

# Common European Data Spaces: Challenges and Opportunities



Simon Scerri, Tuomo Tuikka, Irene Lopez de Vallejo, and Edward Curry

**Abstract** Common European data sharing spaces are essential for the implementation of the European digital market. This chapter addresses the challenges and opportunities of Data Spaces identified by the Big Data Value Association community. It brings forward five independent goals, convergence, experimentation, standardization, deployment, and awareness, each targeted toward specific stakeholders in the data sharing ecosystem and presents a timeframe when the goals should take place. Furthermore, we have proposed actions based on BDVA recommendations and mapped them over the five goals.

**Keywords** Data space · Data ecosystem · Big Data Value · Data innovation

## 1 Introduction

The digital market is essential for Europe to act concertedly and based on European values, i.e., self-determination, privacy, transparency, security, and fair competition. A legal framework is essential to support the emerging data economy to delineate data protection, fundamental rights, safety, and cybersecurity. One of the EU's key policy deliverables is the harmonization of digital markets. The main tangible document is now the European Strategy for Data, released in early 2020. This will

---

S. Scerri  
metaphacts, Walldorf, Germany

T. Tuikka  
VTT Technical Research Centre of Finland, Oulu, Finland  
e-mail: [tuomo.tuikka@vtt.fi](mailto:tuomo.tuikka@vtt.fi)

I. L. de Vallejo  
DisCO.coop, Bilbao, Spain

E. Curry (✉)  
Insight SFI Research Centre for Data Analytics, University of Galway, Galway, Ireland  
e-mail: [edward.curry@insight-centre.org](mailto:edward.curry@insight-centre.org)

be followed by the Data Governance Act, aiming to foster data available for use by increasing trust in data intermediaries and strengthening data sharing mechanisms across the EU. Also forthcoming is the Data Act, which is a legislative proposal that aims to create a framework that will encourage business-to-government data sharing.

Data space has been an emerging concept but especially outlined by the European Strategy for Data, which guides European activities toward data economy. The strategy goes so far that it names nine common European Data Spaces: Health, Industrial, Agriculture, Finance, Mobility, Green Deal, Energy, Public Administration, and Skills. Naming Data Spaces gives direction but does not unfold the nature or essence of Data Spaces, ecosystems, which may have their peculiarities and emphasis on layers considered common to all Data Spaces. The BDVA (Big Data Value Association) community of experts has been addressing Data Spaces for many years to understand and reflect on the complexity of the concept. Distinctly, BDVA envisions European data sharing space, which refers to a space that is composed of or connects a multitude of distinct spaces that cut across sectoral, organizational, and geographical boundaries. Data Spaces can also be considered an umbrella term to an ecosystem, benefiting data sharing technologies, a suitable regulative framework, and innovative new business aspects.

This chapter presents the position of the BDVA community [1]. It first provides a short overview of Data Spaces in Sect. 2 and the common European Data Spaces vision in Sect. 3. Section 4 dives into the critical challenges standing in the way of expected value generated by the identified opportunities. These challenges are technical, business, and organizational, legal compliance, and national and regional challenges. Section 5 provides an overview of opportunities, in business, for citizens and science, and government and public bodies. As these challenges and opportunities reflect a large community, they are very informative and manifest their concerns and hopes over the vision. Section 6 presents BDVA's call to action. BDVA has identified five recommended preconditions for successfully developing, implementing, and adopting a European data sharing space: convergence, standardization, deployment, experimentation, and awareness. These recommendations have been translated with the BDVA community into 12 concrete actions, which are placed in a suggested timeframe until the year 2030. Actions can be aligned with the implementation roadmap of Horizon Europe and Digital Europe Programmes. Finally, the chapter concludes in Sect. 7.

## 2 Data Spaces

Data Spaces is an umbrella term corresponding to any ecosystem of data models, datasets, ontologies, data sharing contracts, and specialized management services (i.e., as often provided by data centers, stores, repositories, individually or within “data lakes”), together with soft competencies around it (i.e., governance, social interactions, business processes). These competencies follow a data engineering

approach to optimize data storage and exchange mechanisms, preserving, generating, and sharing new knowledge. In comparison, data platforms refer to architectures and repositories of interoperable hardware/software components, which follow a software engineering approach to enable the creation, transformation, evolution, curation, and exploitation of static and dynamic data [2, 3] in Data Spaces. Although distinct, the evolution of the two concepts goes hand in hand and needs to be jointly considered, as both can be considered the two faces of the same data economy “coin.” Their complementary nature means that commercial solutions often do not distinguish between the two concepts. For example, the Siemens MindSphere platform relies on MS Azure data solutions; the Amazon solutions embed the EC2 applications (as the platform) and the S3 storage (space) services. Furthermore, due to the particular requirements for the preservation of individual privacy, a distinction between technology and infrastructures that store and handle personal and other data has emerged. The evolution of industrial data platforms (considered key enablers of overall industrial digitization) and personal data platforms (services that use personal data, subject to privacy preservation, for value creation) has continued to follow different paths [4].

