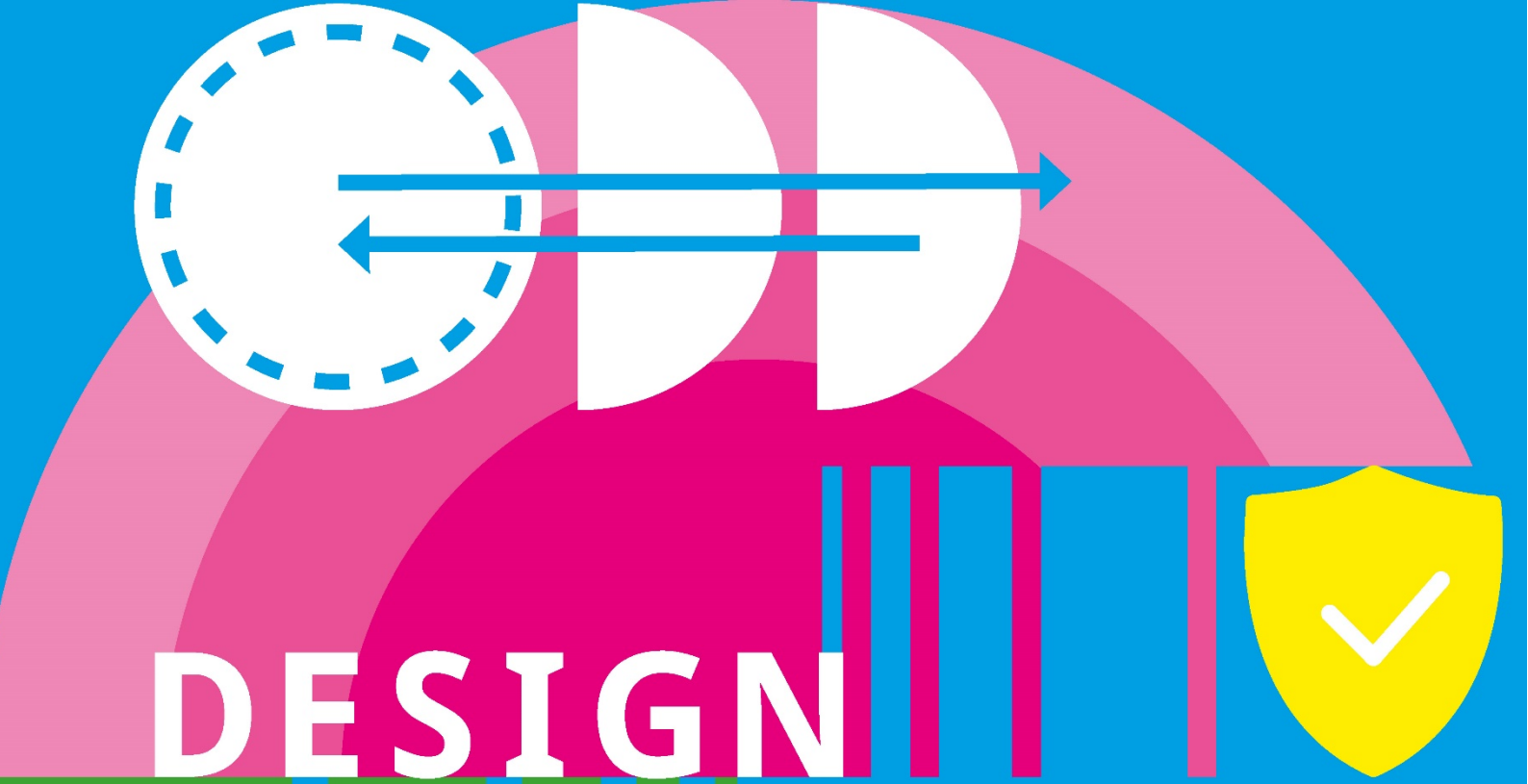
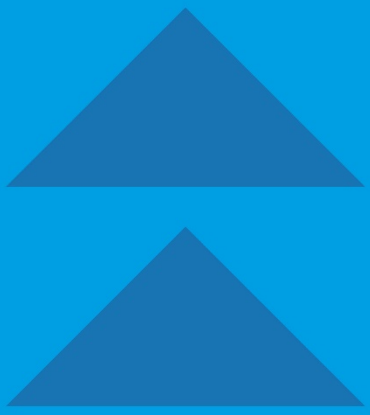
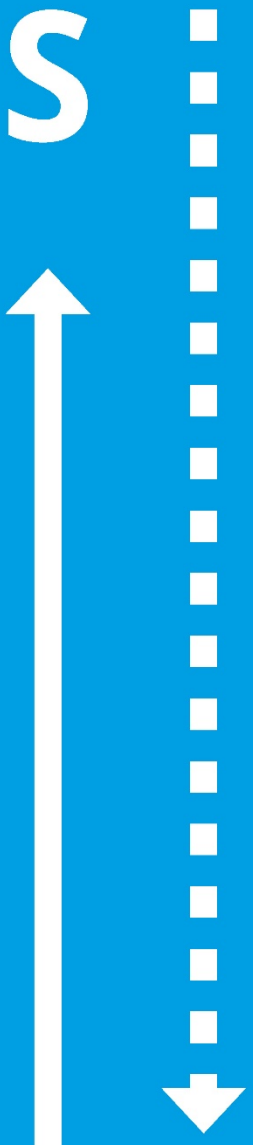


DESIGN PRINCIPLES FOR DATA SPACES

POSITION PAPER | VERSION 1.0 | APRIL 2021





OPENDEI

Data space and industrial domain experts team up to define for the first time cross-sectoral and across initiatives the fundamental design principles to build data spaces.

The position paper underlines the importance of data spaces and though the sovereign sharing of data in creating the future data economy. It has been developed under the coordination and leadership of Task Force 1 lead by International Data Spaces Association of the Horizon 2020 project "OPEN DEI Aligning Reference Architectures, Open Platforms and Large-Scale Pilots in Digitising European Industry" with the collaboration of more than 40 data spaces and industrial domain experts representing more than 25 organisations from 13 Horizon 2020 projects and related initiatives. This is the first approach to define the design principles for data spaces, agreements on the building blocks for a soft infrastructure and governance for data spaces.

Contributing Organisations

Contributing Projects

Publisher

International Data Spaces Association
Anna-Louisa-Karsch-Str. 2
10178 Berlin
Germany

Digital Object Identifier

https://doi.org/10.5281/zenodo.5244997

Editor

Lars Nagel
International Data Spaces Association
Douwe Lycklama
Innopay

Authors & Contributors

Ulrich Ahle, FIWARE
Harrie Bastiaansen, TNO
Kjell Bengtsson, JOTNE
Mallku Caballero, AgriCircle
Silvia Castellvi, IDSA
Alberto Dognini, RWTH
Frans van Ette, TNO
Marianna Faraldi, TCA
Joshua Gelhaar, Fraunhofer
Alessio Graziani, ENGINEERING
Andrej Grguric, ERICSSON
Sergio Gusmeroli, POLIMI
Kristian Helmholt, TNO
Juan Jose Hierro, FIWARE
Denise Hoppenbrouwer, Innopay
Thorsten Huelsmann, IDSA
Srdjan Krco, Dunavnet
Antonio Kung, TRIALOG
Nuria De Lama, ATOS
Oscar Lazaro, INNOVALIA
Angelo Marguglio, ENGINEERING
Maria Marques, UNINOVA
Christoph Mertens, IDSA
Giorgio Micheletti, IDC
Luc Nicolas. EHTEL
Boris Otto, Fraunhofer
Eugenio Perea, TECNALIA
Carmen Polcaro, INNOVALIA
Matthijs Punter, TNO
Jorge Rodriguez, ATOS
John Soldatos, INTRASOFT INT.
Sebastian Steinbuss, IDSA
Harald Sundmaeker, ATB
Anne-Sophie Taillandier, IMT
Mariane ter Veen, Innopay
Francesco Torelli, ENGINEERING
Tuomo Tuikka, VTT
Marko Turpeinen, 1001 LAKES
Luis Usatorre, TECNALIA
Javier Valiño, ATOS

This paper has received funding from the European Horizon 2020 Programme for research, technological development and demonstration under grant agreement n° 857065

Table of content

List of figures

0 Coordinated approach for establishing data spaces to live up to European ambitions for a thriving data economy

0.1 Introduction

In February 2020, the European Commission announced the European Strategy for Data,[1] aiming at creating a single market for data to be shared and exchanged across sectors efficiently and securely within the EU. Behind this endeavour stands the Commission's goal to get ahead with the European data economy in a way that fits European values of self-determination, privacy, transparency, security and fair competition. For this to achieve, the rules of accessing and using data must be fair, clear and practicable. This is especially important as the European data economy continues to grow rapidly – from 301 billion euros (2,4 % of GDP) in 2018 to an estimated 829 billion euros (5,8 % of GDP) by 2025.[2]

The centrepiece of the European Data Strategy is the concept of "data spaces", for which the Commission defined nine initial domains, all driven by sector-specific requirements. From a technical perspective, a data space can be seen as a data integration concept which does not require common database schemas and physical data integration, but is rather based on distributed data stores and integration on an "as needed" basis on a semantic level. Abstracted from this technical definition, a data space can be defined as a federated data ecosystem within a certain application domain and based on shared policies and rules. The users of such data spaces are enabled to access data in a secure, transparent, trusted, easy and unified fashion. These access and usage right can only be granted by those persons or organisations who are entitled to dispose of the data.

As individuals and organisations usually act in multiple ecosystems at the same time, they are not limited to sharing data within a single data silo or data domain only[3]. Thus, data spaces can be overlapping or even nested. To prevent fragmentation of the data economy into multiple, mutually isolated domains, and to create appropriate conditions for setting up an open data ecosystem characterized by mutual trust between participants, a European 'soft infrastructure' is needed, specifying legal, operational and functional agreements as well as technical standards for being widely adopted by users. The total number of all data applications organically emerging over time will then constitute the de-facto infrastructure. This (intangible)

[1] European commission (February 2020) Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions. available at https://ec.europa.eu/info/sites/info/files/communication-european-strategy-data-19feb2020_en.pdf

[2] IDC (2020) Final Study Report: The European Data Market Monitoring Tool Key Facts & Figures, First Policy Conclusions, Data Landscape and Quantified Stories. available at https://datalandscape.eu/sites/default/files/report/D2.9_EDM_Final_study_report_16.06.2020_IDC_pdf.pdf

[3] Big Data Value Association (BDV) (November 2020) *Towards a European-governed data sharing space.* Available at https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf

infrastructure will facilitate both data sovereignty and platform interoperability across multiple domains, with users participating in multiple data spaces and switching from one data space to another in a seamless fashion[4].

Just like other soft infrastructures (e.g. the internet), data spaces are sector-agnostic, with many requirements and functions being similar or even identical across different sectors and data spaces. Therefore, creating a soft infrastructure for data spaces primarily is not so much a technological challenge, as there are plenty of technical solutions and standards available.

Realising interoperable data spaces is more of a coordination challenge: agree on standards and design principles that are accepted by all participants. While making data interoperability work in pilot applications, proof of concepts, and living labs is relatively easy, the real challenge lies in viewing interoperability as the new norm for facilitating mass adoption and scalability. The authors of this position paper expect that a critical mass for irreversible adoption can be achieved within five years, while full adoption will take about a decade.

The soft infrastructure underlying European data spaces should be developed and established in a coordinated way, combining technological, functional, operational and legal processes. Thus, it should be carried by the community of public and private stakeholders – and not by an individual keystone company, as we know it today from leading platform providers. An inspirational analogy can be made here to GSM (the standard for mobile telecommunication), which is also a soft infrastructure bringing together distributed actors and data in a unified user experience across the globe, simply through collaboration and coordination based on well-balanced governance mechanisms. A similar approach to setting up European data spaces will leverage the full potential of the European data economy in line with European ambitions and values.

0.2 Data space design principles

Here is the vision: Five years from now, the EU Strategy for Data will have been fully implemented, multiple data spaces will have been widely adopted across Europe, and European individuals and organisations will have regained the possibility of control over their data and, with that, their rightful and balanced place in the digital world. More initiatives will have been started and more value is captured in Europe. Ten years from now, this is mainstream and the larger audience wouldn't accept it any other way.

[4] Examples of soft infrastructures include the global GSM network, payment networks and the internet. All based on agreements to be implement on hard infrastructures and adopted by users, in a decentralized and competitive fashion.

While the possibilities seem endless, European data spaces basically will bring about three new elements:

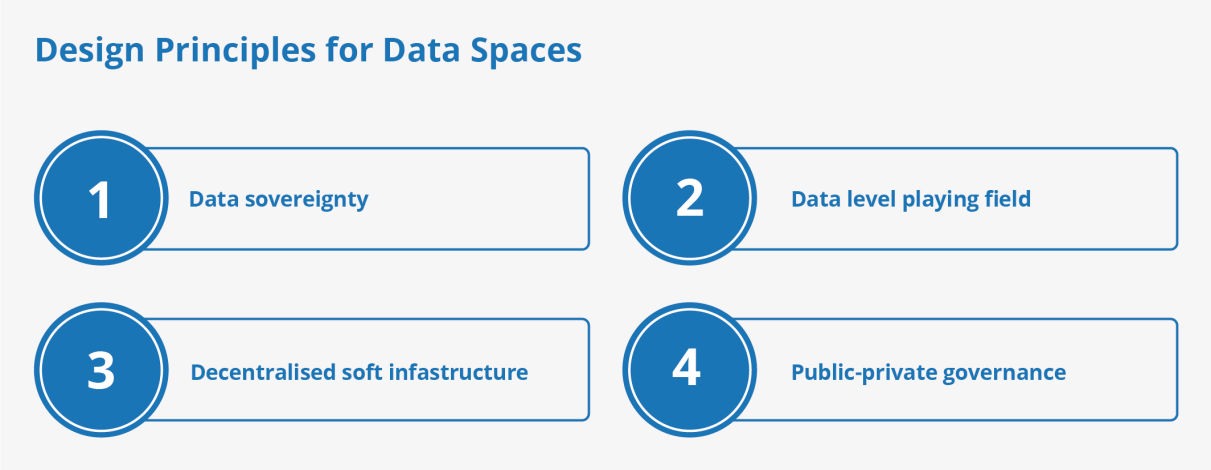
» entirely new services for users, based on enhanced transparency and data sovereignty;

» a level playing field for data sharing and exchange, leading to less dominance of, and dependency on, large, quasi-monopolistic players;

» a new user behavior and digital culture, as users learn to play by the rules and use data (both their own and other users' data) in an ethical way.

In the following paragraphs, these three elements will be elucidated further.

By sketching the vision and the approach to exploiting the potential of data spaces as specified above, the authors of this position paper are proposing four design principles for European data spaces to be built on:



This will be the way for data spaces to become a solid and sustainable foundation for the next growth cycle within the digital economy of the EU.

0.2.1 Entirely new services for users based on enhanced transparency and data sovereignty

While GDPR grants individuals the right to decide what data collectors are allowed to do with their personal data and what not, European data spaces will provide the tools to exert these

rights and stay in control over that data.[5] However, European data spaces will not do so for individuals only, but also for companies/organisations and their data. Driven by sector-specific needs, data spaces will promote the development of tools to share, exchange and access all types of data, including data that is stored in smart objects and things. These tools will empower those entitled to the data to always demand transparency as to where their data is stored and what access rights apply to it. They can use these tools to give or revoke their consent and to change access rights and specify new conditions of how their data can be accessed and used. Furthermore, they can choose to outsource data rights management to third parties (e. g. data intermediaries), just like users (individuals and organisations) today outsource the management of their financial balances to financial institutions (i.e. think investment profiles).

This future scenario highlights Design Principle 1 for data spaces to be built on [6]:

Design Principle 1 for data spaces:

Data sovereignty: is the capability of a natural person or a corporate entity for exclusive self-determination with regard to its economic data goods. This is the innovative and transformative concept underlying data sapces.

With regard to such data sovereignty tools, a nice analogy can be drawn with how users today control information about their bank accounts (balance) through their electronic payment cards. Every time a shopper uses their card for payment, there is a data sharing transaction taking place. The card functions as an instrument to 'prove' to the merchant that the shopper's bank account holds enough balance to buy the goods. With each transaction, the card together with the PIN code acts as a data sovereignty tool for the user, through which a tiny piece of information ('enough balance: yes or no?') is shared by the shopper's bank with the bank of the merchant via electronic data systems. The ecosystems of the actors involved in the process are kept together by a soft infrastructure consisting of rules and agreements of legal, functional, technical and operational nature. Similar to how we have learned to pay with these cards, new functions of sharing and exchanging data will arise from using data spaces and be adopted by users on a large scale.

