BDV BIG DATA VALUE ASSOCIATION

DATA AT THE CORE OF EUROPE'S DIGITAL STRATEGY

BDVA'S INPUT TO THE EUROPEAN DATA UNION STRATEGY

POSITION PAPER – JULY 2025

BDVA welcomes a renewed data strategy for Europe that builds on the foundations of the European Data Strategy 2020[1]. **Data should remain at the core of Europe's Digital Strategy** and the **focus must be on accelerating the speed to generate value out of data for European businesses and society**, making companies and the public sector more efficient, competitive and innovative, creating new business opportunities and equipping professionals and citizens with always more needed strong data skills.

As part of the AI Continent Action Plan feedback, BDVA has called for the development of "AI-data value ecosystems"[2] and for stronger alignment in between data and AI strategies with the spirit of developing **digital ecosystems that accelerate value creation in Europe.** Despite AI and other advancements in digitalisation being the main drivers to update the European Data Strategy, **data remains the cornerstone of innovation, competitiveness and public value creation.** For this reason, updated policies and investments in data (data spaces, data sharing ecosystems[3], data technology and services, data research and innovation, data companies and data skills) are essential not only to support AI and digital strategies in the current context of strategic autonomy, but also **as engine of economical and societal value in Europe**.

The value of the European data market keeps growing and increased by around 10% between 2023 and 2024[4]. Nevertheless, the EU27 is outpaced by global competitors such as China (37.9% growth rate) and the US (18.7%). Moreover, the American market is already four times bigger in absolute value. On the other hand, when it comes to the number of Data Suppliers, the EU demonstrates a higher growth rate and greater absolute growth[5]. For that reason, the European Data Strategy should also focus on **ensuring that the value created with European data remains in Europe.**

[1] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en
[2] https://bdva.eu/download/92/publications/6393/towards-a-european-ai-data-value-ecosystem-v-1-1-2.pdf, p.3
[3] Data space: "interoperable framework, based on common governance principles, standards, practices and enabling services, that enables trusted data transactions between participants" CEN Trusted Data Transaction CWA 18125 (https://www.cencenelec.eu/media/CEN-CENELEC/CWAs/RI/2024/cwa18125_2024.pdf). Data sharing ecosystem: broader concept that includes any other environment where different actors share data, or instruments that support this sharing of data.
[4] European Data Market Study 2024-2026 – First Report on Policy Conclusions (available: The European Data Market study 2024-2026 | Shaping Europe's digital future), p. 31
[5] Ibid.

BDVA's high-level recommendations for the European Data Union Strategy are therefore:

- Data must remain at the core of Europe's Digital Strategy.
- Focus must be on:
  - Accelerating the speed to generate value out of data for European business and society (as described in our paper "Towards a European AI-data value ecosystem")[6].
  - Providing the instruments and conditions so that the value created with European data remains in Europe
  - Making European business and public services more efficient, competitive and innovative and creating new business opportunities in Europe.
    Equipping professionals and citizens with strong data skills.

The EC has structured the Data Union Strategy around 3 objectives as described in the call for evidence[7]: **i) scale-up data use and availability** (with strong focus on data for AI)**; ii) simplification; and iii) international data strategy**. In this paper we provide a set of actionable recommendations linked to these 3 objectives, summarised as follows:

- **Data availability, access and use:**
  - Continue and consolidate public and private investments to achieve sustainable, scalable and interoperable data spaces and data sharing ecosystems.
  - Establish a European coordination body to achieve convergence on current fragmented efforts in setting up data spaces.
  - Introduce incentives for data producers and data intermediaries.
  - Grow the ecosystem of new data companies in Europe.
  - Establish data innovation hubs in all regions of Europe.
  - Invest in synthetic data.

- **Data for AI:**
  - Invest in AI-ready data tools, frameworks, benchmarks and standards.
  - Use Data Labs as a framework to accelerate convergence of AI and data ecosystems.
  - Boost new industrial collaboration models for the creation of vertical foundation models.

[6] https://bdva.eu/news/towards-a-european-ai-data-value-ecosystem/
[7] Ares(2025)4163996 - 23/05/2025

- **Simplification:**
  - Apply a holistic approach to data and digital legislation.
  - Develop automated compliance solutions (RegTech[8]) and new paradigms in regulation for automation.
  - Address legal barriers for research and innovation transfer to market.

- **Development of an International Data Strategy:**
  - Develop and support international standards and protocols.
  - Contribute to the definition of global data governance principles.
  - Invest in collaborative research and innovation, launch pilot projects and support referential lighthouse projects (e.g., International Manufacturing-X[9]).