### 3 Common European Data Spaces Vision

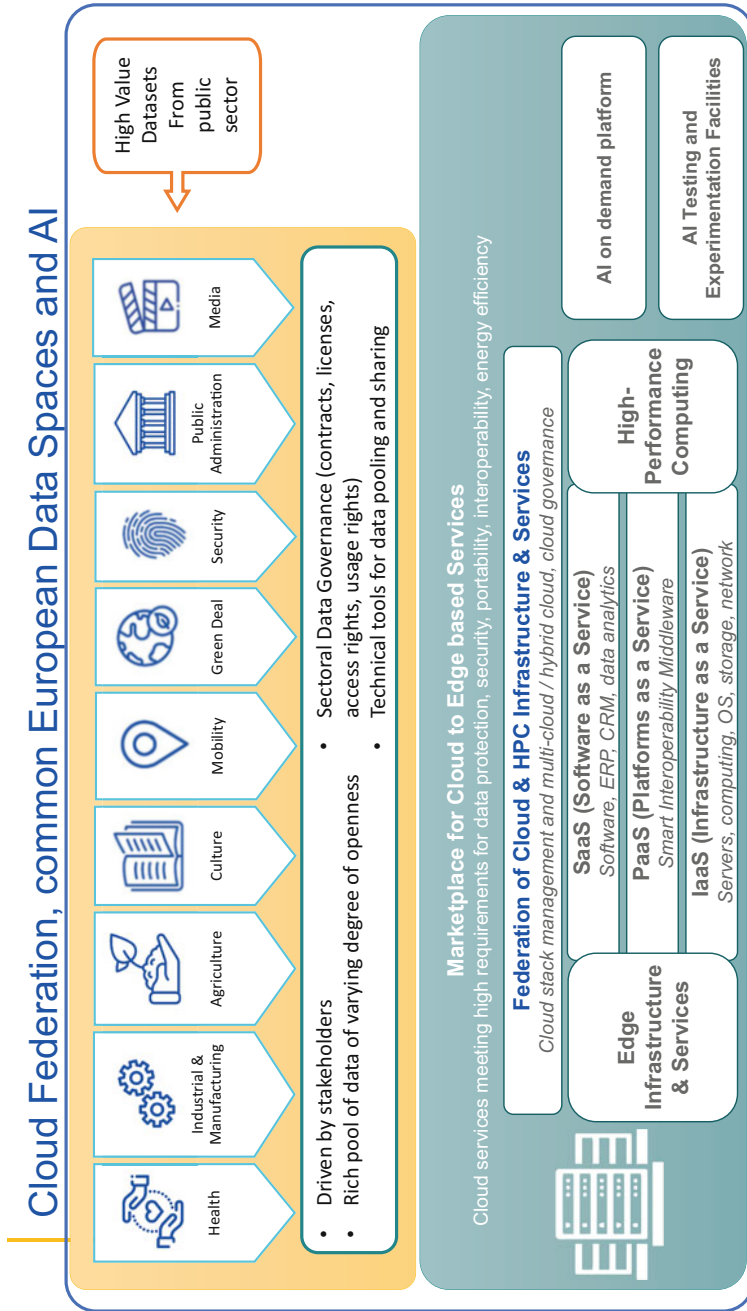
The European strategy for data aims at creating a single market for data that will ensure Europe’s global competitiveness and data sovereignty. The strategy aims to ensure:

- Data can flow within the EU and across sectors.
- Availability of high-quality data to create and innovate.
- European rules and values are fully respected.
- Rules for access and use of data are fair, practical, and clear, and precise Data Governance mechanisms are in place.

Common European Data Spaces will ensure that more data becomes available in the economy and society while keeping companies and individuals who generate the data in control [5]. Furthermore, as illustrated in Fig. 1, common European Data Spaces will be central to enabling AI techniques and supporting the marketplace for cloud and edge-based services.

### 4 Challenges

The BDVA community has identified the most critical challenges (see Fig. 2) that stand in the way of the expected value generated by the identified opportunities [1]. The challenges can be categorized into two main concerns: inter-organizational (lack of suitable data sharing ecosystems) and intra-organizational (issues faced by data producers and consumers, as data sharing participants).



**Fig. 1** Overview of cloud federation, common European Data Spaces, and AI [5]

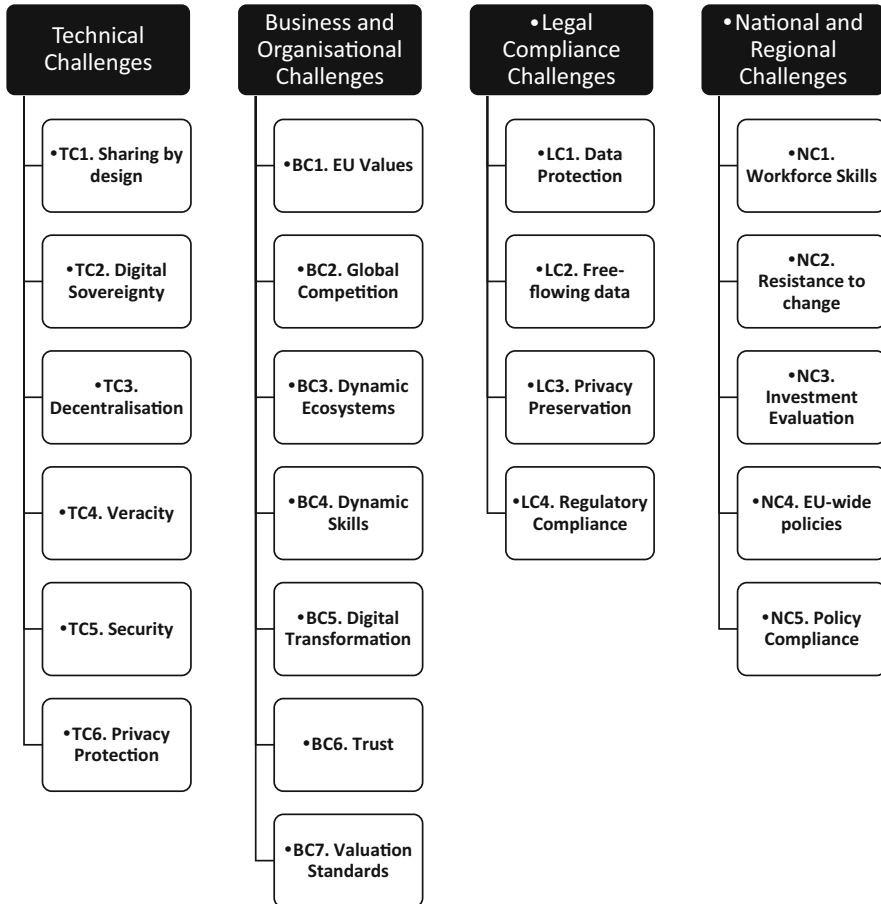


Fig. 2 Overview of challenges for common European Data Spaces

The most pressing inter-organizational concern remains the lack of valuable and trustworthy data sharing ecosystems that inspire immediate large-scale participation. Primary causes include the lack of robust legal and ethical frameworks and governance models and trusted intermediaries that guarantee data quality, reliability, and fair use. This is compounded by the lack of widespread adherence to emerging best practices and standards (e.g., interoperability, provenance, and quality assurance standards), whose maturity pace also continues to fail expectations. From a technical point of view, data sharing solutions need to address European concerns like ethics-by-design for democratic AI, and the rapid shift toward decentralized mixed-mode data sharing and processing architectures also poses significant scalability challenges.