[5] Press conference Margrethe Vestager on 25 November 2020: "You don't have to share all data. But if you do and data is sensitive you should be able to do in a manner where data can be trusted and protected. We want to give business and citizens the tools to stay in control of data. And to build trust that data is handled in line with European values and fundamental rights", *available at* https://ec.europa.eu/commission/presscorner/detail/en/IP_20_2102

[6] Prof. Dr-Ing. Boris Otto, Fraunhofer ISST (2017) INDUSTRIAL DATA PLATFORMS. Available at https://ec.europa.eu/futurium/en/system/files/ged/b2-otto-industrial_data_space.pdf

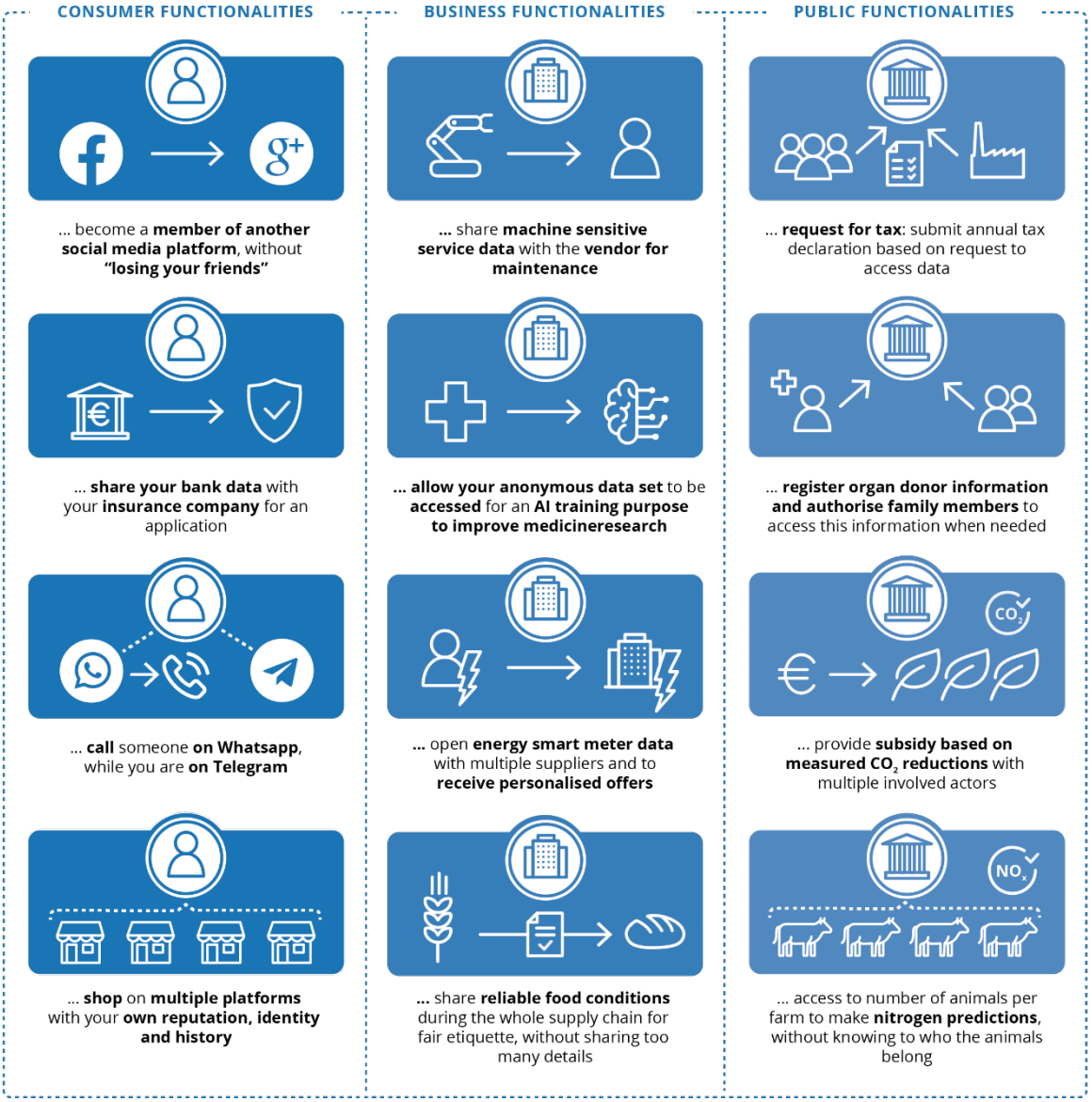


Figure 1 Some expressions of new possibilities in the data economy.

By regaining control over their data, and as data becomes portable between providers on a user-controlled consent basis. Users can switch between providers without losing their data and vendor lock-in will become a phenomenon of the past.

0.2.2 Level playing field for data sharing and exchange, leading to less dominance of large, quasi-monopolistic players and more opportunities for businesses and individuals.

Data spaces are a central element within the European Commission's goal to foster competition and innovation in the data economy by creating a level playing field for data sharing and exchange.[7] With the actual parties that generate the data regaining control, large, quasi-monopolistic players will no longer have the chance to position themselves as exclusive 'data owners'.

Users will be empowered to 'move' to another provider, while being able to take along their data and keep all contact information, chat histories, reputation gained, and so on. Switching social platforms while 'taking your friends' will be just as normal as switching telco providers whilst taking your number. Market entry barriers for new players will be reduced and fair competition will be stimulated. This especially holds for incumbents and SMB, but will also change the position of larger companies in their powerplay with platforms and their ability to capitalise on their (and other's) industrial data by creating new services and innovations.

Design Principle 2 for data spaces:

Data level playing field: implies that new entrants face no insurmountable barriers to entry because of monopolistic situations. When a data level playing exists, players compete on quality of service, and not on the amount of data they control. A data level playing field is a pivotal condition to create a fair data sharing economy.

In the early days of the internet, nobody would have expected that the true innovation of electronic media was how it changed the overall behaviour and daily routines of people. The internet enabled users to be connected 24/7, shop all over the world, and manage multiple aspects of daily life. While new electronic infrastructures initially seemed to address a limited number of use cases, they actually created a new space of endless possibilities through mass adoption within just two decades. Common standards (such as HTML for the internet, GSM for mobile communication) enabled global collaboration. People have become accustomed to using their smartphones to phone, text and chat from everywhere.

In the same way, data spaces will change the behaviour, culture, and etiquette when it comes to sharing and exchanging data. Users will become more mindful about treating their data as an asset. A Netflix documentary called *The Social Dilemma* (2020) highlighted that using social media is not for free ('If you do not pay for the product, you are the product'). Another example

[7] European Commission (2019) Competition policy for the digital era. Available at https://ec.europa.eu/competition/publications/reports/kd0419345enn.pdf

is Amazon paying their customers for providing data on transactions with other retailers. While users already know data has value, they hardly act on it. The absence of valuation standards and tools to control data makes it a challenge to monetise it.[8] Only if users have the means to control their data they are able to treat it as a true asset. The same holds for organisations, data spaces will enable them to capitalise on industrial data developing new services and business models we can't even dream of today.

0.2.3 Need for data space interoperability – the soft infrastructure

With its Strategy for Data, the Commission promotes the development of European data spaces for strategic economic sectors and public-interest domains, starting with the following nine: industrial (manufacturing), green deal, mobility, health, financial, energy, agriculture, public administration, and skills.

While data spaces stimulate higher availability of data pools, technical tools, and infrastructure addressing domain-specific challenges and legislations, the EU Strategy for Data acknowledges that these data spaces should be interconnected and that this challenge requires specific attention. But Europe doesn't need to start from scratch – data sharing and exchange within specific domains and sectors is already happening in existing initiatives. However, each of these initiatives follows its own approach, and therefore they are not interoperable. So, part of the EU strategy should be to include and build upon existing data-sharing initiatives in the quest for interoperability and the specification of future 'soft infrastructure' agreements.

Interoperability between domain-specific data spaces is crucial for two reasons. First, an individual or organisation is never just part of one single space but operates in different spaces simultaneously. If data spaces are organised in silos, users have to adopt different solutions. This results in fragmentation, high integration costs, and monopolistic behaviour of market participants. Second, use cases are not limited to a single data silo. Fragmentation of the data economy must be prevented to reap the maximum value for organisations and individuals in the EU.

Infrastructures in general build the foundation on which all providers can offer their services. In the GSM infrastructure, for example, all telecommunication providers use cell towers to transfer standardised signals, they all use the same identifiers (in the form of telephone numbers) for people to call one another and similar contracts to settle financial balances among each other. For data sharing and exchange, the physical (or hard) infrastructure is available (cables, data centres etc.). However, *how* participants interact with each other is not standardised. Therefore, a soft infrastructure for data spaces is needed that addresses the agreements on how to participate.

[8] *Amazon launches a program to pay consumers for their data on non-Amazon purchases* (October 2020) available at https://techcrunch.com/2020/10/20/amazon-launches-a-program-to-pay-consumers-for-their-data-on-non-amazon-purchases/

Figure 2 shows a selection of essential infrastructures that are a combination of hard and soft infrastructure. Historically, the wealth of people and their nations is built on infrastructures, resulting over time in a balance between public and private interests. For data sharing and exchange, many domains today rely on privately governed solutions characterised by quasi-monopolistic tendencies. In part this goes back to the early days of the internet, which was standardised up till the point of interactions only. Soon in its life, transactions became internet's paramount feature. The standardisation for transactions, which requires elements such as identification, authentication, authorisation (consent) was left to the private sector and quickly taken by the larger quasi monopolistic companies. It is now time to bring back the balance between public and private interests also in this part of the standardisation. Interoperable data spaces are an essential precondition for rebalancing public and private interests.

Selection of essential infastructures

Application	Hard infrastructure	Soft Infrastructure
	Physical	Agreements how to use (e.g.)
Data	Cables, data centres	Data spaces soft infrastructure
Identity	Cables, servers, data centres	Level of Assurance, processes, legal agreements
Payments	Payment cards and POS terminals	Legal agreements, charge back procedures
Email	Fibre optic network, servers, cables	Email clients, pop and imap protocol
Internet	Fibre optic network	Internet Protocol (IP)
Mobile communication	Cell towers	GSM, Mobile number as address, legal agreements, financial
Electricity	Electricity cables	Voltage of 230 V, universal socket design
Rail	Physical rails	Right of use of physical rails, operational agreements
Road	Physical roads	Drive on the right side of the road, road sign meanings

Forms of data

Figure 2 Selection of essential infrastructures.

Infrastructures lay the foundation for interoperable services. Each infrastructure includes hard infrastructure and soft infrastructure elements. The soft infrastructure specifies how to use the hard infrastructure based on functional, technical, operational and legal agreements.

The soft infrastructure provides a level playing field for data sharing and exchange. It is made up of technology-neutral agreements and standards specifying how organisations and individuals can participate in the European data economy and how they need to act and behave in compliance with commonly agreed rules and directives. As all participants implement the same minimal set of functional, legal, technical and operational agreements and standards, they can interact in the same manner, no matter what data space they are operating in. Integral design of these agreements and standards from the start will provide cohesion, as these elements are complementary rather than following up on each other.

The soft infrastructure will specify common functional, legal, operational and technical aspects, such as security, identity, authentication, protocols, metadata etc. In order to establish sector-adequate data spaces, the soft infrastructure needs to be complemented with sector-specific aspects. This will lead to a stacked, role-based architecture.

Key building blocks and roles will be part of each data space (see Chapter 2 of this publication). Some elements of the building blocks will be similar for multiple spaces and will therefore be part of the general soft infrastructure. For instance, standard data models and standard APIs will be an important part of the soft infrastructure for data spaces, just like HTML has been for the internet or GSM for mobile communication. Other elements of the building blocks will need to be customised to work for sector-specific data spaces (see Chapter 3). Figure 3 provides an indicative overview of the stacked architecture of the soft infrastructure for data spaces.

It should be noted that the content of the stacks as shown in Figure 3 is indicative and subject to change. The overview indicates the general building blocks (categorised in interoperability, trust, data value, and governance) in line with data space design principles. The stacks of the architecture offer space for customisation within these building blocks. During the development of the soft infrastructure, agreements and standards will be defined in line with the building blocks.

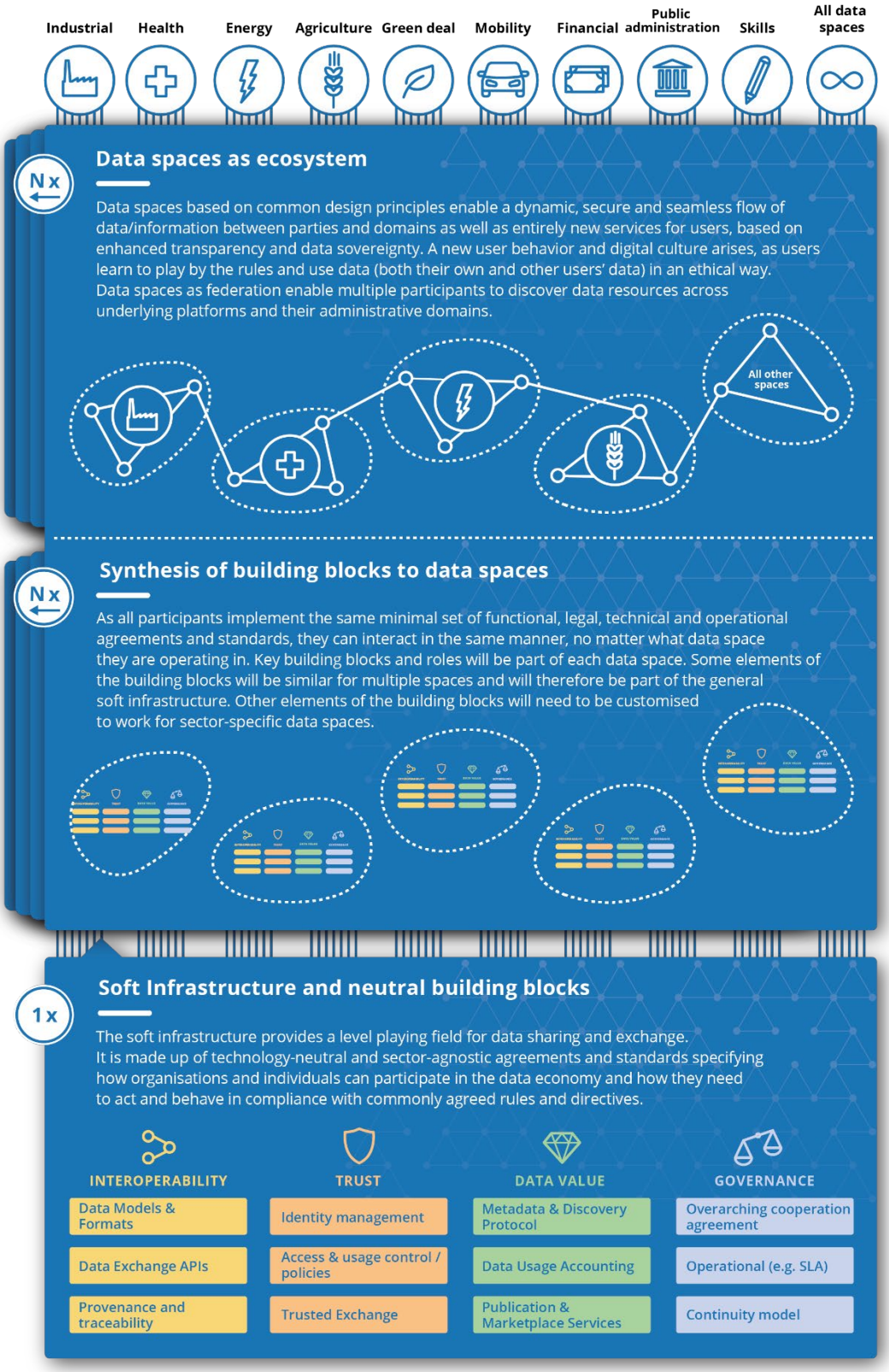


Figure 3 Overview of the stacked architecture of the soft infrastructure for data spaces.

Design Principle 3 for data spaces:

Decentralised soft infrastructure: The data sharing infrastructure is not a monolithic centralised IT infrastructure. Instead, it is the de facto collection of interoperable implementations of data spaces which comply to a unified set of agreements in all disciplines: functional, technical, operational, legal and economic. A 'soft infrastructure', as it is merely invisible and made up of agreements. Out of the principle of data sovereignty follows functional and non-functional requirements of interoperability, portability, findability, security, privacy and trustworthiness.

A soft infrastructure for data space interoperability and data sovereignty of users is the way to prevent that the current mode of operation, which is characterised by a limited number of providers and concentration of 'data power' in a few hands, will prevail. The soft infrastructure will lead to decentralisation and a level playing field for data sharing and exchange.

0.2.4 Public private governance: Europe taking the lead in establishing the soft infrastructure in a coordinated and collaborative manner.

Europe is standing at a historic crossroads, demanding from us to decide about the next evolutionary step in the digital economy. This moment can be compared to the introduction of the GSM standard in the 1980s, which turned out to be the pivotal moment for the natural evolution of telecommunications, towards decentralisation combined with innovation, competition and accelerated adoption.

After 30 years of internet infrastructure driven by private forces, it is time to balance private interests with public interest and create the next 'GSM moment'. Now we know better what we want and what we do not want in terms of our digital economy. The EU Strategy for Data, together with the Data Governance Act, [9] are essential cornerstones of this evolution, which will lead to a new organisation of digital market forces.

Public intervention means indicating the right direction, followed by activation of public and private energy in realising this endeavour.

[9] European parliament and the council (2020) Proposal for a regulation of the European parliament and of the council on European data governance (Data Governance Act). Available at https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=EN

Design Principle 4 for data spaces:

Public-private governance: For the design, creation and maintenance of the data level playing field a sound governance is essential. All stakeholders need to feel represented and engaged. These include users (persons, businesses) or provider of data services as well as their technology partners and professionals.

