remains fragmented in its implementation and has, in some cases, introduced additional complexity for companies seeking to share and reuse data, particularly for AI development. Businesses often face legal uncertainty when navigating overlapping obligations across instruments such as the Data Governance Act, Data Act, GDPR, AI Act, and sector-specific rules like the EHDS. For example, the interaction between data portability rights under the GDPR and access rights under the Data Act is not always clear, especially for mixed data sets that contain both personal and non-personal data. This lack of clarity can deter organizations from engaging in cross-sector and cross-border data sharing, which is essential for training, validating, and deploying trustworthy AI systems**.**

Standardization and interoperability enable effective data sharing and the development of data spaces.

Standards should be developed through close, transparent collaboration with industry, ensuring processes are consensus-based, market-driven, fair, and inclusive. This approach should build on existing standards from ISO/IEC, CEN/CENELEC, and other leading international bodies, while respecting the principles set out in the EU Standardization Regulation (1025/2012). In the context of the Data Act, harmonized standards and common specifications should be designed with similar considerations to foster interoperability and legal certainty, relying on the established European standardization system (CEN/CENELEC and/or ETSI), the international framework, or coordinated efforts between them. This will help ensure that European data-sharing frameworks remain globally compatible and aligned at the global level, because standards as a "common language" are unlike regional legal regulations internationally understandable and interoperable. In this context, we welcome to contribute to

the ongoing discussions as part of the "Trusted data framework" and AI Act standardization activities and encourage a swift adoption once finalized.

The development and implementation of EU standards, along with international frameworks often encounter challenges rooted in differing interpretations of fundamental concepts like what constitutes a "dataset." This highlights an urgent need for harmonization of standardization activities, which in turn requires aligned policies across regions. Even when regulations are nominally the same, they can be understood and applied differently both within the EU and abroad. To address this, regulations should be co-developed by industry and governments, following collaborative models like the "one-stop-shop" approach. Moreover, the increasing "technification of regulation" calls for the showcasing of best practices that help interpret rules and establish compliant solutions. The task of translating regulations for implementation is already complex at the EU level and becomes even more challenging when extended internationally, underscoring the importance of coordination and clarity in regulatory design and deployment.

3. Foster an enabling environment for international data transfers

International data flows are fundamental to the modern data economy

Businesses depend on the ability to transfer data across borders to remain globally competitive, develop innovative products, and address challenges that are inherently international in scope. The forthcoming Data Union Strategy should aim to facilitate – rather than constrain – these flows, promoting cross-border data sharing and maximizing the collaborative opportunities it enables. The vitality of Europe's digital economy rests on the strength of its Digital Single Market, which itself benefits from open, trusted exchanges of data, goods, and services with global partners. As European industrial players increasingly operate as global digital actors, they depend on reliable international data flows to scale their operations, develop AI models, and participate in global value chains. It is therefore crucial that EU policymakers continue to advance interoperable frameworks that enable secure and trusted international data transfers – safeguarding privacy and security while ensuring Europe remains a competitive force in the global data economy.

Engaging with non-European regions

Global societies increasingly face the challenge of enabling secure and efficient cross-border data sharing. As digital transformation accelerates, interoperable data ecosystems are essential – not only for competitiveness but also for addressing shared challenges like climate change and public health. In response, Europe would benefit from engaging with non-European regions – especially Japan and Asian countries – shaping standards and regulatory frameworks. Involving partners as stakeholders before development is finalized and well ahead of deployment helps ensure frameworks are globally accessible and applicable. This benefits both European stakeholders and the wider international community. Japan offers a timely example. In June 2025, it launched the Japan Digital Ecosystem Partnership (JDEP), a public-private

initiative for exploring strategies in data collaboration. Structurally, it resembles European efforts like the Data Spaces Support Centre (DSSC)[6] and the Simpl Programme[7], aiming to build or support the building of socially grounded ecosystems through government-industry coordination. Japan's growing expectations for cross-sector interoperability reflect this commitment.

Given these shared goals, cooperation should go beyond government dialogue to include private-sector participation. As a cooperative and technically aligned third party, Japan can constructively contribute to the evolution of European frameworks – supporting global adaptability while respecting regional sovereignty. Early and inclusive collaboration, especially with private-sector involvement, is key to successful dissemination. Frameworks developed in isolation often face delays due to limited mutual understanding. Inclusive design from the outset is essential to building truly global, interoperable data ecosystems.

4. Key recommendations

We suggest the EU Commission provides practical tools to facilitate stakeholders' engagement in data sharing practices with:

- **Data Literacy Training, sector-specific blueprints for the creation of collaborative data sharing projects** (i.e., technical minimal requirements, standard governance frameworks, contractual templates),
- **Development of European Networks of Data Territories** (i.e., territories that unfold (computing) infrastructures and data ecosystems for competitiveness and well-being).

Regarding the regulatory simplification, legal clarity and standardization opportunities, we suggest the European Commission to:

- **Clarify the interactions between regulations:** Provide clear guidance based on use cases on how new rules (e.g., the Data Act, Digital Markets Act, GDPR) interact, especially where obligations overlap or differ,
- **Accelerate the creation and adoption of harmonized technical standards and open interoperability specifications** for data formats, vocabularies, and APIs; use these standards as an instrument to allow for harmonization, interoperability and specification of regulations by referencing,
- **Create legal testing environments** (i.e., data regulatory sandboxes) for the implementation of cutting-edge tools for international data sharing, to create elective forums for discussion among regulators and data stakeholders in key sectors (i.e., health, climate change) and geographical areas (i.e., Global South),

[6] https://dssc.eu/
[7] https://simpl-programme.ec.europa.eu/

- **Provide common guidance on Standard Contractual Clauses for cross-border data transfers** to stimulate the creation of stable contractual frameworks to govern cross-border data transactions,
- **Simplify and harmonize documentation, notification and reporting procedures:**
    - Harmonize and consolidate reporting obligations across different regulations to prevent businesses from facing conflicting or repetitive requirements,
    - Use "One-Stop-Shop" principle: Authorities should request data from other authorities that have already collected it, rather than duplicating requests to businesses. This reduces the administrative burden and streamlines compliance processes (on both sides, for businesses and authorities),
- **Co-develop regulations together with industry and governments**, following collaborative models like the "one-stop-shop". The cooperation should go beyond government dialogue to include private-sector participation,
- **EU policymakers should continue to advance interoperable frameworks** that enable secure and trusted international data transfers – safeguarding privacy and security while ensuring Europe remains a competitive force in the global data economy.

Prepared and published by

International Data Spaces e. V.
Emil-Figge-Str. 80
44227 Dortmund | Germany
https://internationaldataspaces.org

Contact person

Christoph Mertens | Head of Adoption
christoph.mertens@internationaldataspaces.org

Responsible IDSA Working Group

The IDSA Innovation & Policy Forum