As the European Data Union Strategy paves the way to the upcoming Multi-annual Financial Framework of the European Union (MFF 2028 - 2034), we expect the recommendations in this paper to help guide future investments and priorities in the domain of data.

Call to Action. Role of BDVA

BDVA is a European non-profit association with over 250 members from industry (including SMEs, startups and large companies), research, academia and the public sector, all united by a mission to create value for society and industry through data and AI innovation. The AI Continent Action Plan and the European Data Strategy are central to BDVA's mission and objectives. As the private side to the EC to the former Big Data Value cPPP (H2020), a private member of the EuroHPC JU, a founding member and strong contributor of the AI, Data and Robotics Partnership in HE and a strong player in the data spaces ecosystem of projects and activities, BDVA actively contributes to achieving all foreseen objectives of the European Data Union Strategy.

This position paper reflects the consolidated input of the BDVA community, gathered through targeted workshops, discussions of the Board of Directors, task forces and expert consultations.

BDVA stands ready to support the EC, Joint Undertakings, Member States and other collaboration partners and associations in turning the European Data Union Strategy into tangible outcomes. **Strong, structured collaboration with community-driven initiatives like BDVA will be essential to ensuring impact on the ground.**

[8] https://en.wikipedia.org/wiki/Regulatory_technology
[9] https://imxc.org/

1. Data availability, access and use

Thanks to the European, national and regional investments in the last few years[10] and the work of many data-oriented associations, data spaces and data sharing ecosystems have heavily emerged in multiple sectors and geographies, addressing multiple value chains and data types. In addition, we can observe many operational data platforms, marketplaces and portals that are evolving to interoperate with data spaces. The EC standardisation request for harmonised standards linked to the Data Act and the work of several data associations[11] and projects such as the Data Spaces Support Centre[12], are fostering the convergence of standards at European level in parallel with international efforts. However, we are still far from achieving the vision 2030 of having a "European-governed data sharing space" stated by BDVA its paper "Towards a European-Governed data sharing space: Enabling data exchange and unlocking AI potential"[13] released back in 2019 and updated in 2020.

Data space and data sharing initiatives under implementation vary widely in form, size, maturity and ambition. They involve diverse actors, data types, funding models, governance approaches and technical architectures often shaped by specific sectoral or use-case-driven needs. Despite this dynamic landscape, challenges around scaling participation of data providers and data users, long-term sustainability and interoperability remain. The landscape is rich and very complex and there is a growing risk of fragmentation, as the rapid expansion of project activities outpaces the overall readiness of the ecosystem to integrate and align them.

Additionally, the EU lacks clarity on how much data is needed to train large AI models to meet their specific performance objectives (data sufficiency) and it is therefore difficult to measure and share progress toward goals related to data availability, visibility and quality. Without clear benchmarks with frequent measurements that are widely shared for data sufficiency, policymakers and potential data providers cannot make informed decisions about when to persist with current policies or pivot to new goals.

[10] Linked to the European Data Strategy 2020
[11] E.g Gaia-X, IDSA, BDVA, National initiatives, etc
[12] https://dssc.eu/
[13] https://bdva.eu/download/76/roadmap-strategic-agenda-and-programme/3232/towards-a-european-governed-data-sharing-space-bdva-position-paper.pdf . The envisioned European data sharing space referred to "a space that is composed of, or connects, a multitude of distinct spaces that cut across sectoral, organisational and geographical boundaries"

BDVA suggests the following recommendations on the way forward to scale availability, access and use of data:

1. Continue and consolidate public and private investments to achieve sustainable, scalable and interoperable data spaces and data sharing ecosystems. Data spaces are being deployed to give data holders/providers/rights holders the ability to manage visibility, access and use of any data item and therefore a well-functioning and interoperable ecosystem of robust, reliable and machine-actionable data spaces would strongly contribute to the European Data Union strategy objective of scaling-up data availability, access and use. We need to keep working in this direction.

2. Establish an independent European coordination body that helps accelerating convergence on current fragmented efforts in setting up data spaces and other multiple existing and evolving data sharing ecosystems and data platforms. Coordination, consolidation, full interoperability and a common understanding of the governance of the ecosystem should be the main objective of this body.