The recently proposed Data Governance Act [10] confirms the notion of a governance structure constituted by multiple entities. For European data spaces, it is recommended to have a (domain) governance authority for each data space and a central governance authority overseeing all aspects in connection with interoperability of data spaces, i.e. the de-facto 'soft infrastructure'. This central authority will interact with all data space specific authorities. Therefore, N x data spaces plus one central authority will need to be organised (see figure 4). For this governance structure of the soft infrastructure, authorities on three levels (a strategic, a tactical and an operational level) should operate in close cohesion with each other. The DGA specifies the European Data Innovation Board as the authority on the strategic level. Chapter 4 of this position paper proposes the introduction of an authority on a tactical and operational level, called the 'Data Exchange Board' (DEB).

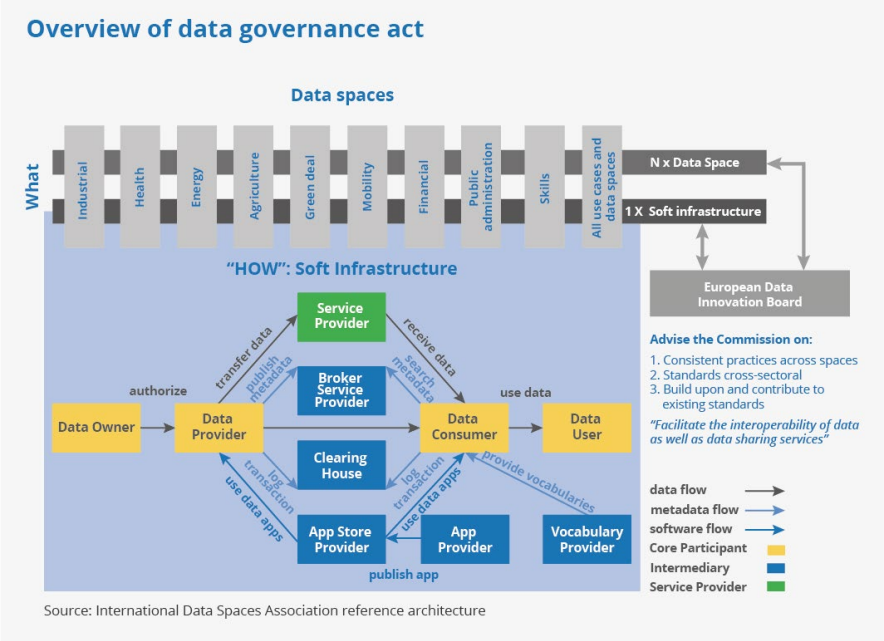


Figure 4 Overview of data governance act.

[10] European Commission (November 2020) proposal for a regulation of the European parliament and of the council on European data governance (Data Governance Act). Available at https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=71222

To live up to the ambitions of the EU with regard to the data economy, the Commission should invest in ten years funding for creation, operation, and mass adoption of a set of agreements needed to establish the soft infrastructure for data spaces. This process will have two key phases: 1) a convergence phase and 2) a deployment phase. During the convergence phase, all stakeholders will need to be aligned in terms of problem identification, specification of a solution, and setting up an action plan. Alignment with other programs would be highly desirable (such as the Connecting Europe Facility program, CEF, [11] as this program is devoted to providing the building blocks for creating digital service infrastructures across Europe). The soft infrastructure underlying European data spaces would be such a kind of digital service infrastructure. Subsequently, three key activities need to converge in order to be able to come up with a first version of the soft infrastructure:

» **Advise 1 –** Set up trusted governance: To facilitate trust, all stakeholders in the project should be included in the program and feel equally important in the co-creation of the soft infrastructure. The initial coalition that will co-create the soft infrastructure agreements should therefore represent different market needs and segments. Furthermore, non-discriminatory and transparent communication as well as decision and escalation lines in governance should be established to ensure that the coalition has trust in the process. Public and private organisations not included in the initial coalition will receive the opportunity to raise their voice in supplemental sounding boards that support and advice the governance bodies established.

» **Advise 2:** *Co-create an interoperable, distributed, public-private soft infrastructure:* It will be mandatory to develop functional, legal, technical and operational agreements and standards that support the most pressing needs of users in various data spaces. In this process, the most eager participants should be in charge of co-creating the initial version of the soft infrastructure. In the past decade, researchers and practitioners across the world have done a lot of thought-work. It is now a matter of agreeing on the optimal, coherent approach across all relevant disciplines and prioritizing the most urgent use cases of businesses and governments.

» **Advise 3:** *Create awareness in the market:* It will be key to create awareness of the rationale, concept, and functional range of the soft infrastructure, before the first version is published. While the coalition will be a fair representation of the market, not all potential participants can be involved in the co-creation process. Creating awareness beyond first movers will support the process of adoption. Ensuring adoption and scalability will be essential for the success of the endeavour. The lack of widespread adherence to current solutions will be a key challenge for European data spaces.[12] To ensure mass adoption, ample attention must

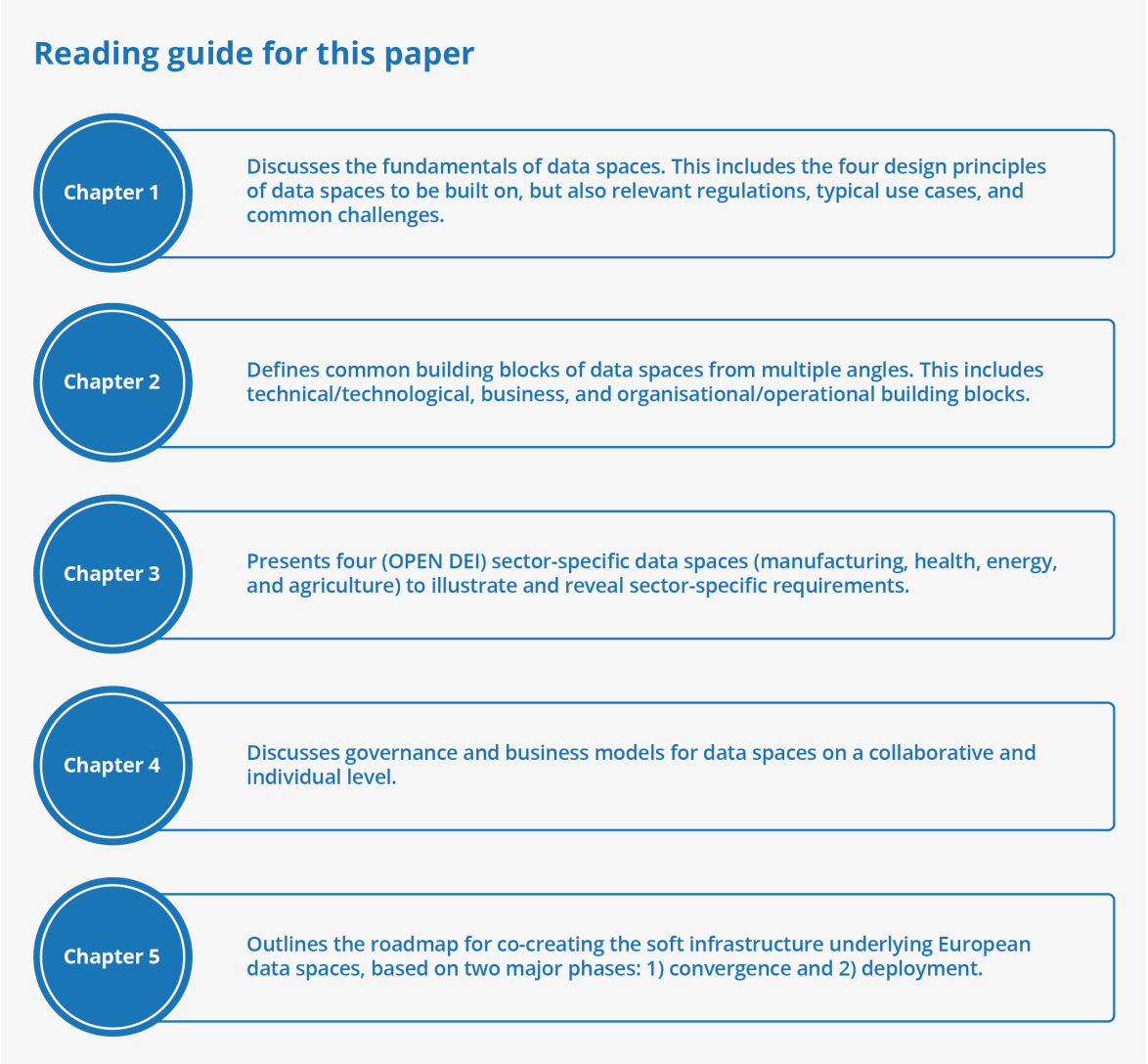
[11] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/CEF+Digital+Home

[12] Big Data Value Association (BDV) (November 2020) *Towards a European-governed data sharing space.* Available at https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf

be given to education, awareness, and inclusion, and appropriate funding should be granted accordingly.

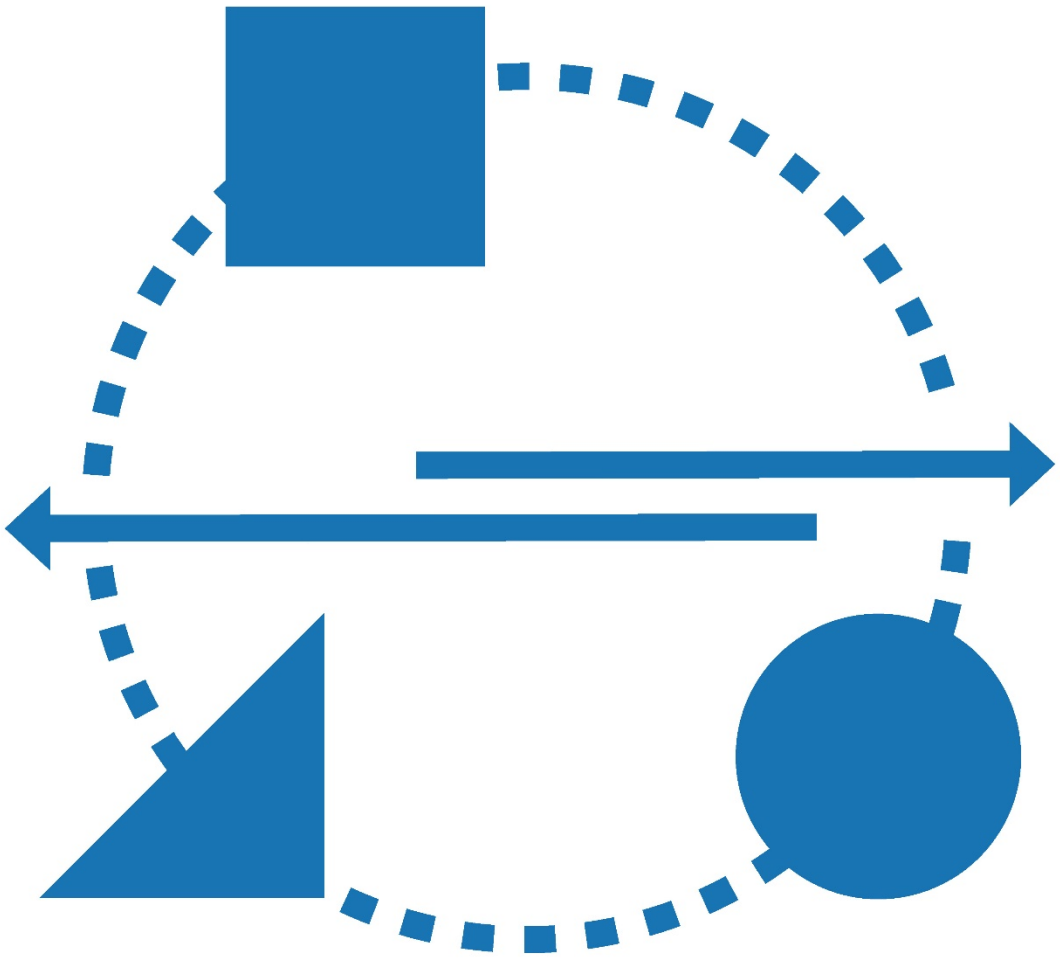
The soft infrastructure will lead to entirely new opportunities in the European data economy. These include opportunities in the AI field, where the access to data is the key to success, usage for manufacturers along industrial supply chain or in use cases in which the individual controls the data flows. But these are merely examples; this soft infrastructure will create additional security and business opportunities for all organisations and individuals across the EU, opportunities we cannot even dream of.

0.3 Reading guide for this paper



Fundamentals of Data Space

1



1 Fundamentals of data spaces

1.1 Introduction

This section of the position paper presents and illustrates the fundamentals of data spaces. The position paper uses an architecture description approach,[13] which promotes a clear separation between the expectations to be met by data spaces and the architecture resulting from these expectations. Figure 5, on the left-hand side, shows how architectures can be specified according to ISO/IEC/IEEE 42010. On the right-hand side, Figure 5 indicates how this approach is adopted for the cause of this position paper, in which this chapter focuses on the expectations, while the following three chapters focus on the resulting architecture.

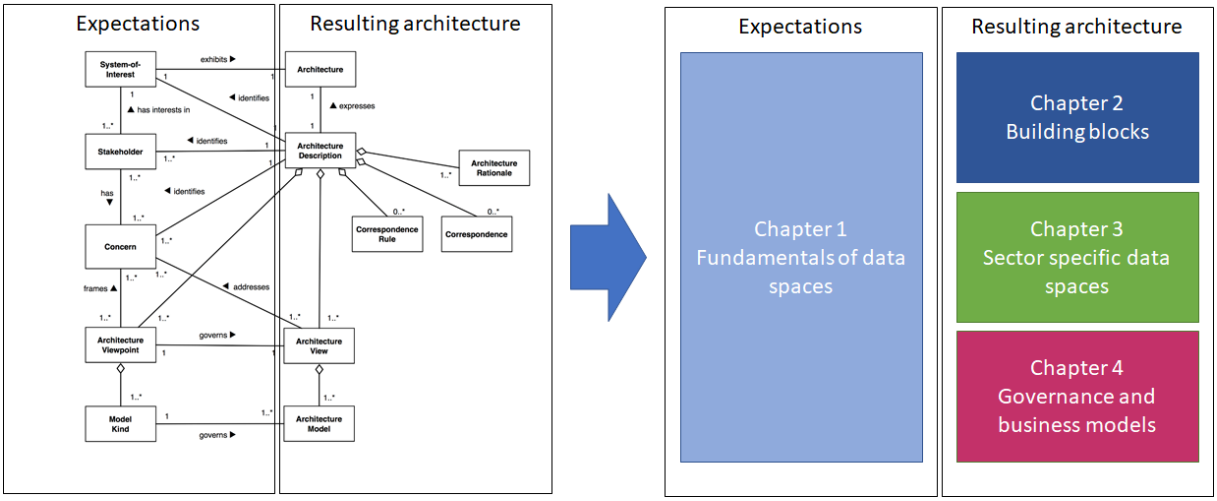


Figure 5 Architecture-based approach of position paper.

Consequently, Chapter 1 addresses:

» the system of interest (i.e. the central building blocks, design principles, etc. of data spaces),

» the stakeholders and their concerns (i.e. the motivation for establishing data spaces),

» resulting expectations consisting of architecture viewpoints and model kinds, including principles on how to deal with data, federation of data spaces, use cases, regulation, and challenges.

[13] Based on ISO/IEC/IEEE 42010 Systems and Software Engineering – Architecture Description, an international standard for describing the architectures of systems and software

1.2 What are data spaces?

A data space is defined as a decentralised infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles.

Data spaces require the following elements:

» building blocks such as *data platforms*, providing support for effective data sharing and exchange as well as for engineering and deployment of data exchange and processing capabilities;

» building blocks such as *data marketplaces*, where data providers can offer and data consumers can request data, as well as data processing applications;

» building blocks ensuring *data sovereignty*, i.e. the ability for each stakeholder to control their data by making decisions as to how digital processes, infrastructures, and flows of data are structured, built and managed, based on an appropriate governance scheme enabling specification of terms and conditions.

Figure 6 shows the example of a high-level architecture of a data space for the mobility sector, as it is currently being developed in Germany.

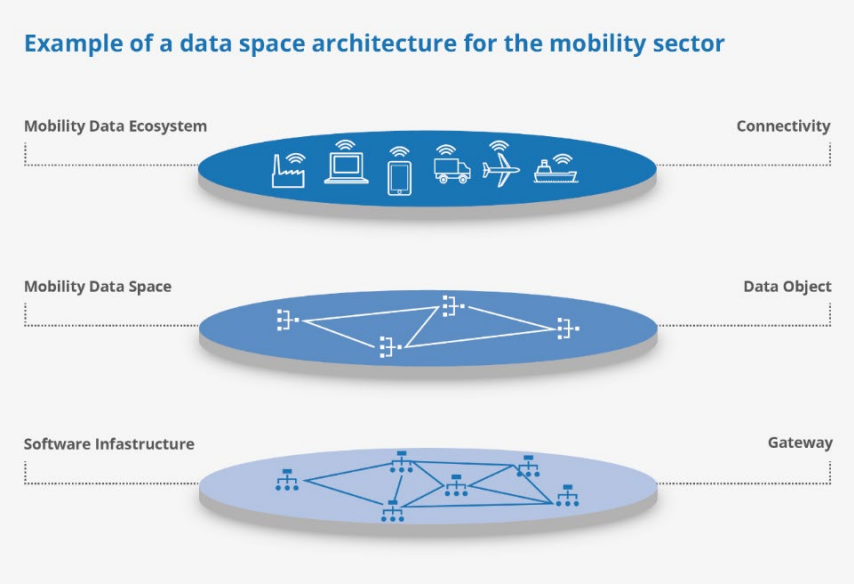


Figure 6 Example of a data space architecture for the mobility sector

The architecture consists of three layers[14]:

1 The top layer is about ensuring connectivity. It represents the mobility data ecosystem, in which the actors of the ecosystem (travellers, public transportation providers, car-sharing companies, etc.) provide or consume end-to-end, data-based mobility services. The goal of these services is twofold: 1) increase comfort for travellers, and 2) optimise traffic and passenger flows.