In terms of intra-organizational concerns, the first significant concern is the difficulty determining data value due to a lack of data valuation standards and

assessment tools. This problem is further compounded by the highly subjective and party-dependent nature of data value and the lack of data sharing foresight exhibited by most producers. The second concern revolves around the difficulty faced by data producers balancing their data's perceived value (after sharing) against risks exposed (upon its sharing) despite adhering to standard guidelines. Specific examples include the perceived loss of control over data (due to the fluid nature of data ownership, which remains hard if not impossible to define legally), the loss of trade secrets due to unintentional exposure or malicious reverse engineering (in a business landscape that is already very competitive), and the risk of navigating around legal constraint given potential data policies breaches (including GDPR and exposure of private identities).

The subsections below offer different perspectives to these challenges but should be considered neither exhaustive nor in isolation. In addition, the fact that some challenges are more prominent in specific sectors is well known and should also be taken into consideration.

#### **4.1 Technical Challenges**

The ambition to realize a cross-border, cross-sectoral sharing data space and enable platforms to process “mixed” proprietary, personal, and open public data introduces new technical challenges and compounds existing ones. The impact of known challenges (e.g., the Vs of Big Data: volume, velocity, variety, veracity) along the data lifecycle needs revisiting following the arising opportunities for data sharing which, in addition to conventional raw data and its transformations along the processing chain, also extend to metadata, models, and processing algorithms. The main challenges are:

- **TC1. Sharing by Design:** Data lifecycle management is not designed around sharing. Most data producers do not yet consider data sharing as a possibility at the data creation stage. Existing data lifecycle management models need to improve how they incorporate all relevant processes, including preparing data for sharing and finding the correct data. The maturity of data services (e.g., cleaning, aggregation) in data sharing ecosystems is as crucial as the availability of the data itself; without them, the data economy will not establish. Furthermore, the above differentiation between multiple kinds of data that can be made available for sharing also raises the complexity of the “variety” challenge, and interoperability solutions need to address this change.
- **TC2. Digital Sovereignty:** Enforcing data usage rights. The realization of a mixed data sharing space will only materialize if data producers are guaranteed to retain their rights as the original owners, enabling them to control who can use their data, for what purpose, and under which terms and conditions. Different ownership models or suitable data rights management frameworks need to be further explored to guarantee digital sovereignty.

- **TC3. Decentralization:** Decentralized data sharing and processing architectures. The need to guarantee that data producers control their data results in setups that forego data porting favoring decentralized data storage architectures. Thus, discussions on data volumes and data velocity (data streams) need to increasingly consider both the scalability of real-time operations over-dispersed data-at-rest in undetermined geographical distributions and the distributed processing of data-in-motion, which requires no intermediate storage. Standard data exchange protocols in decentralized architectures are therefore increasingly required.
- **TC4. Veracity:** Weak verification and provenance support. Data veracity remains crucial for the sustainability of data sharing ecosystems. Data in various stages processing chain will need to carry traceable information about its origins and operations (i.e., metadata about its raw form, algorithms, and operations it was subjected to). Support for advanced provenance is required to raise trust.
- **TC5. Security:** Secure data access and restrictions. To enable a trusted network within which closed (proprietary, personal) data can be unlocked for exchange and sharing, issues like confidentiality and digital rights management must be addressed appropriately. Furthermore, secure access control needs to be guaranteed even within a decentralized peer-to-peer network. Therefore, security solutions and exchange protocols must be standardized across all data sharing space nodes and participants.
- **TC6. Privacy Protection:** Maturity of privacy-preserving technologies for Big Data. Although technical solutions for secure and trustworthy data sharing (e.g., privacy-enhancing and privacy-preserving technologies, including digital identity management) are in continuous development, continued investment is required toward their further maturity to increase their reliability and uptake. More flexible ways of allowing uptake of compliance solutions also need to be explored.

## 4.2 *Business and Organizational Challenges*

Challenges in this category have predominantly been observed in industrial applications (non-personal data sharing). The clarity of the GDPR means that Industrial Data Platforms (IDPs) can potentially be extended to handle de-personalized data for added business value. However, legal compliance constitutes a challenge in itself and is covered in the next section. Thus, foreseen business challenges are related to the socioeconomic sustainability of a pan-EU IDP connecting multiple Data Spaces and offering data marketplaces and include the following:

- **BC1. EU Values:** Difficulty establishing EU IDPs in the global market. EU-designed IDPs need to adhere to values such as democracy, open competition, and egalitarian treatment. These characteristics can distinguish them in the global market and eliminate questionable “shortcuts” to the advantage of global competitors. In addition, new business models need to demonstrate clear business

value in adhering to EU values and their advantage over existing commercial solutions in this setting.