3. Introduce the following set of incentives for data providers and data intermediaries:

a. **Focus on the resolution of problems meaningful to a sector/value chain/ address key societal issues:** Once the main tech and data sharing infrastructure is in place, the funding should focus on the resolution of problems with data sharing (for a sector, a value chain, public or private sector or/and oriented to solve real societal issues). Funded solutions should be reliable, scalable, cross-data space and cross-border and exemplary to massively attract other players.

b. **Support the data-sharing readiness of organisations** (data providers), by offering economic incentives or comparable external support to internal infrastructure investments.

c. **Provide visibility and financial incentives to organisations**, such as data hubs, data intermediaries and operational data platforms, **that actively support smaller players** (e.g., farmers) in accessing and sharing data and data services.

d. **Provide companies with financial incentives to share non-sensitive data.**

e. **Promote Public sector involvement in data spaces as data provider and data user**, not only for open data but for sensitive data as well.

Additional incentives are suggested on the topic of data for AI, in particular linked to the collaboration models.

4. Establish the policy conditions to grow the ecosystem of data companies in the EU with special attention to start-ups:

To further stimulate the European data economy, scale the availability, access, use and value of data, we must not only **attract AI start-ups to data spaces and data sharing ecosystems**, but also **nurture the emerging European ecosystem of data start-ups and SMEs**, in particular:

- TechReg companies
- Synthetic data companies
- Data protection companies
- Data farming companies
- Data space tech companies
- Data intermediaries companies
- Other data services companies

This can be combined with incentives to enable European infrastructure and communication providers, such as the telco industry, to play a central role in the data sharing landscape.

5. Support the establishment of "data innovation hubs" in all regions in Europe: Fully inspired by the BDVA i-Spaces[14], we call for the transformation of EDIHS in "data innovation hubs" and when feasible in full combination with the AI experience centres described in the AI Continent Action Plan. "Data innovation hubs" should exist in all regions in Europe. They should be focused on providing data knowledge, data services and data experimentation infrastructure to their regional ecosystems. This would accelerate scaling participation in data spaces (making more data available, accessible and usable) and would accelerate the adoption of big data and AI technologies required by the Digital decade. By Federating these hubs and integrating them into existing data spaces, we would accelerate data.

6. Advance data literacy, education and skills development programs to scale and ensure inclusive participation in data sharing ecosystems. While awareness of this need has grown, scaling data availability and use requires scaling in the pool of skilled professionals in Europe.

[14] https://bdva.eu/what-are-i-spaces/

2. Data for AI

BDVA has provided detailed input to this objective in its response to the "Data for AI" pillar of the AI Continent Action Plan. BDVA stated that this pillar provides a strong foundation for Europe's AI ambitions, highlighting the importance of facilitating access to trusted and well-organised data in order to unlock the full potential of AI in Europe. The Data for AI pillar of the AI Continent Action Plan also captures some of the essential needs to share and use data across sectors and borders, introduces the "Data Union Strategy" as the response to these needs, leverages on relevant existing data driven initiatives and introduces a new instrument, the Data Labs.

Many of the identified gaps, considerations and opportunities highlighted in BDVA's response to the AI Continent Action Plan equally applies to the European Data Union Strategy and can be summarised as follows:

- As anticipated in previous sections, the strategy should place strong emphasis **on creating incentives (monetary and non- monetary) for potential data providers** (including content creators), to enable scaled-up data use and availability, specifically in the context of data for AI. It should incorporate a clear **economic dimension**, promoting new forms of collaboration, viable business models and the financial sustainability of key stakeholders involved in the AI value chain.

- While high-quality data is important for AI, an even more critical first step is ensuring data is appropriate and valuable for its specific AI-driven purpose (focusing on the right data, not just the highest quality data). The strategy should include **new mechanisms to identify and mobilise additional, more relevant data for AI applications.**

- The strategy should provide the means to **clarify on what constitutes "sufficient data"** to efficiently and reliably train AI models that respond to specific predefined objectives (sectorial or others), **define clear benchmarks** accordingly and support the provision of tools to **enable periodically measurements** about data sufficiency that guide towards the predefined data objectives.

- The strategy should emphasise the **importance of standardisation** to achieve interoperability and trustworthiness when connecting AI and data ecosystems and **the European efforts on this aspect**, especially those towards harmonised standards on Trusted Data Transactions and in support of the AI Act[15] (CEN CENELEC). This can also include other European standards like DCAT (W3C) and NGSI-LD (ETSI).

- Synthetic data plays a crucial role when combining data and AI, with a bidirectional relationship: synthetic generation can be used to complete and enrich datasets for improved AI quality and compliance, while generative AI can be used to produce synthetic data, mostly in specific domains with special scarcity of high-quality data and for underrepresented languages. Therefore, the strategy should **highlight the importance of synthetic data**, by supporting a **comprehensive governance model** for its generation and use, ensuring alignment with privacy requirements, ethical standards and regulatory compliance.