2 The middle layer represents the data space corresponding to the mobility data ecosystem. It creates digital twins of the different real-world objects present in the mobility domain, which is necessary because sharing data in ecosystems requires accessing and combining different data from various sources to finally create and deliver innovative mobility services. In order to share data, data interoperability must be achieved. What is required therefore is a common understanding of shared data objects among the members of the data ecosystem. For example, battery-charging points for electric mobility must be described by a consistent set of attributes (type, location, charging mode, charging levels, etc.) and attribute values.

3 The bottom layer of the architecture specifies the software infrastructure required to support the creation, management, and sharing of the digital twins' data. To meet multiple requirements in different areas of application adequately, a decentralised architecture is better suited than a centralised architecture. However, a decentralised architecture, just like a centralised one, requires standard software components and gateways allowing data providers and users to participate in the data space ecosystem. In addition, common services are needed to ensure that data can be exchanged and shared between the various, distributed software components.

[14] Boris, O. "Creating data spaces based on GAIA-X and IDS"

1.3 Stakeholders and their concerns

In order to create data spaces that meet the intended purpose, it is important to identify relevant stakeholders and their concerns. The following stakeholders and concerns can be identified with regard to data spaces:

» data consumers: they access data spaces to use data;

» data providers: they collect and manage data and make it available in data spaces;

» data producers: they create data;

» data owners: they have rights to grant or revoke terms and conditions for access and use of data;

» data application providers: they provide applications that transform, process or visualise data;

» data platform providers: they provide capabilities that allow for operation of data platforms;

» data marketplace providers: they provide capabilities that allow for operation of data marketplaces;

» identity providers: they provide capabilities for identifying parties.

Instead of the terms used for stakeholders in the above list, other terms can be used as well. Table 1 shows the correspondence of some of the above terms with the categories of stakeholders defined by ISO/IEC 20547-3.[15]

Table 1. Correspondence of different terms for stakeholders

Terms used in this document	ISO/IEC 20547-3 Terms
Data consumer	Big data consumer
Data provider	Big data provider
Data application provider	Big data application provider
Data platform provider	Big data framework provider

[15] ISO/IEC 20547-3:2020 — Big data reference architecture — Part 3: Reference architecture

Summing up the concerns of all stakeholder groups, the following requirements to be met by data spaces can be specified:

» Data spaces should provide a framework for effective and efficient data exchange, supporting the decoupling of data producers and data consumers. This means they should allow for adoption of common APIs and security schemas, as well as adoption of data models that can be represented in data formats compatible with adopted APIs, for the purpose of sharing and exchanging data.

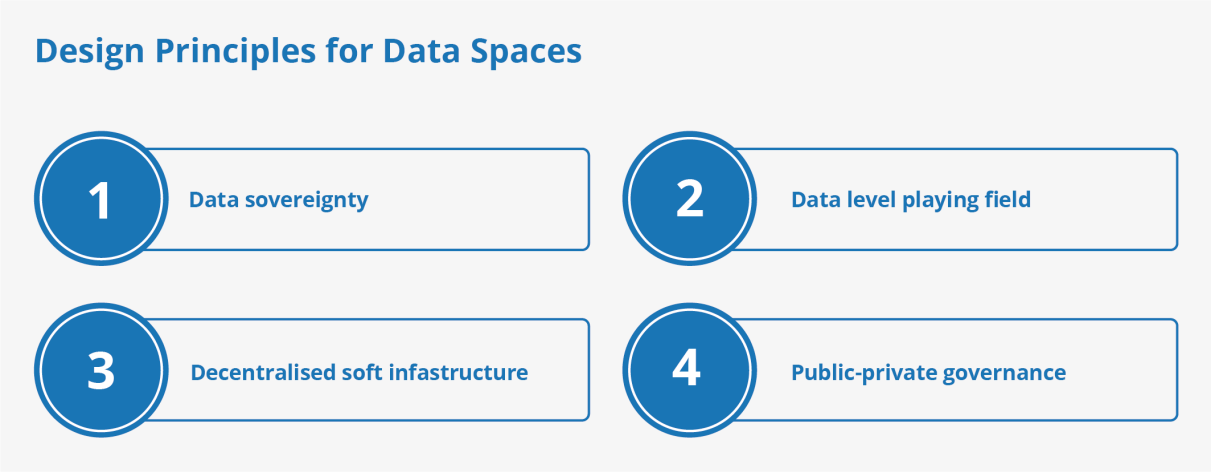
» Data spaces should provide a structure for defining and enforcing agreements on the use of data (including potential monetization of both data provision and data use). This means they should allow for incorporation of capabilities for specifying and publishing data offerings comprising terms and conditions (including pricing) that can be enforced during data exchange transactions.

» Data spaces should provide a structure for trustworthiness, in which data consumers and data providers can share their business interests on the basis of common ethical values. This means they should provide security capabilities to protect data ownership aspects as well as business operations, including capabilities for privacy protection that can be engineered and deployed.

» Data spaces should provide a structure on the basis of which specific policies and regulations can be supported.

1.4 Data spaces design principles

From the stakeholders' concerns outlined in the previous paragraph, the following design principles for data spaces can be derived:



Data sovereignty

Data sovereignty is the capability of a natural person or corporate entity for exclusive self-determination with regard to its economic data goods. This is one of the central innovative and transformative concepts underlying data spaces.

For participants in data spaces, data sovereignty means two things: 1) benefit from enhanced possibilities to view, process, manage, and secure their data, and 2) stay in control over their data when making it accessible to other parties.

Level playing field for data sharing and exchange

Ensuring a level playing field for all data space participants implies that new entrants face no insurmountable barriers (e.g. due to a quasi-monopolistic structure of the data ecosystem) when seeking admission to a data space. On a level playing field, players compete on the quality of their data and services, not on the amount of data they control.

A level playing field for data sharing and exchange can emerge if such an ecosystem is ruled by the idea of cooperation instead of competition. This can be achieved by a sound design and thorough maintenance of the soft infrastructure underlying data spaces.

Decentralised soft infrastructure

The infrastructure for European data spaces will not be a monolithic, centralised IT infrastructure. Instead, it will be made of the totality of interoperable implementations of data spaces complying with a set of agreements in terms of functional, technical, operational and legal aspects. Such a 'soft infrastructure' will be invisible to data space participants. It will entail functional and non-functional requirements regarding interoperability, portability, findability, security, privacy, and trustworthiness.

Viewed from a technical standpoint, a soft infrastructure can be seen as a collection of interoperable, API-based IT platforms, where users control the flow of data through advanced mechanisms of identity and consent management. The design of the soft infrastructure will include mechanisms for economic exploitation of data sharing and exchange transactions (i.e. data monetisation).

The soft infrastructure for data spaces will be technology-neutral, giving maximum freedom to all actors to make their own choices in accordance with their engineering capabilities.

Public-private governance

For the design, creation, and maintenance of the level playing field for data sharing and exchange, sound governance is essential. All stakeholder groups need to feel represented and engaged. This includes businesses, individuals, and governments acting as data users or data providers, as well as their technology partners and IT professionals.

When establishing European data spaces, public values and interests need to be represented as much as private interests. Public values and interests should be ensured through proper legislation and regulation, the foundation of which is already in place (eIDAS, GPDR, PSD). More specific legislation and regulation is currently being prepared (such as DGA,[16] DSA,[17] and DMA).[18]

Such public-private governance is also tasked with promoting broad adoption of European data spaces. A major, long-term effort will be needed for communication of the idea towards all stakeholders. Building up and maintaining a development community is a specific point of attention here.

In the first years of creating and maintaining the soft infrastructure, and establishing data spaces on top of it, the public sector will play an essential role as an early adopter and a provider of financial resources. Once a critical mass is reached, network effects will kick in and adoption of European data spaces will grow by itself.

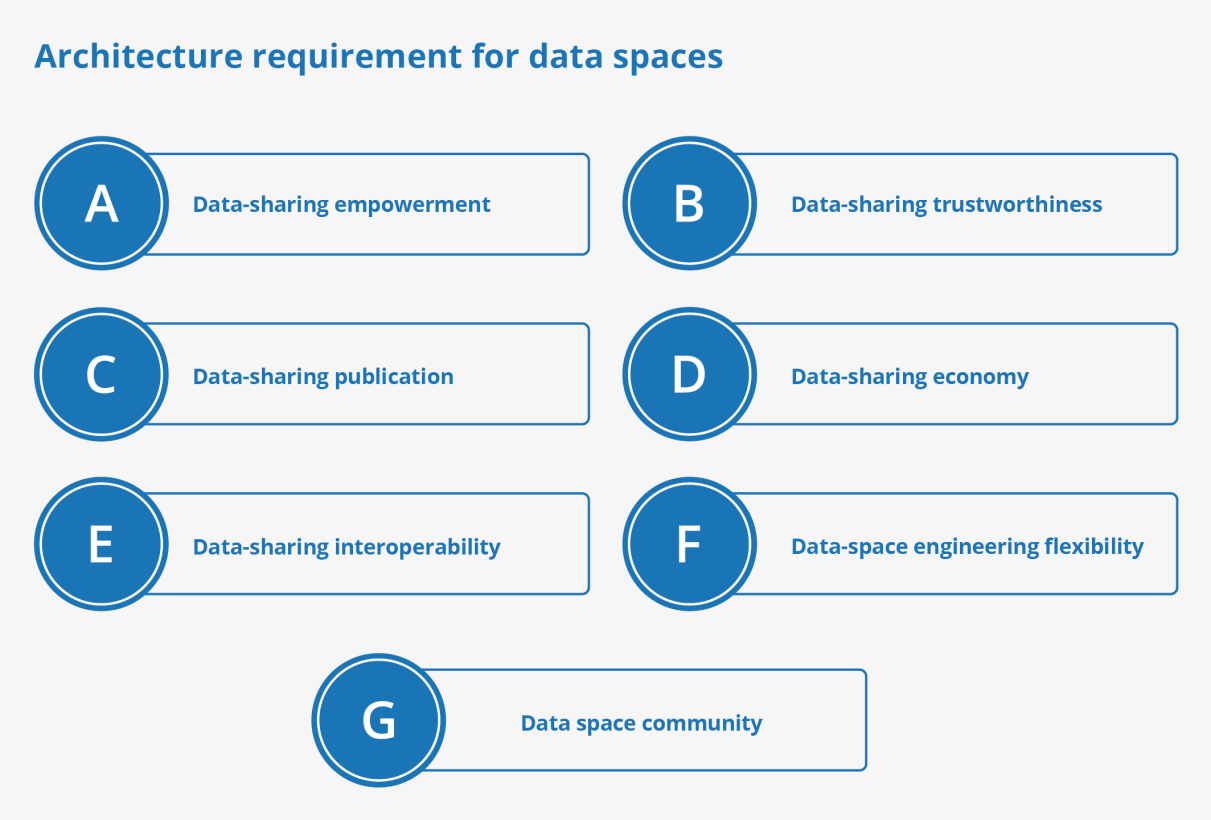
[16] Digital Governance Act; European Commission (November 2020) Proposal for a regulation of the European parliament and of the council on European data governance (Data Governance Act). Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=71222

[17] Digital Service Act; https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package

[18] Digital Markets Act (DMA); https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0842&from=en

1.5 Data spaces architecture requirements

From the stakeholders' concerns and principles outlined in the previous paragraphs, the following seven architecture requirements have been identified for data spaces:



Requirement A: Data-sharing empowerment is about ensuring that decisions can be made by appropriate stakeholders. This means that tools and organisational practices are available for

- governance in data spaces, i.e. the possibility to define and monitor policies in data sharing,
- citizen engagement support, i.e. the possibility for citizens to participate in data sharing and exchange transactions,
- data sovereignty support, i.e. the possibility for stakeholders owning data to govern the use of it, and
- federation, i.e. the possibility to connect several data platforms with each other, with each one retaining control of its own operations.

Requirement B: Data-sharing trustworthiness is about ensuring that data spaces operate according to expected requirements. This means that the development of data-sharing applications must support

» security-by-design, i.e. security of data space assets and support of non-repudiable and unambiguous agreements,

» privacy-by-design, i.e. integration of privacy concerns in the development of data platforms and data-sharing applications, and

» assurance-by-design, i.e. integration of security and privacy assurance requirements in the development of data platforms and data-sharing applications.

Requirement C: Data-sharing publication is about enabling data to be published so it can be easily located by data consumers.

Requirement D: Data-sharing economy is about creating the conditions for data sharing and exchange, requiring

» non-financial incentive mechanisms,

» financial incentive mechanisms, including models to monetise data and methods to determine the value of data, and

» agreement mechanisms.

Requirement E: Data-sharing interoperability is about providing the ability for all applications in data spaces to create, use, transfer and effectively exchange data. This requires the definition of data exchange APIs and data models supporting

» semantic interoperability, ensuring that the meaning of the data model within the context of a subject area is understood by the participating systems,

» behavioural interoperability, ensuring that the actual result obtained from usage of data exchange APIs achieves the expected outcome, and

» policy interoperability, i.e. interoperability while complying with the legal, organisational, and policy frameworks applicable to the participating systems.

Requirement F: Data space engineering flexibility is about providing the ability for engineers to add customised features in data-processing applications and data platforms to enable

» flexibility in terms of interoperability, i.e. extending data spaces with specific interoperability capabilities,

» flexibility in terms of trustworthiness, i.e. extending data spaces with specific security, privacy, and assurance capabilities, and

» flexibility in terms of data processing, i.e. extending data spaces with data-processing capabilities.

Requirement G: Data space community is about fostering maximum reuse of data space solutions. This includes

» open solutions, i.e. ensuring that data space platforms and data-sharing applications are based on open specifications,

» reusability, i.e. ensuring that capabilities of existing data and marketplace platforms as well as data-sharing applications can be easily replicated,

» open source, i.e. allowing free access to data and marketplace components developed by communities, and

» sustainability of solutions, i.e. assurance that solutions will be available and maintained over a long period of time.

Table 2 illustrates how the data space design principles (see Chapter 0 and Chapter 1.4) and the data space architecture requirements are related.

Table 2. Data spaces architecture requirements and design principles

	Data spaces design principles			
Data space architecture requirements	**Data soverei-gnty**	**Level palying field**	**Descentralised soft infraestructure**	**Public-private governance**
Data-sharing empowerment	●	○	○	●
Data-sharing trustworthiness	●	●	●	●
Data-sharing publication	●	○	○	○
Data-sharing economy	●	●	●	○
Data-sharing interoperability	○	●	●	○
Data space engineering flexibility	○	●	●	○
Data space community	○	●	●	○

1.6 Towards trustworthiness in data ecosystems

The principles outlined in the preceding section can be addressed through an ecosystem structure that is based on interoperable and distributed data spaces. Federation is one of the central aspects when it comes to creating trust. It calls for interconnection of multiple data platforms that are operated and controlled individually.

Requirements regarding trustworthiness will have to address the following concerns:

Federated security management consists of having individual data spaces' security management associated with a federated collaboration on global security management. To do so, a common framework is needed that can be constructed using the guidance available in ISO/IEC 27110 (guidelines for cybersecurity framework). Such a framework includes five concepts: identify, protect, detect, respond and recover. Issues to be addressed include access control, usage control, trust management, and identity management.

Federated privacy management consists of having individual data spaces' privacy management associated with a federated collaboration on global privacy management. To do so, a common framework is needed, the construction of which can be based on an extension of ISO/IEC 27101 (e.g. NIST privacy framework[19]), identifying five related concepts: identify privacy, govern privacy, control privacy, communicate privacy, and protect privacy. Furthermore, guidance from current and upcoming privacy standards can be used.[20]

Federated assurance management consists of having individual data spaces' assurance management associated with a federated collaboration on global security and privacy assurance management. We recommend working on the agreement on consistent individual assurance first, before work on the definition of federated assurance starts.

[19] https://www.nist.gov/privacy-framework
[20] Just to name a few : ISO/IEC 29100, ISO/IEC 29134, ISO/IEC 27101, ISO/IEC 27550, ISO/IEC 27570, ISO 31700.

1.7 Towards effective and efficient data sharing in data ecosystems

Creating innovation-driven data space ecosystems requires to address three basic concerns: extensibility, replaceability, and independent evolution.

» **Extensibility** means 1) an ecosystem's ability to dynamically onboard new participants, 2) the ability to create new data value chains based on data provided by existing participants, and 3) the ability to extend existing data value chains, thus bringing about innovation by exploiting value not foreseen initially (otherwise the data value chains supported by data spaces would be limited to those value chains designed by existing participants).