- **BC2. Global Competition:** Competing in the global market through product-service platforms. The combination of data and service economies represents a significant competitive advantage of the EU in the global marketplace. Thus, value-added data-driven services that could make “Made in EU” products competitive globally must be identified. Furthermore, SMEs (99% of the EU industrial fabric) and the role of PPP mediators like the Digital Innovation Hub (DIH) need to be further clarified, and co-opetition models need further investigation.
- **BC3. Dynamic Ecosystems:** Implementing Data Spaces in dynamic business and data ecosystems. In the industrial domain, shared data ecosystems must guarantee data producers complete control over the access and usage of their data. However, ownership is difficult to define legally (see related legal compliance challenge, LC2). Furthermore, there are no clear guidelines or consensus on implementing data sovereignty in flexible and dynamic business ecosystems (rather than in static hierarchical supply chains). It is also unclear how next-generation peer-to-peer networks can guarantee trust and sovereignty without centralized control.
- **BC4. Dynamic Skills:** Effects of disruptive technology challenges on the job market. There are differing views on the exact impact of new data-driven technology and automation on jobs and employment. Short-term actions include the re-skilling and up-skilling of personnel. However, a complete re-definition of workflows, processes, and human-machine interaction patterns (e.g., “collaborative intelligence” between humans and autonomous systems) is required in the longer term. In addition, the current education system is still not geared toward continuously catering for new and unknown professions.
- **BC5. Digital Transformation:** Challenging organizational impact of the 6Ps digital transformation model. Data-driven transformations are needed at the level of products (and services), processes (and organizations), platforms (and spaces, marketplaces), people (and roles), partnerships (and participatory innovation models), and performance (and data-driven KPIs). Methods and tools to support EU industry with this transformation are required. Although disruptive innovation models benefit start-ups and scale-ups, evolutionary innovation models need to be considered alternatives for the broader industrial ecosystem.
- **BC6. Trust:** Lack of data sharing trust and motivation. Data marketplaces rely on an understanding of the commercial value of data produced by industry at all levels. The lack of confidence in the quality of data available for sharing is in itself a challenge. Without quality standards, widespread, automatic data exchanges will not materialize. Attempts at optimizing data accuracy should also extend to algorithms (e.g., algorithm bias). In addition, preparation costs for sharing data (e.g., cleaning, quality assurance) need to be considered, together with risks (e.g., potential access to trade secrets, intellectual property sharing). In addition, sharing personal data in B2B applications needs to comply with the GDPR strictly. The effective application of anonymization and obfuscation meth-



ods can minimize both risks. However, it can generate increasingly synthetic data whose classification can be objective and whose actual value can be extremely low (e.g., critical applications), thus not offering producers an immediate return on investment. Open data models can become a new industry model if the value of open innovation, early involvement of customers, and strategic alliances (even with competitors, as in the case of service ecosystems) are well understood. To set up trusted data networks, ad hoc and on-the-fly B2B data exchange mechanisms and contracts, provided under well-defined data sovereignty principles, must be considered.

- **BC7. Valuation Standards:** Lack of data valuation standards in marketplaces. Data marketplaces introduce new opportunities and business models whose center is valorization or monetization of data assets. New challenges revolve around the pricing of data, e.g., whether this is determined by the producer, by the market demand, or by a broker or third party: whether the value for a specific data asset is universal or depends on the buyer-seller relationship. Guidelines and pricing models need to be established to guide businesses in determining value in participation (refer to last business challenge). New forms of value creation uncovered by new sharing mechanisms need to be explored. In addition, data discovery will need to be better addressed since the value of data assets is materialized upon findability.

### ***4.3 Legal Compliance Challenges***

All the different regulations introduced in the last decade in the context of the digital single market make for a complex landscape of policy for data. However, notwithstanding the inherent complex role of data, an increased understanding is needed of how data regulation interplays and connects within data platforms. The following are the most pressing challenges that need to be addressed:

- **LC1. Data Protection:** Tackling inverse privacy and understanding personal data rights. There is a significant gap between the rights introduced by the GDPR (and its 28 national implementations) and the average citizens' and companies' understanding of their implications, what control they can exercise, and how. New business models should not assume that a sufficient portion of private users have the time, expertise, and interest to comprehend these implications fully, but data rights and consent can change. The practice of inversely private data should be discouraged to provide means for individuals to retain control and access to their private data at all times. More guidance is needed from regulators and data platform developers. Developments such as sticky policies and dynamic user consent offer promising avenues (refer to the related technical "Privacy Protection" challenge).
- **LC2. Free-Flowing Data:** Ownership and other legal blockers. Although we speak of the free movement of data as a fifth European freedom, data is far from

flowing freely. Legal questions surrounding data ownership, access, portability, and retention remain pressing topics of attention, even more so in an AI context. Existing legislation (e.g., database rights) are outdated, hampering the use of data in AI and the development of new business models. Furthermore, data ownership is tough to address in a data marketplace setting, as it is difficult to define legally. In the absence of a “GDPR for non-personal data,” the principle of data sovereignty can be an answer to confidentiality and security requirements but also poses implementation challenges (see TC2).

- **LC3. Privacy Preservation:** Privacy preservation in an open data landscape. Open data initiatives and public blockchains are driving open innovation in multiple ways. Privacy preservation in this openness is a topic that has to be carefully examined not only in technical terms but also regarding legal compliance at national and European levels.
- **LC4. Regulatory Compliance:** General uncertainty around data policies. Data-driven SMEs and companies that aim to develop data platforms still face questions on how to incorporate and adjust for the effects of the regulatory landscape within the digital single market, e.g., how to be compliant; when, where, and which regulation comes into effect; and how to gather knowledge on implementing the regulation.

#### ***4.4 National and Regional Challenges***

Industry and academia adopt new and disruptive technology much faster than member states, and the European Commission can adapt their policies and regulations. Amid an emergent data economy facilitated by the convergence of digital technologies, these challenges need to be high in the political agenda:

- **NC1. Workforce Skills:** Public organizations lack digital skills and resources. Digital technology is developing fast, and public organizations have difficulties keeping up with the pace of development (perhaps more so than business; see also business challenge BC4). At the same time, it is difficult to identify what kind of new skills and education public organizations would need. For instance, new digital skills include planning how data is organized and creating value in society. Organizational and individual skill development are also budget issues, which may not be high on the public agenda. The challenge is to use funding wisely and to avoid waste of resources.
- **NC2. Resistance to Change:** Insufficient support for digital transformation in business by public authorities. Digitization will transform processes, and data, along with AI, will build up knowledge of society. Transforming the organization leads to changing personnel’s work profiles. Roles will change, leading to employment disruptions and the need for re- and up-skilling. New services are an opportunity, but resources for the transformation are limited. Efficiency and