The high-level recommendations listed above are translated into more actionable proposals, organised around three pillars: (i) AI-ready data, (ii) Data Labs as the foundation for AI-data value ecosystems and (iii) Vertical Foundation Models, further detailed in what follows.

2.1 AI-ready data

To fully unlock the power of data for AI, it is not enough to simply make vast amounts of data available. The data must also be of high quality, suitable for its intended use and specifically prepared for use by AI applications. This concept of "AI-ready" data should extend beyond the datasets themselves, also encompassing all the elements of the so-called data product[16], which relies on the FAIR principles to ensure that the datasets can be easily discoverable, understandable, exchanged and consumed by AI practitioners. In this sense, BDVA suggests that the European Data Union Strategy includes the required actions to accelerate this AI readiness at scale. This entails research and innovation actions, deployment, support to standardisation efforts and the necessary legal reviews, including:

[15] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). http://data.europa.eu/eli/reg/2024/1689/oj
[16] https://bdva.eu/blog/ai-ready-data-products/

- **Define of a commonly accepted overarching framework to assess the AI readiness of datasets** and all elements part of the data product, such as metadata description, licences, data quality, APIs and access, regulatory compliance. This framework should be based on established standards, define clear AI-ready criteria with measurable metrics for each aspect and be validated through feedback from AI practitioners and users.

- **Establish a new data quality paradigm tailored for AI, that reinforces the fit-for-purpose approach with AI-centric considerations.** This paradigm should build on existing efforts such as ISO/IEC 5259 series "Data quality for analytics and machine learning" and CEN CENELEC JTC21 (data quality for AI Act). It should cover established data quality dimensions for AI, while evolving to address emerging aspects such as intended purpose, domain knowledge, appropriateness of data, quality of synthetically generated data and other relevant factors.

- However, ensuring AI-readiness goes beyond having a framework, quality criteria, metrics and assessment tools in place. Meeting all these requirements can make data preparation complex and burdensome. Therefore, the strategy should also **support the design and development of innovative, efficient techniques to simplify and automate the preparation of data for AI applications.**

- AI-ready data requires a convergence between the data and AI value chains, introducing an added layer of complexity to data governance. This includes not only "data quality for AI" assessment (as discussed above), but also aspects such as data provenance, lineage and accountability effective for AI and across this combined value chain. The strategy should **support the development of novel tools and methodologies to manage** the complexity coming from **this combined governance**, while also strengthening and leveraging existing instruments, such as data spaces, to enable effective, transparent and trustworthy data governance for AI.

- Preparing data for AI also involves addressing the ambiguity, uncertainty and reluctance of data providers, resulting from the potential responsibilities and liabilities they may face under the AI Act, national and sector-specific regulations and the risks associated with AI applications, including concerns about misuse and accountability. This also includes uncertainty about copyright and intellectual property aspects of the data and content. **The strategy should address these concerns and provide guidance and assurance for data providers when sharing data for AI purposes.**

- Promote the adoption of license types specifically conceived to support the use of data by AI practitioners. These licenses should include terms explicitly designed for AI use cases, offer AI-compatible options for attribution requirements and allow AI-related rights and liability.

- Finally, the strategy should promote **public-private collaboration to map, improve and fund the availability and preparation of high-quality, purpose-fit datasets for AI development**. This includes ensuring clarity around the goals and scope of existing instruments aimed at preparing data for AI, improving understanding their intended purpose and usability, and providing clear guidance on how to leverage available tools. In addition, practical support, such as platforms or services that help users navigate and select the right tools and resources, should be developed to optimise their accessibility and use.

2.2 Data Labs as frameworks to accelerate the development of "AI-data value ecosystems"

The concept of Data Labs was introduced by the EC in its recently published AI Continent Action Plan[17]. Their primary goal is to support AI Factories' users to get the data they need in the format, quality and diversity required, under appropriate governance and technical conditions and to offer additional data services including the provision of secure data infrastructures. Despite acknowledging that the term Data Lab still lacks clarity and, for many, overlaps with existing data space efforts, BDVA welcomed the Data Labs as it offers a first framework for the convergence between data and AI ecosystems.

This section builds upon previous documents[18] and suggests several recommendations that can strongly contribute to the development of the European Data Union Strategy:

- The call for evidence addresses data availability and access in general, but focused on its specific use for generative AI and to "foster a more cohesive, non-discriminatory and trustworthy data and AI market". In this sense, we advocate for Data Labs as the foundational framework for addressing the "last mile" where data and AI can converge, in a combined value chain.