» **Replaceability** means the ability to replace existing participants without affecting the data value chains they were involved in. This will ultimately translate into avoidance of vendor lock-in, and thus better protection of end users' investments.

» **Independent evolution** means that each participant can evolve independently of other participants, as long as the interfaces for interaction with these other participants are respected. Extension of the number of participants in a data space, or replacement of any of them, should not affect the way each participant evolves.

Data providers joining data spaces must be able to publish data resources knowing that data consumers unknown to them a priori know how to retrieve and consume these resources. Otherwise, the above mentioned concerns will not be addressed properly (the World Wide Web works on exactly the same principle, i.e. content providers publishing web pages on web servers knowing that users use web browsers to retrieve these web pages and view the contents).

This means that all participants in data spaces should 'speak the same language', which translates into adopting common APIs and security schemas for data exchange (in the analogy with language: the way of constructing sentences) together with data models that can be represented in data formats compatible with those APIs (the vocabulary used in these constructed sentences). Furthermore, data spaces should include means for publishing and retrieving data resources accessible through the defined APIs and using the defined data models. Data spaces leveraging building blocks as defined under the Connecting Europe Facility (CEF) program, for example, could rely on the ETSI NGSI-LD standard API for exchange of right-time values of digital attributes of twin entities representing real-world assets. Data resources available could be published for retrieval on data portals like the European Data Portal.

1.8 Supportive regulations and recommendations

Regulation in the field of data sharing and exchange will be driven by the FAIR principles (Findability, Accessibility, Interoperability, and Reusability), as exemplified by the Commission's call to foster a trusted environment for sharing and analysing data from all publicly funded research.[21] This will have a far-reaching impact on data spaces and federation capabilities. For establishing European data spaces, the following EU regulations and recommendations should be taken into account:

» **GDPR** (Global Data Protection Regulation) is an EU regulation on data protection and privacy, which became effective on May 25th, 2018.[22] GDPR's primary goal is to give individuals control over their personal data and simplify the regulatory environment for international business. Applied to data space ecosystems, there will be a need to ensure compliance with GDPR for each stakeholder (data consumers, data providers, data application providers, and data space capability providers).

» **eIDAS** (electronic IDentification, Authentication and Trust Services) is an EU regulation on electronic identification and trust services for electronic transactions in the European Single Market, coming into effect July 1st, 2016.[23] eIDAS oversees electronic identification and trust services for electronic transactions in the European Union's internal market. A further initiative of the European Commission called Connecting Europe Facility[24] is funding a set of digital service infrastructures [25](DSI), including the Context Broker and European Blockchain Services Infrastructure (EBSI).

» **PSD2** (Payment Services Directive) is an EU regulation[26] directed towards payment services and payment service providers in Europe. The goal of PSD2 is to foster a more integrated European payment market, based on safer payments and protection of consumers.

» **Context Broker** (CEF Building Block)[27] is recommended by the EC for the integration and sharing of collected data, including insights for further exploitation under the Connecting Europe Facility program.

» **EBSI (**European Blockchain Service Infrastructure**,** CEF Building Block)[28] is a network of distributed nodes across Europe to deliver cross-border blockchain/DLT services that citizens, governments, and businesses may rely on during their interaction.

[21] https://ec.europa.eu/info/research-and-innovation/strategy/goals-research-and-innovation-policy/open-science/eosc_en
[22] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679
[23] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=uriserv:OJ.L_.2014.257.01.0073.01.ENG
[24] https://ec.europa.eu/inea/en/connecting-europe-facility
[25] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/All+CEF+services
[26] https://eur-lex.europa.eu/eli/dir/2015/2366/oj/eng
[27] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Context+Broker
[28] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/EBSI

1.9 Challenges for European data spaces

There are three types of challenges to be addressed when establishing European data spaces: technical, organisational and economic.

Technical challenges mainly relate to specifying and supporting the data-sharing lifecycle, managing data ownership policies, incorporating data provenance mechanisms, and defining a decentralised architecture agreed upon by all relevant stakeholder groups:

» *Specifying and supporting the data-sharing lifecycle* requires specific building blocks. What is needed is a common understanding of the different data lifecycle phases (i.e. data acquisition, preprocessing, processing, analysis, storage, and removal[29]). For this, a common framework for data-sharing agreements[30] needs to be developed, as well as a common understanding and definition of the data lifecycle (template definition, authoring, analysis, mapping, enforcement, disposal[31]).

» *Managing data ownership policies* is not subject to debate from a legal viewpoint only,[32] but requires also agile and standardised ICT capabilities when such agreements are to be negotiated dynamically. What needs to be established is semantic interoperability of policies (i.e. both parties in a data exchange transaction understand the meaning of a given policy) and behavioural interoperability of policies (i.e. both parties implement the policy with the same understanding).

» *Incorporating data provenance mechanisms* is key to a data-sharing ecosystem based on data sovereignty and trustworthiness. It is mandatory to provide information about the creation, update, transcription, abstraction, validation, and transfer of data ownership.[33] Furthermore, data provenance is a critical enabler of security and privacy in case there is a need for attribution of actions (e.g. if it is important to qualify the origin of data used to train artificial intelligence systems). For establishing European data spaces, existing models characterising data provenance can be adopted. To do so, it would be possible to leverage the work of W3C[34] or to draw upon the ISO/IEC work item to define a data provenance reference model.[35]

[29] ISO/IEC 20547-3:2020 — Big data reference architecture — Part 3: Reference architecture
[30] A standard framework is currently being developed: ISO/IEC 23751 Data sharing agreement (DSA) framework. See https://www.iso.org/standard/76834.html
[31] See the following contribution from the cococloud FP7 project: https://ercim-news.ercim.eu/en100/special/engineering-the-lifecycle-of-data-sharing-agreements
[32] See analysis from prof Thomas Hoeren, https://www.uni-muenster.de/Jura.itm/hoeren/veroeffentlichungen/Big_Data_and_the_Ownership_in_Data.pdf
[33] ISO 8000-2 :2020 Data quality – part 2 : vocabulary
[34] https://www.w3.org/TR/2013/REC-prov-dm-20130430/
[35] A preliminary work item has started in ISO/IEC JTC1/SC27 WG4. It is considering layers such as the physical layer, the provenance link layer, or the attribution layer.

» *Defining a decentralised architecture agreed upon by all relevant stakeholder groups* requires to combine individual data space architectures into a federated architecture. Combining different architectures requires agreement on how to describe architectures[36] and how to use common building blocks. While we can leverage significant work on architectures,[37] there is a need to create a global consensus on how individual architecture specifications can be combined. To address this topic, we can leverage the work carried out by ISO/IEC JTC1.[38]

Organisational challenges mainly relate to supporting practices for trust management and adapting data spaces to the specific needs of discrete ecosystems:

» *Supporting practices for trust management* means supporting practices for security management, privacy management, and assurance management. Security management should mainly refer to governance issues (i.e. security policies, security risk analysis, security-by-design, situation awareness, preparedness, and cybersecurity incident management). Privacy management should mainly include governance issues, privacy management policies, privacy impact assessment, privacy-by-design, preparedness, and privacy breach management. Assurance management should mainly include practices for audits and certification.

» *Adapting data spaces to the specific needs of discrete ecosystems* depends on the given requirements of a domain, sector, or territory. For instance, a data space for the energy industry should take into consideration the specific needs of the stakeholders of that domain (distribution system operators, transmission system operators, etc.), and it should support the constraints of the smart-grid cyberphysical system (e.g. when building a digital twin, so that data exchange has an effect on grid operation). Likewise, a data space for a smart city should take into consideration the multitude and variety of possible cross-domain interactions (energy, transport, health, etc.), as well as citizen engagement practices.

[36] For instance, ISO/IEC/IEEE 42010 Architecture description language
[37] For instance, IDSA reference architecture, Gaia-X reference architecture and H2020 projects.
[38] ISO/IEC JTC1/AG8 Meta Reference Architecture and Reference Architecture for Systems Integration

Economic challenges mainly relate to finding the balance between ethical/societal concerns and economics, providing agile support of data monetisation models, and providing mechanisms for incentivising data sharing and exchange:

» *Finding the balance between ethical/societal concerns and economics* is a key factor for acceptance, which goes beyond mere compliance with regulations. Furthermore, it is a moving target, which means that concerns can change over time. On this topic, several reports are available, such as the Ethics Guidelines for Trustworthy AI issued by the Commission,[39] or the ISO/IEC standards on ethical and societal concerns.[40] It is important to note that this balance can also depend on the specific requirements of smart-city ecosystems, where citizen engagement schemes can influence data-sharing policies.[41]

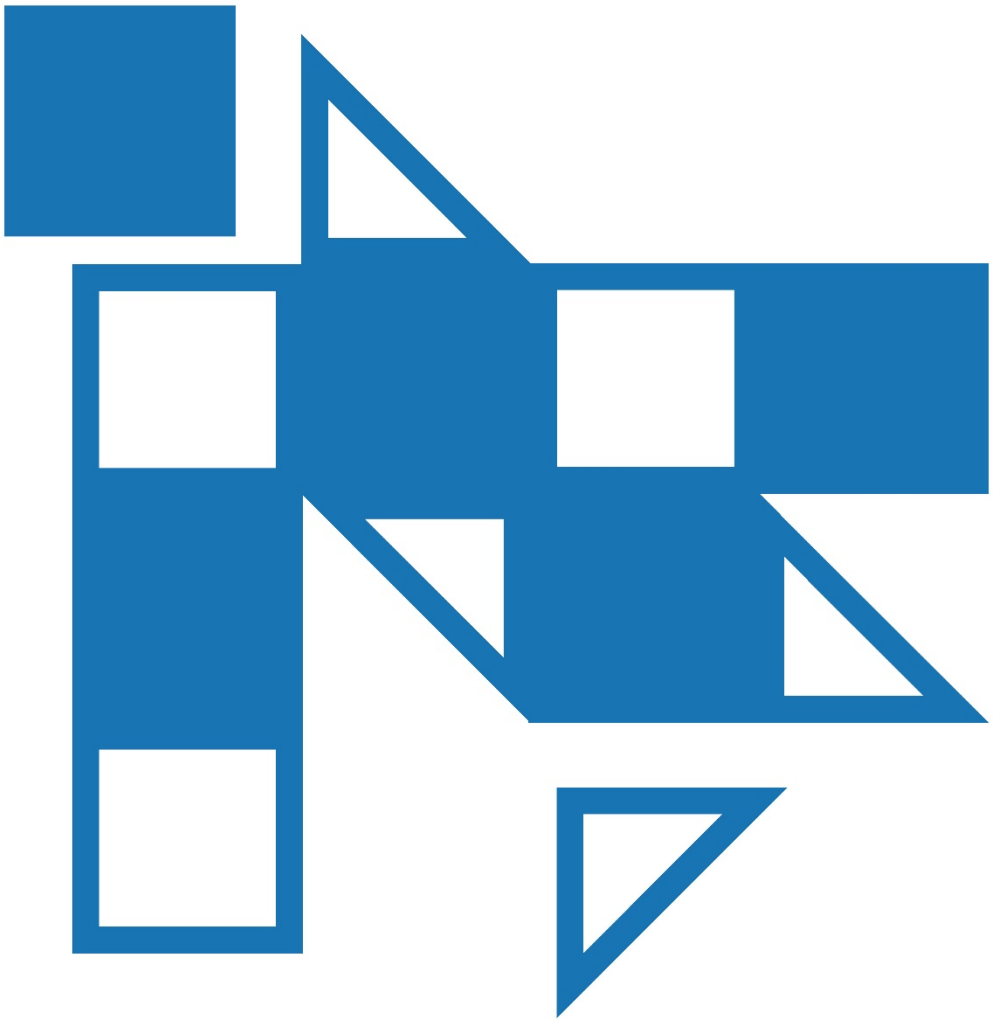
» *Providing agile support of data monetisation models* is a prerequisite for data spaces to strive for. Even if organisations and individuals recognise the possibility of extracting value from data, they often lack understanding and insight when it comes to assessing the potential market (also because there can be as many data monetisation models as there are business cases).[42] A spectrum of business models could include data as a service, insight as a service, and analytics-enabled as a service.[43] European data spaces should provide capabilities for agile support of such new data monetization models.

» *Providing mechanisms for incentivising data sharing and exchange* is necessary for allowing high-volume data sharing. This could involve schemes such as sponsoring the bootstrapping of a data-sharing ecosystem (e.g. in a smart city), inventing specific data monetisation models (e.g. an electricity utility sponsoring data providers), or facilitating crowdsourcing.[44] Despite the fact that many digital business models can only be successfully implemented through cross-company data sharing and exchange, many companies still are not aware of the added value that is generated as a result of sharing data not just with cooperation partners, but also with competitors. Therefore, the principle of quid pro quo should be taken into account by all stakeholders.

[39] https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai
[40] ISO/IEC 24368 – Artificial intelligence – overview of ethical and societal concerns, under development (https://www.iso.org/standard/78507.html)
[41] See for instance ISO/IEC 27570 Privacy guidelines for smart cities that will be publish shortly (https://www.iso.org/standard/71678.html)
[42] J.Baecker, M.Engert, M.Pfaff, H.Krcmar. Business strategies for data monetization: deriving insights from practice. March 2020. DOI: 10.30844/wi_2020_j3-baecker
[43] S.Gandhi, B.hota, R.Kuchembuck, J.Swartz. Demystifying Data Monetization. November 27, 2018. https://sloanreview.mit.edu/article/demystifying-data-monetization/#:~:text=The%20three%20primary%20external%20data,enabled%20platform%20as%20a%20service.
[44] J.Deichmann, K.Heineke, T.Reinbacher, D.Wee. Creating a successful Internet of Things data marketplace. McKinsey, October 2016. https://www.mckinsey.com/~/media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Creating%20a%20successful%20Internet%20of%20Things%20data%20marketplace/Creating-a-successful-Internet-of-Things-data-marketplace.pdf

Building Blocks

2



2 Building Blocks

2.1 Introduction

Now that we understand the fundamentals of data spaces, it is key to understand which elements together form the building blocks of data spaces. In this chapter, we will address a broad range of general building blocks that enable technical, business, operational and organisational capabilities of data spaces from two perspectives: 1) the perspective of an essential soft infrastructure and 2) the perspective of the services that form data spaces within and across domains.

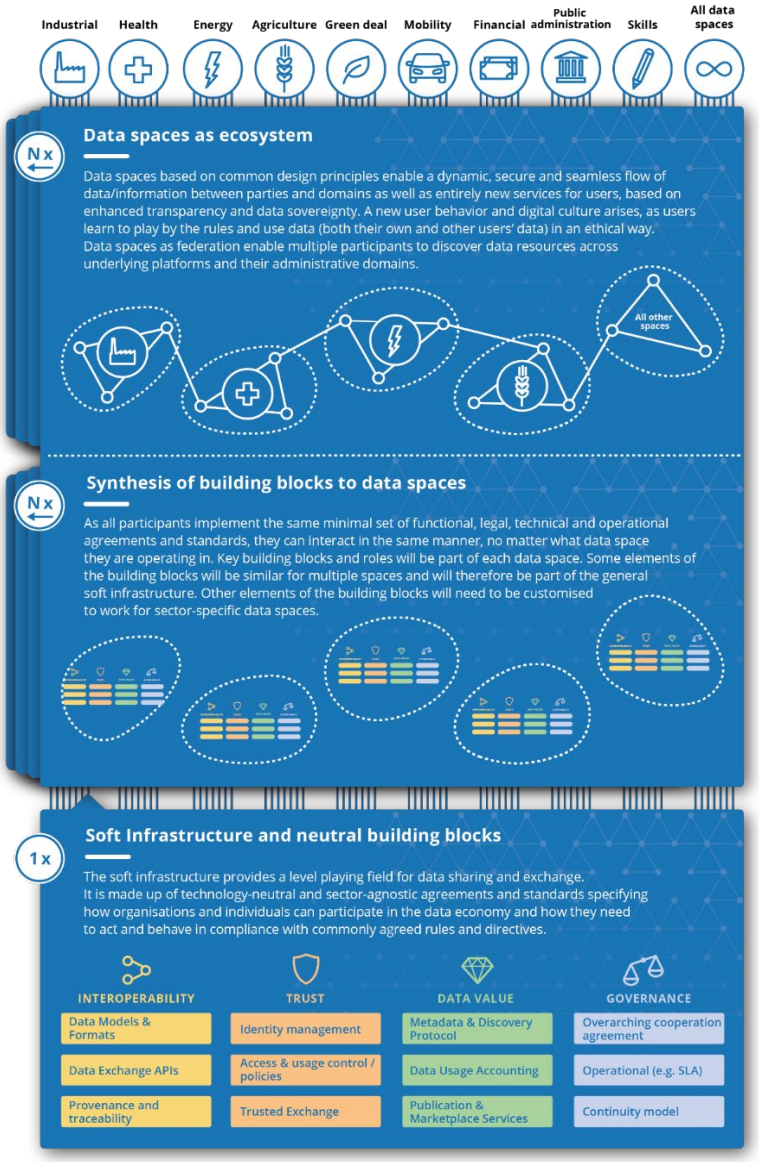


Figure 7 General soft infrastructure stack

2.2 Concept and taxonomy of building blocks

The design and implementation of a data space comprises a number of building blocks, which fall under two types: the technical building blocks and the governance building blocks (as illustrated in figure 8):

Technical building blocks.