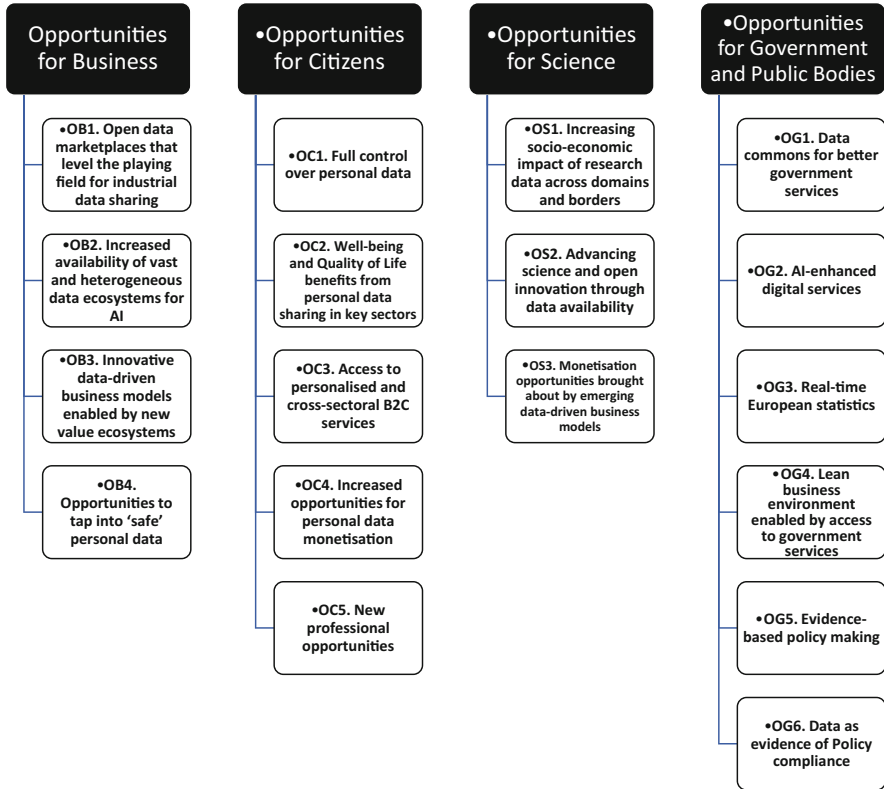
transparency need data sharing but also investments in order to create new Data Spaces.

- **NC3. Investment Evaluation:** Evaluating public organization efficiency and economic impact in the data era. Public organizations serve society, both citizens and industry alike. The constant requirement of efficiency and impact improvement motivates governments to find out new services based on data. However, decisions on development investments are difficult to make, and quite often, investments are considered risky. Nevertheless, public organizations and their services are an essential part of society and one of the starting points of emerging data ecosystems. From a governmental point of view, the challenge is to evaluate investment in data-centric organizations and ensure that economic results impact the whole society.
- **NC4. EU-Wide Policies:** Lack of common innovation policies. Stepping up from regional innovation policies to EU-level comparisons is challenging. Data provides a means to measure the impact of innovation policies, but regions find it difficult to compare due to varying requirements. For instance, simple dataset timescale variation may give odd results depending on the region.
- **NC5. Policy Compliance:** Translating European-wide policies into tangible measurements. To enable the possibility of real-time, data-driven policy compliance verification, further investments in infrastructure and the certification of data from devices such as IoT appliances and edge nodes are required. Furthermore, when data is needed as evidence for compliance with specific regional and European policies, standard or common approaches recognized and accepted by the respective policies are required to map data, e.g., from IoT device measurements, into compliance levels.

## 5 Opportunities

As indicated in the previous section, in recent years, considerable interest has been observed by major industrial players, national and European legislative institutions, and other key stakeholders in:

- Alignment and integration of established data sharing technologies and solutions, avoiding reinventing the wheel and supporting scale.
- **Architectures, standards, protocols, and governance models aiming to unlock data silos**, over which (a) fair and secure data exchange and sharing take place, (b) protection of personal data is paramount, and (c) distributed and decentralized solutions enabling new types of data value chains can be explored.
- **Business models that can exploit the value of data assets** (including through the implementation of AI) bilaterally or multilaterally among participating stakeholders that are not limited to industry but include local, national, and European authorities and institutions, research entities, and even private individuals.



**Fig. 3** Overview of opportunities for common European Data Spaces

- **Fostering and accelerating the uptake of data technologies** and the data economy within sectors whose business models are not yet data-driven.
- **Enabling data analytics across a European data sharing ecosystem**, including research centers, industry, government, and multi-national bodies leveraging existing pan-European initiatives and networks (e.g., DIH and i-Spaces).

This section extrapolates current trends to outline opportunities arising over the next decade for common European Data Spaces. As illustrated in Fig. 3, the opportunities are categorized into four primary societal spheres of industry (business), private users (citizens as customers), research and academia (science), and local, national, and European government and public bodies (government). However, the identified opportunities have a broader socioeconomic value, often benefitting multiple spheres, and should therefore not be considered solely within the confines of these categories. Indeed, the possibility to boost the data economy will lead to an improved general economy, thus benefiting society.

## 5.1 *Opportunities for Business*

Both SMEs and large industries stand to gain from the following emerging opportunities:

- **OB1. Open data marketplaces that level the playing field for industrial data sharing:** The emergence of large-scale marketplaces whose participation is open to all kinds of data producers and consumers will complement the currently dominant, exclusive data sharing agreements and solutions designed for specific parties. Thus, industrial data can be shared inside and outside of a value network, with guarantees that the producers retain data sovereignty and receive fair compensation. European industrial players of all sizes, who provide both tangible and immaterial services, can tap into data sources that are made available within the rules set by the data producers.
- **OB2. Increased availability of vast and heterogeneous data ecosystems for AI:** Advanced AI applications play a fundamental role in several industries' most critical business processes. Availability of valuable datasets is key for unleashing the potential of AI added value, especially in key industries like business services, manufacturing, wholesale, retail, and infrastructure providers (e.g., 5G operators). Emerging data marketplace infrastructures enable data to be shared and traded in a trusted, secure, and transparent manner that respects ownership. Blockchain technology enables data sharing at scale. Federated analytics on distributed data sources enable the sharing of derived insights without sharing the original data, thus balancing privacy, autonomy, and IP protection. An example of this approach is behind the Collaborative Cancer Cloud. These developments will make data available for AI development in new ways, offering unprecedented opportunities to scale the ecosystem.
- **OB3. Innovative data-driven business models enabled by new value ecosystems:** Moving from “data for business intelligence” to “data for AI” also means transitioning from internal processes to more collaborative and participative cross-domain ecosystems. The most innovative data-driven business models show a wide variety of value creation possibilities, from direct data monetization to access-based valorization of data assets on sharing platforms. Mobilizing data from business processes and services can offer a myriad of new opportunities, where data-driven often also implies engagement with strategic partners and the development of data ecosystems around these opportunities. A prominent example is a drive by the Mobility Open Blockchain Initiative's (MOBI, [www.dlt.mobi/](http://www.dlt.mobi/)) consortium to accelerate the data sharing for autonomous vehicle technology.
- **OB4. Opportunities to tap into “safe” personal data:** The GDPR sets precise requirements for the safe use of de-personalized data outside its original purpose. Personal data will increasingly be considered for cross-sectoral applications following explicit consent and true anonymization (a technical challenge). Driven by the private customers' use of digital services, cross-sectoral services will continue to emerge. The banking industry, for example, was forced to rethink

its market position following the Payment Services Directive, and based on transaction data, new services are being provided across insurance, banking, and health service providers.

## 5.2 *Opportunities for Citizens*

European citizens will benefit from easy and secure data sharing in various ways:

- **OC1. Full control over personal data:** Under the GDPR, data platforms must guarantee legally compliant data privacy and sovereignty, affording individuals higher control and traceability of their data. In this ideal scenario, individuals will be able to monitor what data they share, which space it is stored in, and who can access or use it while retaining the right to alter these decisions. However, in addition to the management of inversely private data, which remains a challenge in itself, the need for individuals to fully comprehend the implications of these rights remains. Therefore, the BDVA is active in gathering and disseminating cutting-edge developments in the area of privacy-preserving technologies for Big Data, contributing to a better digital single market and increased end user data protection.
- **OC2. Well-being and quality of life benefits from personal data sharing in key sectors:** GDPR-compliant data sharing platforms enable emerging European technology to perform data analysis for a clear personal (or social) benefit in important sectors such as health. Furthermore, the safe management of legally compliant personal health data records allow for broader analysis (see “data for research” opportunity, next section) of health, wellness, and life data for improved understanding, risk identification, and prevention of diseases directly benefiting private users.
- **OC3. Access to personalized and cross-sectoral B2C services:** Digitization brings production closer to consumers, offering more choice and personalization independent of geographic location. Interoperable data ecosystems are required to enable bundling services during digital transactions. The business opportunity introduced by tapping into personal data will benefit private citizens, e.g., banking and insurance services offering consumers alternative product sales and choices between the most cost-effective options.
- **OC4. Increased opportunities for personal data monetization:** New European legislation incentivizes individuals to share their data, introducing C2B business models that allow them to remain in control of their data while directly receiving fair monetary or economic benefits.
- **OC5. New professional opportunities:** Further innovation will define new career pathways and generate additional jobs whose education, re-skilling, and up-skilling will continue to be supported by national and regional authorities that understand their long-term value.

### 5.3 *Opportunities for Science*

Academia is expected to benefit from the following opportunities:

- **OS1. Increasing socioeconomic impact of research data across domains and borders:** Converging standards for data and metadata representation, sharing models, licensing, and exchange protocols will make it increasingly easier to discover, integrate, or otherwise jointly process and analyze data in other scientific domains. This broadens collaboration opportunities between different scientific fields and promotes the value of making generated data available, at least partially, as open data, for the broader good. In addition, the establishment of appropriate guidelines can promote collaboration between scientific and national bodies to address societal challenges better, effectively reducing data access bottlenecks faced by European researchers.
- **OS2. Advancing science and open innovation through data availability:** Access to data for research purposes remains limited since companies need to invest effort in preparing data for little or no apparent gain. After clear business incentives for data exchange and digital infrastructure that removes friction in the process are in place, little additional effort will be required to make the data available (different license agreements) to academia. In return, it will be easier to involve researchers with access to this data in open innovation activities, thus accelerating innovation in companies. Due to the data access conditions, data challenges remain common in academia. However, while platforms such as Kaggle are very successful, they tend to be limited to specific challenges with little flexibility in the evaluation metrics. An increase in data availability enables academics to design and run more complex challenges, thus improving scientific crowdsourcing to advance science and identify solutions benefitting industry. This can help eliminate the imbalance between European and non-European industries when providing data to scientists as a core part of their business; examples from the USA and China include Google, Amazon, and Alibaba. Aside from industry data, science will also benefit from the GDPR-enabled availability of de-personalized “personal” data from millions of European citizens, offering new research opportunities in societal challenges such as healthcare and transport.
- **OS3. Monetization opportunities brought about by emerging data-driven business models:** Providing controlled access to research data will enable scientists, universities, and research institutes to have the opportunity to exchange or monetize their research data by making it available in a controlled way to other institutes and companies. This will strengthen cooperation between research and industry, enable different data to be integrated and analyzed, and thus introduce new revenue opportunities for academia.