[17] https://digital-strategy.ec.europa.eu/en/library/ai-continent-action-plan
[18] https://bdva.eu/news/towards-a-european-ai-data-value-ecosystem/

For this, they will have to align needs from both data and AI communities, including language, tools and interfaces, in order to attract practitioners and lower the barriers to adoption. Data Labs are initially conceived to provide services to AI factories and therefore to the pre-training and training (or fine-tuning) of GenAI models, but not so much to application deployment, inference and model maintenance (so not applicable to the overall data-driven AI lifecycle). Nevertheless, we think **Data Labs can be taken as a reference to learn from**, in order to extend this "last mile" along the entire AI lifecycle.

- The role of data spaces and other data sharing ecosystems to provide access to data on a cross-border and cross-sector basis, under clear governance rules that guarantee trust in the data sharing process, is essential for the European Data Union Strategy. Therefore, the integration of their capabilities with those of Data Labs in a seamless way emerges as an extraordinary enabler to fulfill the objectives of this strategy, among others "unlock industrial data and data from under-represented languages" to train large AI models. This will require **additional efforts, on top of already ongoing activities, on interoperability at different levels**, including governance, data provenance and liability and business alignment.

- Data Labs are expected to offer a range of services to AI Factories' users, some of which are particularly well-positioned to support the objectives of this strategy:
  - Data discovery, acquisition and collection services from different data providers, including data spaces. By doing so, Data Labs will help facilitate access to valuable datasets, especially to SMEs and start-ups.
  - Data pooling services, enabling the aggregation and integration of datasets from diverse sources to enhance data richness, interoperability and usability.
  - Synthetic data generation services, supporting the creation of high-quality, privacy-preserving datasets.
  - Services to guide and ensure regulatory compliance.

The strategy should **leverage this range of services** from Data Labs for their own objectives, while **harmonising with similar services provided by other data sharing ecosystems**.

- As explained in the previous "AI ready data" section, to prepare data for AI is not a trivial matter and should follow a structured approach. We recommend that **services for this purpose in Data Labs are organised around a comprehensive AI-readiness framework** that includes:
  - Technical data preparation for AI, including tasks such as data enrichment, augmentation, outliers' detection or feature optimisation
  - Tools to bridge metadata models, enabling the conversion of metadata commonly used in data domains (e.g., DCAT) into formats adopted by machine learning communities (e.g., Croissant, ML-DCAT-AP).
  - Data quality assurance, aligning with machine learning-specific standards (e.g., ISO 5259 series) and incorporating fit-for-purpose metrics tailored to AI applications
  - Data compliance mechanisms, ensuring adherence to current and emerging AI regulations, particularly the EU AI Act
  - Licensing models tailored to AI, such as RAIL (Responsible AI Licenses), which provide clarity and legal certainty for AI developers and users
  - Data governance (including provenance and trust)
  - Incentives and safeguards for data providers

- When sharing data for its use from AI applications, responsibilities and liabilities of data providers emerge as a key concern. When talking about incentives, as the value chain enlarges, how to distribute shared value in a fair way is also a major issue. The **interplay between Data Labs and other data sharing ecosystems should foster synergies, avoid overlapping and duplications and ensure collaboration**, in order to eventually lower barriers and provide incentives to all actors in the value chain, especially those at both ends, namely data providers and service end users.

- To fully unlock the value of data in the EU, European Common Data Spaces are implementing strategies to create synergies and interconnect through a federated model that enables seamless data sharing across borders and different sectoral domains. To effectively support the European Data Union Strategy, **this federated approach should be mirrored by Data Labs,** ensuring that users have access to AI-ready data made available by others AI Factories ecosystems.

To address all previous recommendations, we suggest fostering dialogue among the AI Factories on these aspects. The European Commission is stimulating the creation of working groups among the AI factories, which we believe is necessary. As a EuroHPC JU partner focused on data and AI, **BDVA is in a unique position to work with the AI Factories and support them** in the implementation of all the listed recommendations.

2.3 Vertical Foundation Models

The European Data Strategy, published in 2020, wisely identified key sectoral domains as priority areas for unlocking and sharing data. It proposed the creation of sector-specific European Data Spaces to facilitate data sharing within each domain while addressing the unique needs and requirements of their respective stakeholders.