The building blocks subsumed under this category enable the implementation of the technical architecture of a data space. They include network protocols, middleware components, (standardized) APIs, and more, facilitating the sharing of data between different parties in a secure and trustworthy fashion. A variety of technical components for building data spaces have been developed or adopted by different initiatives in Europe, such as FIWARE[45], Platform Industrie 4.0[46], CEF Digital[47], or the International Data Spaces Association[48].

These components address most of the technical concerns associated with the creation of data spaces, linked to:

Data interoperability, covering aspects such as data exchange APIs, data representation formats, as well as data provenance and traceability;

Data sovereignty, covering aspects such as identity management, trustworthiness of participants, as well as data access and usage control;

Data value creation, covering aspects such as publication of data offerings, discovery of such offerings based on metadata, and data access/usage accounting, which are essential to handle data as an economic asset.

Technical building blocks enable (plug & play) integration of different systems and platforms used by data space participants beyond the security limits of each participant. Additional technical building blocks may optionally be considered for facilitating creation of systems plugged into a data space (e.g., for implementing big-data analysis, supporting data visualization and analytics, or providing an interface with IoT networks). These building blocks enable data usage in data spaces beyond current business capabilities of participants, and lead to new business cases and data usage scenarios.

[45] https://www.fiware.org/about-us/
[46] https://www.plattform-i40.de/PI40/Navigation/EN/Home/home.html
[47] https://ec.europa.eu/cefdigital
[48] https://internationaldataspaces.org

Governance building blocks:

The building blocks subsumed under this category refer to business, operational and organisational agreements among data space participants. These agreements are enforced through legal frameworks participants have to adhere to, or via technical building blocks:

Business agreements comprise service level agreements (SLAs), data usage and access control policies as well as accounting and pricing/billing/payment schemes, which data service providers may specify in connection with their offerings to govern their interaction with data consumers. Such agreements specify the terms and conditions that regulate the sharing and exchange of data between parties. To do so, smart contracts can be used that connect legal and organizational agreements to technically enforceable and measurable agreements.

Operational agreements regulate policies that need to be enforced during data space operation. For example, they comprise terms and conditions dealing with the ever-growing importance of compliance with mandatory regulations like GDPR (General Data Protection Regulation) or the 2nd Payment Services Directive (PSD2) in the finance sector.

Organisational agreements comprise terms and conditions regarding governance bodies and procedures established for a data space.

For integration of building blocks in data spaces, different sets of structuring principles can be applied to different architectures, depending on domain-specific requirements or technical requirements (e.g., streaming of data, high-frequency data, or event processing). Nevertheless, there are some guiding principles that need to be respected for all implementations, such as decentralisation, scalability, collaboration support, federation, interoperability, compatibility, trust management, and auditability.

Figure 8 illustrates how a data space can be created through synthesis of a collection of building blocks, which are integrated in line with the technical architecture, the business structure, and the policy requirements of the data space.

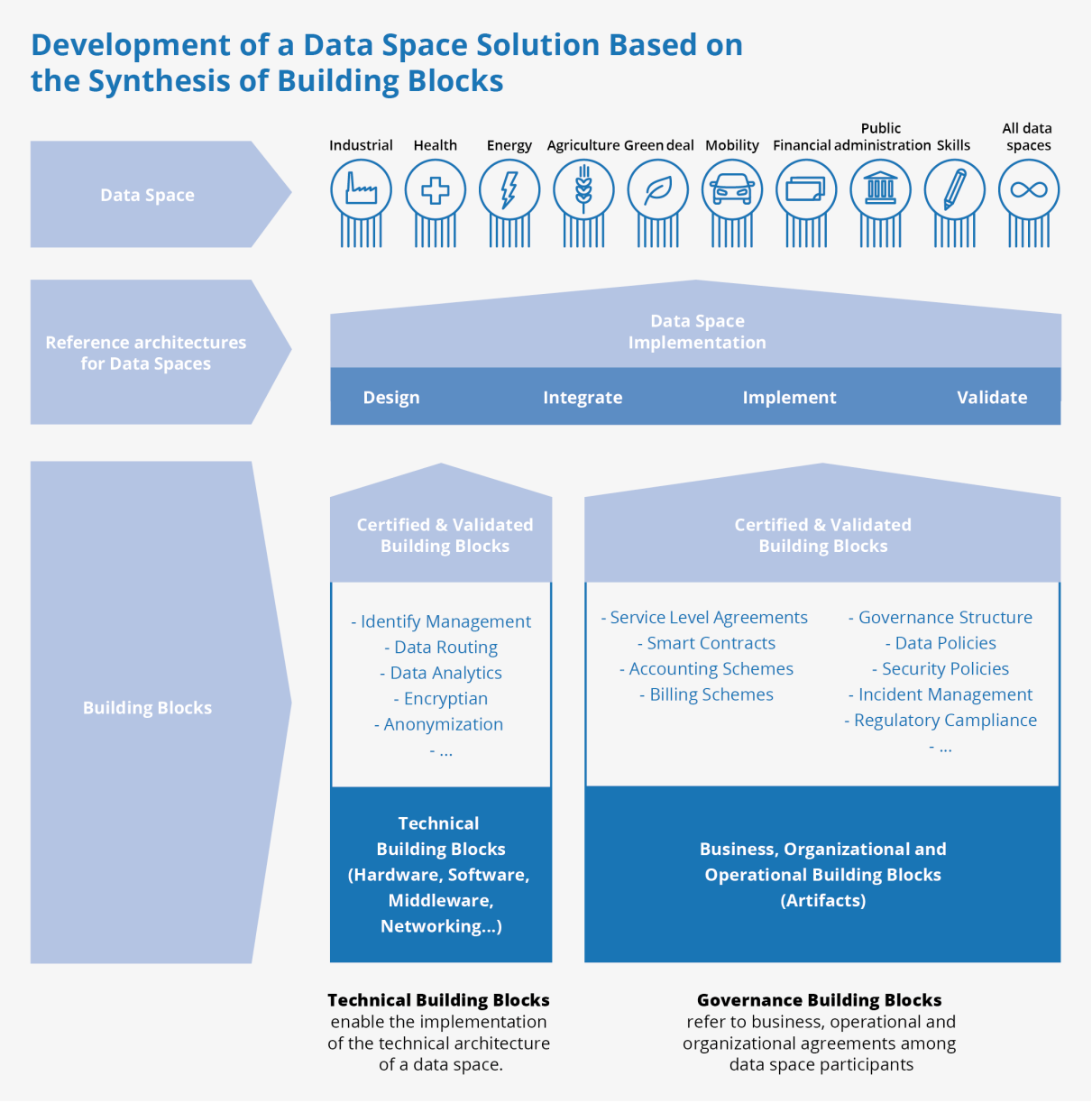


Figure 8 Data space solution based on the synthesis of building blocks

The different building blocks can be specified and developed independently of each other. When doing so, existing norms, standards, and best practices should be used to ensure cohesion of building blocks. Each data space solution can integrate multiple building blocks, as long as they are in line with data space reference architectures (e.g., the IDS Reference Architecture Model[49]). The building blocks presented in this chapter are core elements of any data space. As such, they can be considered sector-agnostic. Nevertheless, they can be used in sector-specific scenarios (see e.g. seaport scenarios).

Data space stakeholders may also define additional building blocks to support innovative features and functions. For instance, data space architects may introduce building blocks that enable novel types of data space architectures combining centralised and decentralised approaches. Likewise, business stakeholders may introduce building blocks that enable novel forms of smart contracts to be agreed upon by participants of a data space, thereby facilitating business model innovations. Hence, the building blocks presented in this chapter are not exhaustive, but rather indicative of the elements of a data space.

In general, each building block consists of reusable, generic components (i.e. which can be used across domains and industries) and more specific components (i.e. to meet requirements and regulations that are specific for certain industries, domains, or even concrete use cases). This allows individual participants to join different data spaces, use data in multiple contexts and scenarios, and be part of multiple data value chains.

The taxonomy of building blocks constituting a data space will benefit various stakeholders:

» **Data space architects and integrators** will be provided with a structured way for identifying the components required for their specific architecture.

» **Building block developers** will be a given a clear description on how their components can fit into the specific architecture of a data space.

» **Business managers** will be able to identify business model candidates for data sharing and exchange (including data monetisation) in the context of a specific data space.

» **Data space policy managers** will have access to building blocks that can be used to specify and enforce data space related policies (e.g., data access policies, privacy control policies).

[49] Reference Architecture Model for the Industrial data space Otto, Boris & https://www.researchgate.net/publication/316854975_Reference_Architecture_Model_for_the_Industrial_Data_Space

The picture below identify the building blocks classified by four categories: interoperability, trust, data value, governance. Furthermore, the building blocks can be of three types: the technical building blocks; business building blocks and operational building blocks.

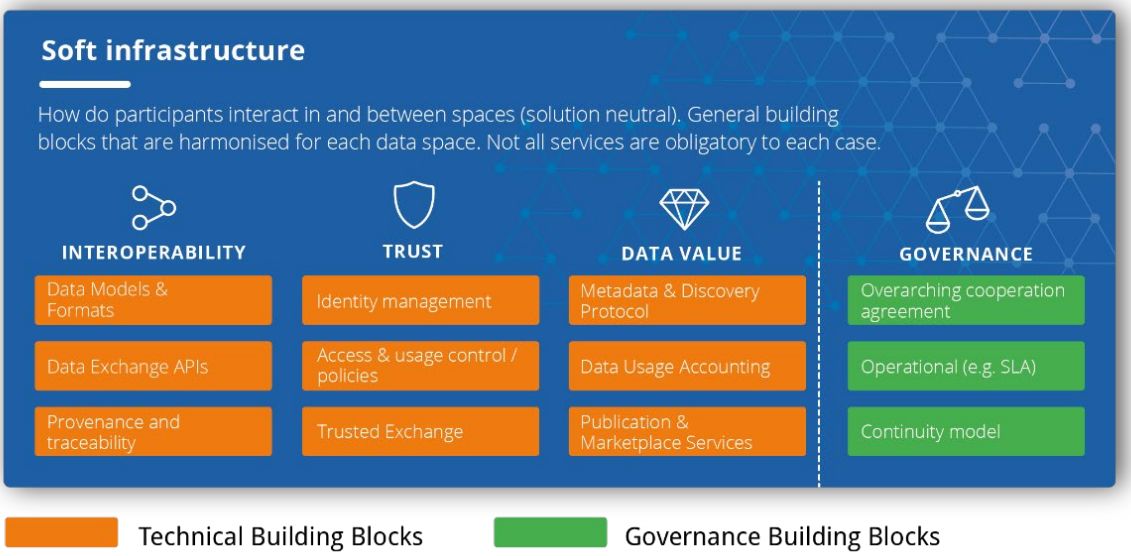


Figure 9 Data spaces building blocks

In the following three subchapters, the different types of data space building blocks will be illustrated in more detail.

2.3 Technical building blocks

From a technical perspective, a data space can be understood as a collection of technical components facilitating a dynamic, secure and seamless flow of data/information between parties and domains. These components can be implemented in many different ways and deployed on different runtime frameworks (e.g., Kubernetes). They can be categorised as follows:

Building blocks facilitating data interoperability:

These building blocks should be deployed by all data providers and data consumers participating in a data space. This way, each data provider can be sure that any data published can be technically consumed by any data consumer entitled to do so, while each data consumer

can be sure they are able to technically access and use any data made available by any data provider selected.

The following building blocks belong to this category:

» **Data Models and Formats:** This building block establishes a common format for data model specifications and representation of data in data exchange payloads. Combined with the Data Exchange APIs building block, this ensures full interoperability among participants.

» **Data Exchange APIs:** This building block facilitates the sharing and exchange of data (i.e., data provision and data consumption/use) between data space participants. An example of a data interoperability building block providing a common data exchange API is the 'Context Broker' of the Connecting Europe Facility (CEF)[50], which is recommended by the European Commission for sharing right-time data among multiple organisations.

» **Data Provenance and Traceability:** This building block provides the means for tracing and tracking in the process of data provision and data consumption/use. It thereby provides the basis for a number of important functions, from identification of the lineage of data to audit-proof logging of transactions. It also enables implementation of a wide range of tracking use cases at application level, such as tracking of products or material flows in a supply chain.

Building blocks facilitating data sovereignty and trust:

» **Identity Management (IM):** The IM building block allows identification, authentication, and authorisation of stakeholders operating in a data space. It ensures that organisations, individuals, machines, and other actors are provided with acknowledged identities, and that those identities can be authenticated and verified, including additional information provisioning1, to be used by authorisation mechanisms to enable access and usage control. The IM building block can be implemented on the basis of readily available IM platforms that cover parts of the required functionality. Examples of open-source solutions are the KeyCloak infrastructure[51], the Apache Syncope IM platform[52], the open-source IM platform of the Shibboleth Consortium[53], or the FIWARE IM framework[54]. Integration of the IM building block with the eID building block of the Connecting Europe Facility (CEF)[55], supporting electronic identification of users across Europe, would be particularly important.

[50] https://ec.europa.eu/cefdigital
[51] https://www.keycloak.org/
[52] https://syncope.apache.org/
[53] https://www.shibboleth.net/
[54] https://github.com/FIWARE/catalogue/tree/master/security
[55] https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/eID

Creation of federated and trusted identities in data spaces can be supported by European regulations such as EIDAS.

» **Access and Usage Control/Policies:** This building block guarantees enforcement of data access and usage policies defined as part of the terms and conditions established when data resources or services are published (see 'Publication and Services Marketplace' building block below) or negotiated between providers and consumers. A data provider typically implements data access control mechanisms to prevent misuse of resources, while data usage control mechanisms are typically implemented on the data consumer side to prevent misuse of data. In complex data value chains, both mechanisms are combined by prosumers. Access control and usage control rely on identification and authentication.

» **Trusted Exchange:** This building block facilitates trusted data exchange among participants, reassuring participants in a data exchange transaction that other participants really are who they claim to be and that they comply with defined rules/agreements. This can be achieved by organisational measures (e.g. certification or verified credentials) or technical measures (e.g. remote attestation).

Building blocks facilitating data value creation:

» **Metadata and Discovery Protocol:** This building block incorporates publishing and discovery mechanisms for data resources and services, making use of common descriptions of resources, services, and participants. Such descriptions can be both domain-agnostic and domain-specific. They should be enabled by semantic-web technologies and include linked-data principles.

» **Data Usage Accounting:** This building block provides the basis for accounting access to and/or usage of data by different users. This in turn is supportive of important functions for clearing, payment, and billing (including data-sharing transactions without involvement of data marketplaces).

» **Publication and Marketplace Services:** To support the offering of data resources and services under defined terms and conditions, marketplaces must be established. This building block supports publication of these offerings, management of processes linked to the creation and monitoring of smart contracts (which clearly describe the rights and obligations for data and service usage), and access to data and services.

Based on the technical needs, the corresponding backend processes for rating, clearing, and billing can be executed. The building block thereby facilitates dynamic enlargement of data spaces with more stakeholders, data resources, and data-processing/analytics services (such as big-data analysis services, machine learning services, or services based on statistical processing models for different business functions). It should comprise capabilities for publishing data

resources following the broadly accepted DCAT (Data Catalogue Vocabulary) standards, and for harvesting data from existing open-data publication platforms.

Additional technical building blocks

Additional technical building blocks may optionally be considered to facilitate connection of additional systems to data space (e.g., to establish data value chains):

» **System Adaptation:** One of the main functions of a data space is to facilitate the transfer of data to and from participants' systems (i.e. database systems, data-processing systems, enterprise systems [like CRM, ERP, MRP or MES systems], but also cyberphysical systems and IoT-enabled systems). Regardless of the system, there is a need for a System Adaptation building block that interfaces with the various data resources exported by the system and performs the necessary transformation of the data formats adopted for data exchange within the data space (see 'Data Exchange APIs' building block). The interface depends on the nature of the system: For example, IoT protocols (e.g. CoAP [Constrained Application Protocol] or MQTT [Message Queuing Telemetry Transport]) can be used to interface with IoT resources, database protocols (e.g. JDBC [Java Database Connectivity] or SQL [Structured Query Language]) can be used to interface with databases, and API protocols (e.g. RESTful services) can be used to interface with enterprise systems and applications. To maintain confidentiality and privacy when transferring data from participants' systems to the data space and vice versa, data encryption and anonymization may be required. Additional data and metadata may also be incorporated in order to transport relevant information required for other data space building blocks (e.g., on data accountability/traceability or usage control) to work.

» **Data Processing:** Systems connected to data spaces via system adapters are able to process shared data to enforce data usage restrictions. The organisations' rules or legal contracts can be substituted or at least accompanied by technical solutions. However, such processing adds additional complexity to data usage control by data providers or data space operators.2 In order to exert wider control, data spaces may incorporate different stages of usage control, as provided by the concept of the usage control onion, (e.g. for data accountability/traceability or access/usage control) with data-processing technologies.

» **Data Routing and Preprocessing (DR&P):** There may be a need for dynamic routing of data to the proper data-processing node (as part of a dynamic data-routing function). The building block for data routing and preprocessing is usually a data middleware platform (or a combination of two or more such platforms). These address different technical requirements, depending on the nature of the data that is collected and routed (e.g. streaming data, data at rest). For instance, stream-processing middleware platforms (e.g. Apache Kafka) can be used to support the routing and preprocessing of streaming data. Data routing needs to consider technical aspects, like horizontal and vertical scalability, but

also aspects resulting from data usage policies, like jurisdiction for data processing, data egress, or combination with other data.