## 5.4 *Opportunities for Government and Public Bodies*

These opportunities will benefit all levels of government, national, and European public authorities:

- **OG1. Data commons for better government services:** Opening public domain datasets and systems offer opportunities to improve existing services or develop new ones. Moreover, it can increase accessibility and simplification of e-Services. The Single Digital Gateway Regulation (SDGR) promises to make company and citizen data available across Europe in a one-stop-shop manner. The once-only principle makes it easy for companies to settle and set up businesses across Europe. Harmonizing and opening up government data also enables governments to act as data platforms offering digital services to both companies and citizens.
- **OG2. AI-enhanced digital services:** AI-boosted digital services can help predict and analyze national and European data in a privacy-preserving and ethical manner. Collaboration, piloting, and information sharing between government agencies support data platform exploitation. Based on working national examples, EU government bodies can form information sharing Data Spaces to support emerging EU-wide data service management. For example, Finland has a network of government agencies to share best practices of introducing AI to organizations.
- **OG3. Real-time European statistics:** An integrated European data sharing space can provide real-time monitoring across key sectors at both national and EU levels. Examples include economy, security, and health: customs statistics can monitor imports and exports and help with cross-border security, whereas health data can make disease outbreaks visible to all relevant healthcare authorities.
- **OG4. Lean business environment enabled by access to government services:** Public governmental services can be connected with industrial data for leaner business planning. For example, financial planning can be based on real-time information on the effects of rules and taxation regulations. In addition, better-integrated information systems enable automation of taxation, leading to cost-saving and predictable national budgets.
- **OG5. Evidence-based policymaking:** Data for policy, sometimes also referred to as evidence-based policymaking, uses Big Data in the policymaking process. It allows policymakers and governmental bodies to use public sector data repositories and collaborate with private actors to improve and speed up policy cycles and explore new areas of policymaking in a data-driven way.
- **OG6. Data as evidence of policy compliance:** Policymakers and regulators will increasingly depend on data-driven policy compliance solutions. In a data-driven society, traditional compliance mechanisms are challenged due to the increasing velocity and complexity of regulating cyber-physical systems. Advances in open data, the Internet of Things, and edge computing create a wide array of new data to be used by public organizations, smart cities, manufacturing and production lines, and logistics. These data can serve as evidence for validation of

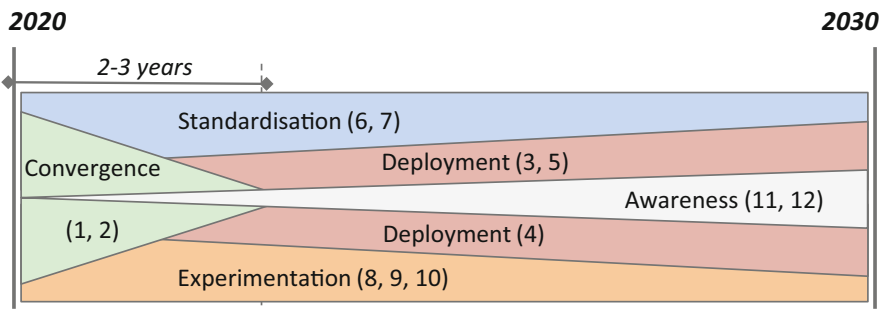


whether specific policy-related conditions, implied by European-wide regulation and policies, are met. This can open new horizons on how certifications on organizational procedures can be provided. Automating compliance and sticky policies [6] can already be witnessed in APIs or blockchain for smart contracting in data markets.

## 6 Call to Action

BDVA has identified five recommended preconditions for successfully developing, implementing, and adopting a European data sharing space [1]. Following widespread consultation with all involved stakeholders, the recommendations have been translated into 12 concrete actions. These can effectively be implemented alongside the Horizon Europe and Digital Europe Programmes [7]. This call for action is aligned with the European Commission’s latest data strategy [5]. The recommended actions are categorized under five independent goals, convergence, experimentation, standardization, deployment, and awareness, each targeted toward specific stakeholders in the data sharing ecosystem. The implementation of the five goals should take place within the timeframe shown in Fig. 4. Assuming the convergence of initiatives required over the next 3 years will yield satisfactory outcomes, deployment efforts can be scaled up with experimentation acting as a further catalyst. Other deployment efforts need to go hand in hand with intensified standardization activities, which are key to a successful European-governed data sharing space. Activities targeted at greater awareness for all end users can initially target organizations, entities, and individuals that can act as data providers and then extend to all potential consumers as solid progress is achieved.

To catalyze the convergence of existing national and regional concepts, efforts, priorities, and strategies:



**Fig. 4** A suggested timeframe for implementing the recommended actions over the next decade (x-axis). The y-axis illustrates the relative, rather than absolute, effort distribution between the five phases. The absolute effort value is expected to fluctuate, reaching a peak around the 3–5-year mark [1]

- **A1:** Launch coordination actions to map existing initiatives at EU, member state, regional, and municipal level (leveraging existing EC CSAs) and identify the common building blocks to create an impact and foster standardization actions (see A6). Efforts should particularly seek to collect, publish, and systematically analyze use cases (see A8) and align legal (including tax treatments) and governance structures necessary for safe and fair level European-wide data sharing (see A5).
- **A2:** Further invest in a Euro-wide skills strategy to identify major skill and expertise gaps in the European workforce, devise Euro-wide re-skilling and up-skilling roadmaps, advise education and training bodies with remedial actions, and offer further scholarships and fellowships in crucial areas of study.

Given the nature of the above actions, they are intended for all stakeholders, but particularly the EC (and the CSA instrument), consortia behind existing EC CSAs, government at member state, regional and municipal level, industrial and research associations, standardization bodies, consumer organizations, as well as (in the case of A2) educators and industry as employers.

Assuming an acceptable degree of convergence is achieved, the deployment of successful trusted data sharing framework and governance structures can be achieved by:

- **A3:** Funding the development of rules and guidelines for iterative, multi-phase design, creation, scale-out, or merging of existing Data Spaces.
- **A4:** Launching research and innovation actions (including EC R/IAs) to evolve high-impact technology (privacy-, interoperability-, security-, quality-, and ethics-by-design) emphasizing an interoperability initiative across existing spaces and the deployment of trust solutions for data sharing assurance.
- **A5:** Developing EU-wide Data Governance practices to clearly define rules of conduct for the fair use of exchanged data after testing different approaches to assessing the impact of data sovereignty, campaign for the revision of relevant legislation, and explore means for conformity assessment (e.g., voluntary, or licensed certification, data sharing assurance agreements) that guarantee trust.

The above actions can be primarily implemented by the EC (RIA instruments), governmental organizations, regulators, industry associations, direct data suppliers and end users, technical and sector-oriented R&D associations, as well as (see A5) standardization bodies, assurance stakeholders, and consumer/end user organizations.