Similarly, Vertical Foundation Models (VFMs) are AI models specifically tailored to address sector/domain-specific challenges. Although still in their early stages, VFMs can offer significative advantages with respect to generic models, among which training datasets specific to the targeted sector and providing optimised performance and more relevant outputs. Especially in some cases, such as language-specific datasets, where industrial data may exist only in small quantities, it is difficult for innovators to train or continuously improve foundation models tailored to European sectors or languages. Besides, given their concrete scope, VFMs would entail a collaborative model that encompasses all actors along the combined AI / data value chain mentioned in previous sections, in this case focused on a specific sector: data providers, data intermediaries, data spaces and sharing ecosystems, AI Factories and AI practitioners.

As an example in the health domain, the European Health Data Space Regulation[19] can offer relevant inspiration, particularly in its governance and technical infrastructure, for enabling the secondary use of electronic health data for VFMs. This is especially true where access to electronic health data is granted for purposes such as developing or innovating products and services, or for training, testing and evaluating algorithms (Article 53(1)(e) EHDSR).

VFMs can provide the following advantages:

- Contribute to fully unlocking the power of data in specific domains.
- Stimulate the availability and use of industrial or sensitive data on this domain (data providers), with a clear purpose / application in mind.
- Provide incentives for data providers coming out from value generation from the VFMs.

[19] 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.
http://data.europa.eu/eli/reg/2025/327/oj

In order to fully achieve these objectives, **BDVA calls for:**

- Identify and support current initiatives in this direction.
- Include VFM as a relevant use case in all sectorial European Data Spaces.
- Provide guidance on the application of competition law to data pooling arrangements in order to have legal certainty for sector-specific collaboration.
- Provide guidance to ensure that industrial data pooled for VFMs appropriately balances protection of intellectual property and trade secrets with the need for openness and reuse within trusted ecosystems.
- Assess whether existing regulatory frameworks, like EHDSR in the health domain, could support the development of VFMs in their specific domains.
- Assess whether certain elements of the EHDSR's secondary use framework could be adapted to other sectors to enable trusted data sharing under restricted-use policies.
- Develop pilots and showcase projects that help to implement all the above and understand the benefits from VFMs.

3. Legal simplification

For this objective, the EC suggests: "1) streamlining existing rules and 2) develop digital tools to reduce administrative burden". The BDVA community stresses the value of **combining efforts in automated compliance solutions (RegTech) and in new approaches for regulators and market players to benefit from that automation**.

On the automation part, we would like to highlight the role of data spaces, that can offer or integrate many of the necessary components to streamline compliance procedures and reduce administrative burdens. From a technical perspective, data spaces can address critical aspects of trust and governance in data transactions[20], many of which must adhere to predefined policies and contractual agreements. Therefore, several technical tools within data spaces can be leveraged or adapted to simplify compliance with regulatory frameworks. Additionally, data spaces incorporate capabilities designed specifically to ensure regulatory compliance[21]. These include features like decision trees that could be automated to simplify complex compliance processes. Finally, data spaces can enable trusted reporting by using standardised, secure formats and, where possible, verified data. A good example in this direction is the Green Deal Common European data space (project SAGE[22]) addressing Corporate Sustainability Reporting Directive and Digital Product Passport.

[20] See https://dssc.eu/space/BVE2/1071255699/Data+Sovereignty+and+Trust
[21] See: https://dssc.eu/space/BVE2/1071253931/Regulatory+Compliance
[22] https://www.greendealdata.eu/

3.1 New paradigms in regulation and RegTech for operationalising regulatory framework

The interest in **automated compliance solutions (RegTech)** reflected in the second objective of the Strategy is very welcomed. While the Strategy's focus on reducing administrative burden through RegTech is important, BDVA sees even greater potential for those solutions.

- Actions should be taken to support the uptake of RegTech **for operationalisation and navigation** of complex regulatory frameworks. RegTech can help with tackling the complexity of simultaneously applicable legal frameworks. BDVA welcomes the embedding of digital reporting measures in existing and future legislation, but similarly believes that **legal automatability and automated compliance would benefit from a more ambitious consideration during the legislative process itself**. A broader commitment by the regulator to consider parallel drafting where appropriate (i.e., the simultaneous drafting of a legal provision and a computational model of the same provision) can increase both the automatability and the fitness-for-purpose of legal norms that govern digital subject matter.

- **Increased research and investment** into the development of RegTech tools and their role in enabling efficient, scalable and trustworthy compliance across legal domains is needed. BDVA will continue working with the community to map and classify different RegTech solutions based on criteria such as stakeholder audience, function, degree of automation, relevant legislation and underlying technologies.

- Effective **legal simplification does not have to be based on amendments** of the currently binding legal acts. For example, in the context of a possible revision of the GDPR[23], it could provide greater value to the data economy if regulators **facilitated tools and standards for automated documentation,** such as RoPAs, rather than simply raising exemption thresholds from 250 to 750-employee enterprises.