» **Data Analytics Engine (DAE):** Many data space use cases allow analysis of multi-source, multi-stakeholder data based on methods like statistical analysis, machine learning, deep learning, and other data-mining techniques (e.g. for demand forecasting in an industrial use case, which must synthesise and analyse multiple data flows coming from different platforms the data space is comprised of). A function like that requires analysis of multiple data flows, which is why it needs to be supported by a 'Data Analytics Engine' building block. Depending on the nature of the data, this building block can take different forms (such as streaming analytics, cloud-based analytics, machine learning, or complex event processing [CEP]).

» **Data Visualisation:** Data spaces should also provide data presentation and visualisation features. A building block offering these features can take various forms, from a simple dashboard to augmented analytics (e.g. implemented on the basis of frameworks like Kibana or Grafana).

» **Workflow Management Engine (WME)** Data-processing use cases usually involve interaction of multiple data sources, data consumers, and data services. This interaction must be properly orchestrated by means of structured and acknowledged workflows (including data extraction, transformation, and analysis, as well as data presentation and visualization).

A data spaces is defined as a federated data management (i.e. federation of the data platforms/sources of the various business actors). Such federation enables multiple participants to discover data resources across underlying platforms and their administrative domains. It can be implemented on the basis of some of the above-listed building blocks, like 'Data Resources and Services Publication and Discovery'. Federate data management enables data spaces to operate in a federated and decentralised fashion. In a circular economy, multiple federation of different participants/stakeholders should be able to share and exchange data over the circular supply chain. The data spaces decentralised soft infrastructure principle establishes and supports sector-specific federations (e.g. in the automotive or the plastics industry). Furthermore, it is possible to configure data spaces to establish cross-sector federations (e.g. the plastic industry reusing waste produced by the automotive industry).

Table 2 provides an overview of the various technical building blocks, along with some examples of using them in specific business scenarios. To justify our claim that the building blocks are not focused on single sectors, the given examples span multiple sectors (including healthcare, manufacturing, circular economy, smart cities, and digital finance).

Table 2: Overview of technical building blocks of data spaces

Technical Building block	Role and scope	Example
Data Models and Formats	Facilitates a common format for data model specifications and representation of data	The Smart Agrifood domain needs a common representation of agronomic data (e.g. crops, senso data from the field, multispectral imagery from UAVs, geolocation data, fertilisation logs, …). This common data model shall be used for all data exchanged between software components.
Data Exchange APIs	Facilitates data sharing and exchange between data space participants, ensuring semantic interoperability of data and data sources	A smart city needs to calculate its environmental performance on the basis of a collection and aggregation of information about all the sustainability projects in its urban environment. This information is shared by different stakeholders, who use different formats and semantics to report CO_2 emissions and other indicators. The building block enables syntactic and semantic harmonisation of the different data sources, as well as effective exchange of data using a common data exchange API to enable the calculation of the KPIs (key performance indicators) needed.
Data Provenance and Traceability	Enables traceability of provenance, access, and usage of data shared and exchanged in a data space	In the scope of a circular supply chain, there is need for providing end-to-end traceability of the status and conditions of key circular entities, like products or materials. The building block allows authorised participants to query on the status of specific products and materials, and to receive detailed information about their status and location in the circular chain.

Technical Building block	Role and scope	Example
Identity Management	Provides authentication and authorisation of data space participants	A user within an organisation registered with a data space provides his/her log-in credentials to the IM module in order to gain access to the data of the data space in line with his/her role in the organisation.
Access, and Usage Control/Policies	Enforces different data access and usage policies that ensure trustworthiness of data sharing and exchange between participants	Enforcing Data Protection Regulations in Health Care Applications. When a company is processing patient records for the sake of accounting an billing as a service to doctors and insurances, it is thus in the interest of the company to ensure that it complies to those regulations.
Trusted exchange	Facilitates trusted data exchange among participants	Trust is a necessary feature in any data-sharing environment, i.e. also for predictive maintenance. Unfortunately, predictive maintenance is difficult to achieve, as algorithms used are still not as effective as desired, and the quality of outcome often is not sufficient, due to a lack of reliable data. Nevertheless, integrating and leveraging data from partners – and even from competitors or companies from different sectors (OEMs, maintenance equipment producers, energy companies) – can be of great benefit for all participants. To overcome the lack of trust currently still prevailing, data sovereignty concepts and services should be employed.[56]

[56] Critical Success Factor for the Manufacturing Industry https://internationaldataspaces.org/download/21213/

Technical Building block	Role and scope	Example
Metadata and Discovery Protocol	Enables publication of offerings centred around data resources and services, making use of common descriptions of resources, services, and participants.	A data space participant analyses the terms and conditions linked to a given data resource and acquires the corresponding access/usage rights in line with these terms and conditions.
Data usage accounting	Facilitates the basis for accounting access to and/or usage of data by different users	The clearing house in the Smart Connected Supplier Network (SCSN) is used to send purchase-to-pay information in a business-to-business scenario. This information can be highly confidential and it is mission-critical for their day-to-day business. If dispute arises the clearing house is used as trusted third party to resolve this issue by comparing the fingerprint of the messages and identifying the error.[57]
Publication and Marketplace Services	Provides a directory of the various data assets for dynamic access and discovery as well as management of established contracts	A data space user queries the data resources publication platform on specific data assets (e.g. based on content, theme, industry, etc.). Upon selecting the dataset she/he wants to access, she/he receives a link (e.g. an URL) to the dataset chosen.

[57] New business models for data spaces, https://internationaldataspaces.org/wp-content/uploads/IDSA-Position-Paper-New-Business-Models-sneak-preview-version.pdf

Technical Building block	Role and scope	Example
System Adaptation	Supports connection of participants' systems to the data space, enabling them to provide and consume data	A smart city connects its system to a data space to publish information on air quality. At the same time, it connects to the data space for receiving weather forecast information in order to predict evolution of air quality parameters and take appropriate action. The building block interfaces with the smart-city system to enable access to the data resource providing the air quality information and to transfer the data on weather forecasts.
Data Analytics Engine	Supports execution of data analytics with regard to data shared and exchanged over the data space	In a data space in the field of digital finance, banks and other financial organisations can make credit risk assessments of participants (e.g. citizens, businesses, financial institutions) using statistical methods. To this end, they leverage machine-learning techniques and AI algorithms, the latter being implemented through the data analytics engine.
Data Visualisation	Provides data presentation and visualisation features for data shared and exchanged over the data space	In a data space in the field of digital finance, there is a need for visualising credit scores and other parameters of participants assessed. Diagrams and figures can then be presented over a proper dashboard, which can be set up with the help of this building block.

Technical Building block	Role and scope	Example
Workflow Management Engine	Enables implementation of entire data-driven business processes involving multiple interactions	In a manufacturing supply chain, data from various stakeholders (e.g. distributors, retailers) are aggregated for production forecasting. A function like this, which requires interaction between multiple data space participants, can be represented as a workflow. This workflow can then be executed by the workflow management engine.

As already mentioned, the set of technical building blocks listed and specified above is not exhaustive. Data space architects may incorporate more building blocks into their specific architecture. Furthermore, the above descriptions do not aim to dictate a specific technical implementation. That is the reason why we do not delve into the low-level implementation details of the building blocks, except for some indicative examples. Nevertheless, the types of building blocks specified provide a good indication of the technical elements required to build a data space infrastructure.

As outlined in the initial subsection of this chapter, the various building blocks require proven compliance (e.g. certification schemes) in order to ensure trusted operation of the data space. This means that the implementation of the various technical components should be available for auditing to data space participants. Such auditing can be performed by neutral and approved evaluators in a structured process specified by a certification scheme. Open-source software cannot just provide such auditing insights to a closed group of trusted evaluators, but puts a whole community in the position to review and audit (parts of) a system.

2.4 Governance building blocks

2.4.1 Introduction

It has often been often said that data is the "new oil". What is for sure is that data is the force pushing digital transformation of societies and economies, and the asset on top of which many new business models are built. The AI domain, for example, is completely dependent on data, from building models and analysing data to providing insights and creating new business value. Data is being continuously gathered everywhere around us and by a variety of entities using a plethora of methods. The importance of data starts to be acknowledged not just by

companies and individuals from the technological sector, but also by industry at large, from large corporations to small enterprises.

However, as data is a new type of asset that can be used and reused in different scenarios generating more or less business value (depending on context, availability, accuracy, etc.), common business models are not capable of adequately supporting the growing needs of business. It is therefore important to define and create a business framework capable of supporting new business constellations. This would help the stakeholders of an ecosystem understand both the potential relationships between each other and the underlying business model(s). Jointly, with adequate rules and policies for sharing and using data in place, these data-driven business ecosystems will be able to create new business value more rapidly.

2.4.2 Governance related roles in a data space

In general, the following groups of stakeholders can be identified in data-driven business ecosystems:

» **A Data Owner** is an entity which has the authority to decide how its data can be used by third parties. Depending on the service/solution and the business model put in place, this entity can acquire data on its own (manually or by using cyberphysical systems) or use tools and services of third parties to acquire data. Data can be stored on premises, on the edge, or in the cloud. Data owners can decide to keep their data private for internal use (for improvement of own processes, creation of new business value for commercial advantage, etc.) or share it publicly or with a limited number of third parties. In case data is made available to third parties in some form (mostly done with the help of a data provider), it includes rights and obligations (Data Usage Policies) and terms and conditions. Data Owners can make data available to third parties for free (for example, to help advance science and innovation with open data) or offer it against a fee, depending on the business model.

» **A Data Acquirer or Data Provider** is an entity responsible for collecting and preprocessing data and providing it to others on behalf of a Data Owner (often as part of a business-related service provided to a Data Owner). For example, a company providing asset monitoring services (e.g. for fleet management) deploys tracking devices and collects data on the client's vehicles (time data, location data, etc.). While the data is collected, processed, and stored by the Data Provider for the benefit of the client (i.e. the Data Owner), subsequent use of the data usually remains under control of the client. Recently, new business models have emerged on the basis of which Data Providers/Acquirers offer their services at reduced prices in return for the opportunity to use their client's anonymized data to further improve existing services or create new services generating new business value. The Data Provider offers technical means to the Data Owner to enable trustworthy data

exchange and data sharing with other participants in the data space (including data usage monitoring, if requested).

» **A Data Processor** is an entity responsible for, and interested in, using certain types of data to create new services to be offered in the market. The spectrum of such services is very broad, ranging from domain-specific use cases to cross-domain applications. The value of the data used for creation of new services depends on the data's accuracy, availability (i.e. the number of data providers offering it), and how important it is for the processing algorithms used. It is usually estimated and agreed upon upfront, what to some extent limits the data owner's ability to achieve maximum monetisation of their data, as they have no sound understanding of the additional value created on top of their data and/or the value the new services have for users. As data usage control is based on conventional contract documents set up by the parties involved, leading to dependency on manual/back-office operations, utilisation of data is further complicated, thus slowing down full exploitation and monetisation of data.

» **A Data Marketplace Operator** is an entity providing different kinds of infrastructure (e.g. soft infrastructure, [cloud] hardware, data-processing tools). Furthermore, it is responsible for marketplace governance by providing support services, defining terms and conditions, and deciding on admission and withdrawal of datasets or participants. As the importance and potential of data is more and more acknowledged, data marketplaces are emerging as a new type of business offering. Their goal is to make data usage possible in a seamless and automated fashion, bypassing the need for complicated back-office contracts and agreements. A data marketplace can be cross-domain or domain-specific (i.e. dealing with data of interest to specific use cases and industries). Their main duty is to make data easily discoverable (based on a set of standardised data models) and to provide transparent tracking of all data-related transactions (from who used what data at what point in time to revenues generated from sharing data). Data Marketplace Operators need to put in place mechanisms ensuring compliance with data usage policies (e.g. with regard to the time and count the data was used, or fields of application in which certain data cannot be used).

Figure 10 illustrates the relationships and interactions between these entities and the flow of data taking place between them.

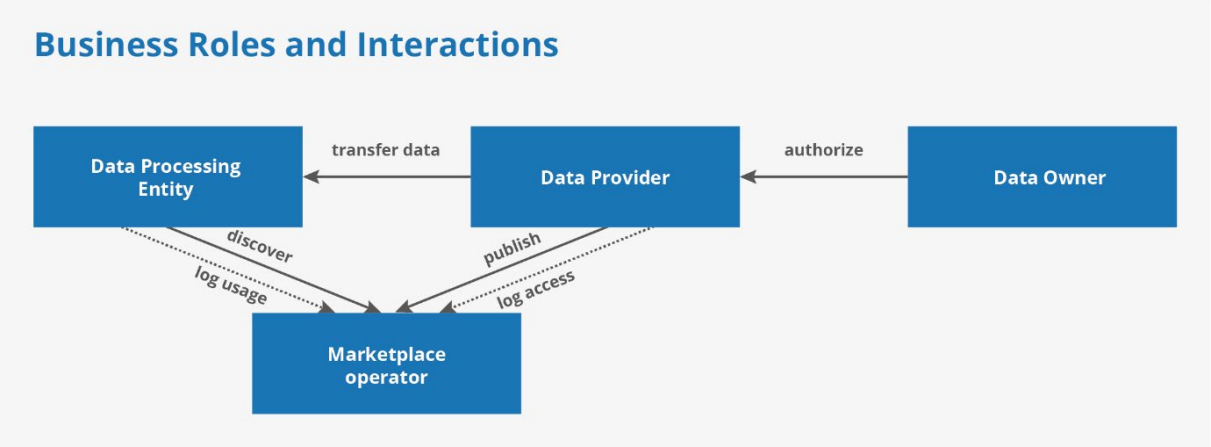


Figure 10 Business roles and interactions

As data is the most important factor of a data-driven business, it is crucial to use quality data in terms of accuracy, reliability, resolution, availability, etc. Responsibility to ensure that data is of the quality required and agreed upon starts with the Data Owner and then goes over to each party involved in the data value chain (e.g. from ensuring that sensors are properly installed and calibrated to making sure that adequate data-processing algorithms are applied properly). To ensure an uninterrupted and reliable data value chain, service level agreements (SLAs) can be put in place between the stakeholders. As SLAs are typically bound to service provisioning, they can be part of a smart contract between Data/Service Provider and Data/Service Consumer (in addition to the data usage policy specified).

An important aspect of business relationships is compliance with the rules specifying how data can be used by third parties. Beyond backoffice agreements, the ability of Data Owners to control how their data is being used is rather limited. Enforcement of data usage restrictions can be implemented in different ways.[58] Organisational rules or legal contracts should be accompanied by technical solutions and vice versa. Comprehensive usage control requires control points in different layers of the data-processing systems. Implementation of such control points can be of different complexity (e.g. control of data in motion is not as complex as control of data being stored, visualised on a screen or printed).

[58] Usage Control in the International Data Spaces, https://internationaldataspaces.org/download/21053/

2.4.3 Business building blocks

Business building blocks are artifacts that regulate the business relationships between the above listed roles.

Overall, a data space comprises the following business building blocks:

Operational Service Level Agreement (SLA): An SLA is a contract between an entity providing data services and an entity using these services within the scope of a data space (e.g. between a Data Provider and a Data Consumer, or between a Data Owner and a Data Provider). It describes the services to be provided by the Data Provider, along with the standards that the provision of the services should meet. The services provided must meet specified standards, while the user of the services is entitled to demand that these standards are met.

» **Accounting Scheme:** This artifact details the accounting practices and reports that should be produced as part of the operation of the data space and in line with the underlying business models. It specifies the parameters that should be logged and reported for every business actor and transaction of the data space.

» **Billing/Charging Scheme:** Leveraging accounting data and reports, data spaces provide the means for billing of services and transactions. In this context, specification of the billing/charging scheme is important. This artifact specifies how billing/charging is to be performed. Commonly used billing/charging schemes are schemes relying on the data volume provided (i.e. volume-based), the number of requests for, or connections to, a service (i.e. I/O based) or the time period the service can be used (i.e. time-based). While in some cases flat billing/charging schemes may be an adequate solution (as they are simple to set up and use), it is also possible to combine the above-listed schemes into a hybrid scheme.

» **Data valuation method:** data evaluation is concerned with methods to estimate the value of data shared by organizations in the data space.

» **Smart Contracts:** They provide a protocol for implementation of agreements between two or more parties (mainly the Data Provider and the Data Consumer). As such they specify data usage policies, legal aspects, SLAs and other agreements in a machine-readable and cryptographically signable manner.

These business building blocks can be implemented through the technical building blocks presented in the previous subsection. This means that specific artifacts (e.g. business agreements between parties) impose different configurations of the technical components of

the data space. In some cases, these artifacts may also impose restrictions regarding the way different technical components interact with each other (e.g. for reflecting centralised or decentralised interactions between the different actors).

The 'Data Resources and Services Marketplace' building block, for example, will typically provide the artifacts for auditing and processing protocols of data exchange transactions taking place between two or more parties. More specifically, it will keep track of the parties' interactions, audit their transactions against applicable rules (e.g. as specified by the terms and conditions of an SLA), identify deviations and remedial actions, and settle payment and billing. In this way, it will reduce the risk of participation in a data space in general, and the risk of engaging in a certain data exchange transaction in particular. It will also help ensure that each party lives up to its contractual commitments. To do so, it may audit proper implementation of certain terms and conditions via smart contracts, specifying transaction protocols that automate the execution of actions required for provision of a service by one or more parties. If a smart contract is used, it will operate in line with the terms and conditions specified by one or more SLAs (i.e. it will comprise legally relevant events listed in a contract or in the SLAs) and data usage policies.