The widespread adoption of deployed accountable, reliable, and trustworthy Data Spaces will only succeed if organizations at the forefront of relevant research, development, and innovation activities bolster standardization activities by:

- **A6:** Proactive coordinated engagement with international standardization bodies to establish key performance indicators (KPIs) that drive standardization of common building blocks (see action no. 1), successful technology (see action no. 4), methods (e.g., data rights management as ownership solution), and best

practices (e.g., guidelines for international data valuation, privacy preservation, and FAIR principles for non-research data) behind trust-enabling solutions.

- **A7:** Assurance of sufficient conditions (assurance, market demand, government procurement) to enable widespread adherence to established standards, rules, and guidelines and promotion within networks.

The above actions are primarily intended for technology and infrastructure providers (e.g., BDVA i-Spaces), standardization bodies in the information and communications technology (ICT) domain, sector-oriented (vertical) associations, data marketplace operators, direct data suppliers, and consumers.

Deployment activities should be supported and guided to exploit various experimentation instruments to make maximum use of safe environments for testing evolutionary steps of a European data sharing space by:

- **A8:** Investing in piloting to test-drive identified data sharing use cases in safe and dynamic regional and European-wide cross-sectoral scenarios (pairing member states with different levels of progress in data sharing infrastructures).
- **A9:** Engaging with major initiatives offering experimentation activities that rely on data sharing (EDIH future network, BDVA i-Spaces network) to jointly explore market capabilities for sustainable new business, cross-sectoral opportunities, and organizational models (e.g., for data valuation and organizational transformation) and create innovative solutions.
- **A10:** To encourage innovation and motivate data sharing, setting up European regulatory sandboxes for trusted data sharing and safe exploration of risks.

The joint realization of the actions requires the attention of funding bodies within member states, industrial targets across Europe, start-ups, entrepreneurs and technological providers, direct data suppliers, consumers, investors, venture capitalists, and incubators in coordination with governmental organizations and regulators. In parallel to all actions, greater awareness of the opportunities offered by an open, fair, and ethical data economy needs to be achieved. To this end, we call for the following supplementary actions:

- **A11:** Launch a campaign that influences organizations to revisit their data strategy and lifecycles to ensure sharing-ready data by design.
- **A12:** Launch an EU-wide citizen-oriented campaign for an open, democratic, and fair data economy and the right to the free flow of safe and trusted data.

Drivers of activities for greater awareness include the EC, government at member state, regional and municipal level, sector-based industrial associations, entrepreneurs and technology providers, and consumer/end user organizations.

## 7 Conclusion

This chapter described initial challenges, opportunities, and calls to action to set the scene for common European Data Spaces. Setting a conceptual basis using a community vision is a necessary first step. However, progress toward actionable Data Spaces will require an iterative learning curve. Therefore, action and learning from action is essential, creating a feedback loop for development. Our proposed roadmap delineates and estimates progress, while framing the roadmap further can encompass more elaborate timeframes of technological architectures laid out by the key stakeholders.

There are many critical points which the proposed action items mitigate as the objective is very ambitious. Ensuring the EU's pole position in data sharing and Data Spaces requires investment and strategic cooperation between European organizations in addition to technical competence development. The digital transformation will be a reality. There is no turning back. The competitiveness of the EU depends on the successful implementation of common European Data Spaces.

**Acknowledgments** We would like to thank the contributors from the BDVA Task Force on Data Sharing Spaces, including Simon Scerri (Fraunhofer), Tuomo Tuikka (VTT), Irene López de Vallejo (DisCO.coop), Martina Barbero (BDVA), Arne Berre (SINTEF), Davide dalle Carbonare (Engineering), Oscar Corcho (UPM), Edward Curry (Insight Centre for Data Analytics), Valerio Frascolla (Intel), Ana García Robles (BDVA), Robert Ginhör (Know-Center), Sergio Gusmeroli (POLIMI), Allan Hanbury (TU Wien), Jim Kenneally (INTEL), Antonio Kung (TRIALOG), Till Christopher Lech (SINTEF), Antonis Litke (NTUA), Brian Quinn (INTEL), Dumitru Roman (SINTEF), Harald Schöning (Software AG), Tjerk Timan (TNO), Theodora Varvarigou (NTUA), Ray Walshe (DCU), and Walter Weigel (HUAWEI). We would like to acknowledge the comments from BDVA members and external communities received.

## References

1. Scerri, S., Tuikka, T., & Lopez de Vallejoan, I. (Eds.) (2020). *Towards a European data sharing space*, Big Data Value Association, Brussels.
2. Curry, E. (2020). *Real-time linked dataspace*. Springer International Publishing. <https://doi.org/10.1007/978-3-030-29665-0>
3. Curry, E., Derguech, W., Hasan, S., Kouroupetroglou, C., & ul Hassan, U. (2019). A real-time linked dataspace for the Internet of Things: Enabling 'pay-as-you-go' data management in smart environments. *Future Generation Computer Systems*, 90, 405–422. <https://doi.org/10.1016/j.future.2018.07.019>
4. Zillner, S., Curry, E., Metzger, A., Auer, S., & Seidl, R. (Eds.) (2017) *European big data value strategic research & innovation agenda*. Big Data Value Association. Available: [http://www.edwardcurry.org/publications/BDVA\\_SRIA\\_v4\\_Ed1.1.pdf](http://www.edwardcurry.org/publications/BDVA_SRIA_v4_Ed1.1.pdf)
5. *Communication: A European strategy for data*. (2020). Available: [https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020\\_en.pdf](https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf)

6. Pearson, S., & Casassa-Mont, M. (2011). *Sticky policies: An approach for managing privacy across multiple parties*. *Computer*, 44(9). doi:<https://doi.org/10.1109/MC.2011.225>.
7. Zillner, S., Bisset, D., Milano, M., Curry, E., García Robles, A., Hahn, T., Irgens, M., Lafrenz, R., Liepert, B., O'Sullivan, B. and Smeulders, A., (eds) (2020) *Strategic Research, Innovation and Deployment Agenda - AI, Data and Robotics Partnership. Third Release*. September 2020, Brussels. BDVA, euRobotics, ELLIS, EurAI and CLAIRE.

**Open Access** This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