- **RegTech tools could enhance cooperation among competent authorities and that way**, improve law enforcement. Without a coordinated framework, competent authorities will continue to face barriers to effective collaboration.

[23] 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). http://data.europa.eu/eli/reg/2016/679/oj

Existing confidentiality requirements already constrain cooperation between authorities such as competition and data protection. The emergence of new competent authorities under recent data and digital legislation risks adding further fragmentation.

- **Development of RegTech solutions should go in parallel with work on semantic technologies**. Expressing and enforcing constraints on provenance, data access control and usage control is a shared challenge across all digital value chains, including DGA[24]/ODD[25] data re-use and B2B data sharing and governance as regulated under the DA[26]/AIA. Semantic RegTech solutions can help meet this challenge by enabling data sharers and recipients to semantically represent provenance information, access and usage policies and key data quality attributes in machine-readable form and to integrate this information along the digital value chain. Since many compliance obligations apply downstream across the digital value chain rather than on a per-entity basis in isolation, research into semantic interoperability will play a key role in ensuring legal compliance in data-driven ecosystems. Semantic solutions such as knowledge graphs could also enable data-driven businesses to query their compliance against authoritative governmental representations of amenable parts the law in knowledge graph form, thereby greatly reducing legal ambiguity and the compliance burden. The importance of open-source paradigms across both middleware and end-user applications (such as dedicated compliance assistance software) in ensuring interoperability, transparency and accountability can be stressed.

The enhanced semantic solutions would have very tangible effects on the utility of data under the scope of the DGA/ODD, as its value is in part contingent upon the ability of public sector bodies to process requests for data access and re-use in a timely manner, as well as upon the ability of potential re-users to guarantee compliance with any imposed conditions for re-use (e.g., ensuring that IPR restrictions and confidentiality in requested data are respected even after transmission to a third country as per DGA Art. 5(10)). Collectively, these measures can help create an integrated, future-proof regulatory environment that facilitates both innovation and compliance across the EU's data economy.

[24] 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 (Data Governance Act). http://data.europa.eu/eli/reg/2022/868/oj
[25] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast). http://data.europa.eu/eli/dir/2019/1024/oj
[26] 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 (Data Act). http://data.europa.eu/eli/reg/2023/2854/oj

3.2 Holistic approach to data and digital legislation

The current legislative framework governing data-sharing ecosystems is widely regarded as complex and fragmented, leading to uncertainty around compliance obligations and responsibilities. This complexity stems from the interplay between horizontal rules and sector- or country-specific regulations, resulting in overlaps and gaps. Additional legal requirements may also vary depending on the type of data and the actors involved, but it is often unclear whom these obligations apply to and under what conditions.

To address these challenges, the Data Union Strategy must:

- **Adopt a holistic approach to the EU data and digital legal framework**. While the consultation rightly identifies the Open Data Directive (ODD), the Free Flow of Non-Personal Data Regulation[27] (FFNPD) and the Data Governance Act (DGA) as priorities for evaluation, this effort should form part of a broader drive to ensure coherence across the entire body of data-related legislation. Revisions should be informed by other key instruments and take into account interactions with adjacent areas of law, such as copyright and competition law. Given the GDPR's role as a cornerstone of the data and AI regulatory landscape, any simplification or amendment must be approached with caution, acknowledging its far-reaching interconnections with other legal frameworks.

- Simplification should not be limited to reducing administrative burdens. It must focus on the **clarification of the interplay between multiple simultaneously applicable legal frameworks**, so that obligations are coherent, understandable and practicable in real-world data-sharing contexts.

In addition, the relevance of private international law warrants further attention and development, for example, in determining applicable law for data-sharing agreements or assessing the legal implications of different organisational models for data spaces.

[27] 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. http://data.europa.eu/eli/reg/2018/1807/oj

3.3 Research and experimentation

As recommended by a recent study of the EU legislative framework, confining research exceptions to non-commercial uses is hardly justifiable in the current research ecosystems, where private-public partnerships are encouraged[28]. This challenge is well illustrated by two examples concerning the training of AI models by research organisations.

First, while research organisations can rely on the text and data mining (TDM) exception in Article 3 of the DSM Directive[29] to train AI models, this exception does not extend to the subsequent commercialisation of the model or its integration into an AI system. If a research organisation collaborates with private entities to market the results of its research, it may lose its status as a research organisation for the purposes of the exception, potentially making Article 4's opt-out regime applicable. Under the DSM Directive, the ability of public-private partnerships to benefit from the Article 3 exception depends on the absence of 'decisive influence' by the commercial partner over the research organisation—for instance, there must be no 'preferential access' to the research results for the commercial entity[30].

Second, Article 3 only covers acts of reproduction during the training phase. It does not address any subsequent acts, such as the communication to the public of outputs generated by the AI model, including situations where copyright-protected material may appear in the output in response to a user prompt.

Similar issues arise under the AI Act. While the legislation is considerate on avoiding that scientific research and development activities on AI models are not negatively affected prior to placing them on the market[31], the full force of the regulation would apply once the fruits of the research find industrial or commercial application and research organisations will be considered providers of general-purpose AI models, thus subject to obligations under section 2 of the AI Act.

[28] Margoni and others, 'Improving access to and reuse of research results, publications and data for scientific purposes' (Commission study, 2024) p 156.
[29] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC. http://data.europa.eu/eli/dir/2019/790/oj
[30] Recital 12 CDSM.
[31] Art 2(8): 'This Regulation does not apply to any research, testing or development activity regarding AI systems or AI models prior to their being placed on the market or put into service'. See also Recital 25 AI Act.

The cascade of compliance measures and the legal uncertainty, which prompts research organisations in partnership with private entities to carry out difficult legal assessment of their status and with little guidance, can represent a significant administrative burden for research institutions, potentially discouraging collaboration. Therefore, attention should be paid in the upcoming Guidelines on AI Act to clarify the extent to which the research organisations can engage in private-public partnerships, without becoming providers under AI Act Art. 3(3) or without falling out of the scope of the exemption AI Act Art. 2(6).

To address these challenges, we support the overarching recommendations of the recent Commission Study and suggest two concrete actions for the Data Union Strategy:

- Develop a clear, unified legal definition of "research data" that clarifies its treatment across different legislative instruments and reduces fragmentation between Member States.
- Harmonise and modernise access regimes for research purposes, ensuring that these regimes recognise the reality that commercialisation is often an integral part of the research value chain and support this transition explicitly.

By clarifying these boundaries and reducing unnecessary legal complexity and uncertainty, the Strategy can better support experimentation, research and innovation in the EU data economy.

4. International dimensions of data. Development of an International Data Strategy

BDVA welcomes the objective of developing an international data strategy as part of the European Data Union Strategy. B2B value chains don't stop at EU borders and on the contrary, in a globalised economy, they are of an international and global nature in most cases. The Digital Product Passport (DPP) is a very clear example of a trans-national initiative that is not unique to the EU, as many other countries and jurisdictions are pursuing similar initiatives.

Introducing additional regulatory restrictions on international transfers of non-personal data, which are typically governed by bilateral contractual arrangements, including provisions ensuring appropriate handling, confidentiality and security of data, does not appear to be the most effective approach to facilitating cross-border data exchange.

BDVA recommends that the international dimension of the European Data Union Strategy is supported by a combination of proportionate safeguards and regulatory clarity, together with the following key elements:

- Development of international standards and protocols and global interoperability principles for secure and trusted data transactions. This also includes the development of a global trust governance architecture that would enable interoperability of data sharing ecosystems with different regulatory environments.
- Definition of data global governance principles in collaboration with other regions of the world and relevant international organisations such as UN agencies, economic and policy groups (e.g., OECD, G20, World Bank) and multistakeholder partners such as the World Economic Forum.
- Launch bilateral pilot projects on data sharing with other regions or countries in the world, driven by real business or societal needs. This would help to address challenges in data exchanges between regions with heterogeneous regulatory regimes.
- Identify and support lighthouse industry-led initiatives setting-up global data-sharing ecosystems such as IMX and the international IMX-council[32].
- Export control measures should be implemented when data (e.g., from Earth observation) may be used in sensitive industries, such as the defense industry. Export control could also serve the purpose of ensuring that when open-source data is used, the services based on it are not shared inappropriately.
- Increase investment in collaborative data research and innovation initiatives with many other countries and regions of the world, including those already covered by the Digital Partnerships (Japan, Republic of Korea, Canada and Singapore), India, Australia, New Zealand, Latin America, Africa and constantly assess collaboration with other countries in the world including middle east, China and US.
- Piloting the international expansion of common European data spaces, including use cases that welcome international industrial users. Supporting the global deployment of data spaces as a common framework for trusted data sharing can position the EU as a global leader in data sharing and data governance.

[32] https://imxc.org/

BDVA
Data, AI and Robotics (DAIRO) aisbl
Avenue des Arts, 56
1000 Bruxelles
Belgium

BDVA.eu
info@bdva.eu