The following table provides an overview of the various business building blocks, along with examples of using them in different sectors.

Table 3: Overview of business building blocks of data spaces

Business Building block	Role and scope	Example
Service Level Agreement (SLA)	Provides specification of a service and the standards that it should meet	In a data space in the field of digital finance, a data provider offers credit risk assessments as a service. The service is offered according to an SLA specifying the amount and depth of information that each credit-scoring report should comprise.
Accounting Scheme	Specifies data-sharing parameters to be recorded and reports to be produced	In a data space in the agriculture sector, a set of parameters needs to be logged and tracked for each interaction between business actors, including e.g. the volume, the type, and the locality of the data provided by a data provider to a data consumer. The accounting scheme details the information logged, along with any relevant reports produced.

Business Building block	Role and scope	Example
Billing/Charging Scheme	Specifies rules that lead to the billing/charging of services provided over the data space	In the scope of a green-deal data space, earth observation experts provide value-added services to urban planners and insurers, such as information on climate change. These urban planners and insurers pay for access to this information. The building block specifies the billing/charging scheme, which can be e.g. per query, per report, or according to the volume of the information provided.
Smart contract	Provides a protocol for implementation of an agreement between two or more parties	Collaborative supply chain risk management is required to help companies exchange information and sensitive data. In the automotive industry, automation of risk reports based on sensitive data allows companies to react more quickly and efficiently. In this context, a smart contract allows the data owner to retain control over the data, which can be stored decentrally. Furthermore, the data owner can specify the terms and conditions of data access and how data consumers can use the data.

As already outlined for the technical building blocks, the list of business building blocks too is indicative, not exhaustive. Additional business actors can be defined as part of a scheme that is derived from specific domains, business requirements, or regulations.

2.4.4 Organisational/operational building blocks

From the perspective of a Data Provider and/or Data Owner, data control capabilities for ensuring data sovereignty, trust and security in a data space are necessary to prevent the misuse of data shared and exchanged. As such, they can be considered sine qua non conditions for organisations willing to share and exchange sensitive data and information.[59]

Data sovereignty is a natural person's or corporate entity's capability of being entirely self-determined with regard to its data. This means, data sovereignty allows a legal person to exclusively decide about the usage of its data as an economic asset. In practice, it allows organisations and individuals to stay in control over the terms and conditions under which their data is made available to others, and how it may be used and processed by them.

To ensure data sovereignty in data sharing and exchange, both organisational and operational agreements must be in place. While such agreements support and enable data usage policies, they also bring trust into the whole data ecosystem, as they function as a trust anchor connecting the physical and the digital world. Interoperability of the system as a whole relies on agreements that ensure interoperability between all participants. An appropriate interoperability scheme has to be continuously maintained and synchronised between all parties. In addition, governing bodies need to provide a frame for all business transactions in data spaces, including reliable maintenance of all underlying agreements.

Building blocks related to interoperability

To ensure interoperability between all data space participants, the technical measures required for facilitating interoperability need to be continuously maintained. This includes general agreements and domain specific models. The continuity model provides measures for change, release, and version management.

» **Domain Data Standard:** The Domain Data Standard represents the language for data sharing in a specific sector or domain. To achieve specific goals, multiple such standards can be used in combination.

Building blocks related to trust

In addition to the technical implementation of building blocks related to trust, operational and organisational measures create a trust anchor for the overall system. The main purpose of the trust anchor is to connect the physical and the digital world. A legal or natural entity requires a digital identity that enables reliable identification and authentication.

[59] The IDS reference architecture model, https://internationaldataspaces.org/use/reference-architecture/

» **Unique Identifiers:** Unique and trusted identifiers enable reliable identification of legal and natural entities (including things) across domain specific or country specific identification schemes. Such identification has to be extended with value-adding attributes (e.g. commercial register number or tax identification number). Such additional information must be provided by trusted parties.

» **Authorisation Registries:** To unambiguously identify each data space participant, special authentication registries must be in place. These registries need to be established in accordance with operational agreements (i.e. policies) concluded within the data space. These registries itself must be approved and monitored by a neutral body. Authentication of a participant requires a structured admission process including a compliance assessment to set up the trust anchor of each identity at the registry.

» **Trusted Parties:** On the basis of authenticated identities, trusted parties can verify and validate participants' capabilities. This includes two aspects: 1) acquisition or evaluation of capabilities in a structured process and 2) verification of these claims against a digital identity. While the first aspect is typically covered by certifications or registrations, the second aspect is often carried out by commercial services. A trusted party therefore provides digital evidence of specified and measurable criteria. The content of those criteria is specified by regulations or by (sector-)specific agreements.

Building blocks related to data space administration, organisation, and guidance.

The fundamentals of all business transactions are frameworks that provide agreements between all actors. All technical and functional agreements are part of this and must be agreed and monitored by a special body.

» **Data Space Boards:** Data Space Boards provide governance for data spaces in terms of decision-making, guidance, steering, and conflict resolution.

» **Overarching cooperation Agreements:** All data space participants need to agree on certain functional, technical, operational and legal aspects. While some agreements are reusable in a generic or sector-specific way (e. g. rule books), others are use-case specific.

» **Continuity Model:** The Continuity Model describes the processes for the management of changes, versions, and releases for standards and agreements. This also includes the governance body for decision-making and conflict resolution.

» **Regulations:** Regulations refer to laws or administrative rules, issued by an organisation, used to guide or prescribe the conduct of the members of that organisation or countries.

All operational and organisational building blocks rely on the existence of governing boards or sector specific bodies that provide common and accepted rules. Those rules need to be monitored by neutral and independent entities. General acceptance of these rules, together with measures for monitoring of the rules by neutral parties, is the foundation for ensuring trust (i.e. a trust anchor) for the overall system. Enforcement of agreements is typically done by authorities or independent evaluators. This separation of powers is a fundamental aspect for ensuring governance in data spaces and promoting the idea of a soft infrastructure.

Organisational/ operational Building block	Role and scope	Example
Unique Identifiers	Identification of legal entities, natural persons, or things in terms of a unique identifier and other information about an entity	Tax identification numbers, legal entity identifiers
Authorisation Registries	Verification and validation of digital identities and their mapping to real-world objects	eIDAS qualified seals provide a mechanism to verify and validate identities. The providing (national) registry implements the policy as defined by the eIDAS regulation.
Trusted Parties	Provide neutral evidence on specified facts based on predefined criteria	A trusted party is an independent and accredited evaluator of a certification scheme (e.g. ISO 27001).
Domain Data Standard	Provides the syntax and semantics for data exchange and data sharing on different levels	In manufacturing data spaces, a combination of different standards is used to describe the syntax and semantics of data transactions (e.g. ISO 10303, Asset Administration Shell, eCl@ss).

Sector specific Data Spaces

3

3 Sector-Specific data spaces

3.1 Introduction

The goal of this chapter is to introduce sector-specific needs and inform about the status of implementation of data spaces in selected industries. The vision of the EU to *become an attractive, secure and dynamic data economy* is materialised by an ecosystem of common European data spaces in strategic sectors and domains of public interest (Pillar IV). Of the nine sectors mentioned in the EC communication and by inspiration of the Digital Platforms focus area of the H2020 ICT Work Programme 2018–2020, the Open DEI project is addressing four through an ecosystem of 35 (and counting) H2020 actions: manufacturing, agri-food, healthcare, and energy.[60] These four sectors were selected because of their strong position in global competition, and because they are system-critical for a working economy and modern society.

For each selected sector, we 1) describe how the four design principles introduced in Chapter 0 can be applied; 2) identify high-value business scenarios, characterised by their opportunities and challenges as a materialisation of the fundamentals described in Chapter 1; and 3) introduce embryonic data spaces as reference implementations of multi-stakeholder business and governance agreements as described in Chapter 4. The European Commission has defined embryonic data spaces as an initial ecosystem that they develop from platforms to ecosystems. In this chapter, we make an initial approach to identify candidates to embryonic data spaces in the four sectors. Furthermore, we expect more efforts from the projects and initiatives to identify data spaces and define their maturity level.

This chapter describes best practices in selected domains, but also problems and obstacles impeding progress. We aim to identify promising approaches towards the deployment of realistic data spaces in terms of security, privacy, trust, business interest, and competition. Some sectors may be aiming at rather disruptive approaches, while others aim at continuous improvement.

Stakeholders can also be assessed with respect to their preparedness for data space innovation. In the manufacturing industry, for example, a so-called data space pathway[61] with five levels of maturity has been used in various dimensions of data spaces: technology (i.e. data platforms and data models), business (i.e. data-sharing agreements and revenue models) and organisation (i.e. data competencies and data business processes).

Successful implementations of sector-specific data spaces are not just to be based on design principles, but also need to be accompanied by convincing value propositions and the key concept of trust: in the validity of the data itself and the algorithms operating on it; in the

[60] The selection is based on the focus of the Open DEI project, which addresses four out of nine sectors through an ecosystem of more than 35 different Horizon 2020 actions funded by the EU.
[61] https://www.connectedfactories.eu/connectedfactories-information-sharing-and-analysis, Connected Factories CSA.

entities governing the data space; its enabling technologies; and in and amongst its wide variety of users (organisations and private individuals as data producers, data consumers or intermediaries). [62]

One important aspect is the infrastructure and platform to use and exchange data in these sectors (see Chapter 1). Such an infrastructure for industrial and/or personal data should be highly distributed, scalable, interoperable and compliant to open standards in order to avoid vendor lock-in. When looking at the different sectors, there is currently a rather wide spectrum of different infrastructure implementations with limited interoperability, especially across sectors.

In the process of creating and establishing a data space, another important aspect is the ability to involve and engage a significant number of data providers and data consumers sharing and exchanging data to achieve business benefits and derive key performance indicators. This is another motivation for conducting maturity assessments and ensure personalised provision of training services for managers (data culture), technicians (data technology), and users (data skills).

Furthermore, an appropriate data ecosystem governance mechanism is a fundamental condition that ensures mutual, reciprocal trust in B2B models, code of conduct fulfilment, confidentiality, as well as non-competition and collaboration agreements under the principles of data sovereignty and data access/usage control.

Data spaces need to find agreed and shared methods for data valuation and monetisation in the network, so that interoperability, trust, and governance principles can materialise in a fair and transparent marketplace. Data value needs to be expressed and modelled via open and standard metadata frameworks and shared among all network participants, both the provider side and the consumer side. By enriching them with value-oriented metadata, datasets become FAIR (Findable, Accessible, Interoperable and Reusable) for data consumers and ready to be shared and exchanged according to different business and revenue models. On the other side, full awareness of the value of data is a fundamental precondition for data providers to properly and conveniently protect and disclose confidential data and knowledge in a network.

[62] The new 2020 edition of the BDVA position paper "TOWARDS A EUROPEAN-GOVERNED DATA SHARING SPACE - Enabling data exchange and unlocking AI potential" proposes a new reference model for elaborating a value proposition for data spaces: the Data-Sharing Value Wheel.

3.2 Manufacturing

3.2.1 Data space scenarios, opportunities, and challenges

The Connected Factories H2020 CSA has recently elaborated a set of three 2025 pathways to indicate the future directions of R&I in the field of data-driven ICT applications for manufacturing:

» Towards networked enterprises in complex, dynamic supply chains and value networks – the Hyper-Connected Factories pathway: Implementing data spaces in this Grand Scenario represents a big opportunity for implementing new business models such as "manufacturing as a service" or "one-of-a-kind production" through e.g. additive manufacturing. Main obstacles and challenges are the great variety and veracity of data and the need to harmonise different, heterogeneous and often cross-national policies and regulations.

» Towards optimised and sustainable manufacturing including advanced human-in-the-loop workspaces – the Autonomous Smart Factories pathway: Implementing data spaces in this Grand Scenario will support high levels of automation (Level-4) and self-capabilities of the production lines with a new role for human operators (Industry 5.0, the Connected Worker). Advanced coordination of smart factory data spaces will support the development of digital twins interacting with digital personas in the engineering of zero-defect and zero-impact manufacturing plants (Green Deal). Main challenges are large data volumes, data velocity aspects, and the need to build highly distributed data spaces along the computational continuum between embedded, edge and cloud infrastructures.

» Towards data-driven product-service engineering in knowledge intensive factories – the Collaborative Product-Service Factories pathway: Implementing data spaces in this Grand Scenario means to be able to follow the product lifecycle and to provide a consistent ecosystem of accompanying services around it. In complex products (e.g. aeronautics, shipbuilding, automotive, machine tools), data spaces will often encompass several years (if not decades) of time with a very diverse pool of technologies involved, several professional figures and human skills exercised, and several administrative and security domains to be integrated. Typical services to be provided are in the fields of fleet management and optimisation, monitoring and diagnosis, or predictive and prescriptive maintenance. In the consumer goods industry (e.g. food, beverage, fashion, jewellery, leisure, retail), data spaces are able to digitally integrate the pedigree of goods, providing consumers with advanced services such as certification of origin, environmental impact assessment, or ethical principles respectfulness. Circularity and twin transition (digital/green) are often common challenges in this Grand Scenario, where data spaces often address all four dimensions of big data (i.e. volume, velocity, variety, and veracity).

Implementation of data spaces in the manufacturing sector has to cope with new requirements coming from new trends in manufacturing, such as:

» from mass production to mass customisation

» reliable and resilient supply chains

» servitisation business models

» circular manufacturing

All these objectives rely on data sharing and exchange and cannot be accomplished without the manufacturing sector being fundamentally digitalised. Several efforts have been made to define the digitalisation pathway for different types of factories. On a lower level, there is data capture, while on an upper level are data analysis and decision support systems. RAMI 4.0 (Reference Architecture Model for Industrie 4.0) has analysed data transmission both horizontally and vertically.

The situation in the manufacturing sector currently is as follows: on the one hand, information and data remain in silos in the different factory departments, leading to strong vendor lock-in; on the other hand, strong confidentiality issues prevent extensive diffusion of open data generated by critical manufacturing processes.

Today's manufacturing companies are analysing the benefits of data spaces and the data economy. The data economy allows establishing new business models that are challenging today's way of doing business. Manufacturing companies therefore need to explore new approaches. Also to be taken into account: the more manufacturing companies go digital, the more they are exposed to cyber-criminality; this is a high-priority concern for this domain.

3.2.2 Instantiation of data space design principles

Decentralised soft infrastructure: A decentralised soft infrastructure requires that all data space participants comply with a set of functional, technical, operational and legal agreements. From a technical standpoint, a soft infrastructure can be seen as a collection of interoperable technologies and standards, ensuring management of security, identity, authentication, protocols, metadata etc. Interoperability among actors, sensors, and heterogeneous systems is a crucial factor when it comes to turning the Industry 4.0 vision into reality.[63] In the manufacturing sector, an ever-increasing number of technologies and concepts support interoperability and digitalisation, such as RAMI 4.0 and Asset Administration Shell (AAS) of Plattform Industrie 4.0, or OPC-UA enabling interoperability on the communication and information model level. Relevant work is currently ongoing in the domain of data standards; especially in the Platform Standardization Council Industrie 4.0 [SCI 4.0] and in the CEN-

[63] The Industry 4.0 Standards Landscape from a Semantic Integration Perspective

CENELEC-ETSI Coordination Group on Smart Manufacturing [SMa-CG]) and metadata / ontology in the RAMI 4.0 and the Industrial Ontologies Foundry.

Data sovereignty: To ensure data sovereignty principles for all applications and across the entire value chain, reference architectures and industrial data platforms must be established. DIN SPEC 27070 ("Requirements and reference architecture of a security gateway for the exchange of industry data and services"[64]) is a standard that specifies requirements to be met by a security gateway for data exchange with regard to the gateway's architecture and cybersecurity measures required for establishing virtual cross-company data spaces.

A level playing field for data sharing and exchange: In the manufacturing sector is required to facilitate collaboration between OEMs, logistics service providers (tier-1, tier-2, …), IT providers etc. This will lead to the establishment of data ecosystems facilitating collaborative business scenarios (e.g. circular economy, collaborative configuration management, distributed manufacturing networks/marketplaces, collaborative condition monitoring [CCM] and so on). If data can be shared seamlessly and in a controlled manner, vendor lock-in will be prevented. At the same time, easier access to data will provide more opportunities for SMEs (such as AI-algorithm providers), while enhanced data sovereignty mechanisms will allow data providers to stay in control over their (sensitive) data.

Public-private governance: Means to establish win-win business and governance models for data spaces in the manufacturing sector. Multi-stakeholder business models, methods and tools (such as the BM Navigator) need to be applied to all stakeholders identified in the ecosystem, and their legitimate expectations in terms of business KPIs need to be harmonised and integrated. A typical example is in predictive maintenance of complex machineries working on shopfloors: The product manufacturer, the machine tool manufacturer, the platform provider and the developer of advanced (often AI-based) applications need to come to business agreements ensuring their IPRs are respected and business expectations are met. Governance rules are also necessary to preserve ownership and confidentiality of data through multi-lateral contracts and agreements. Similar processes need to be put in place when extending the small-scale ecosystem to larger scales (i.e. when multiple representatives of the same role [service providers, equipment providers etc.] and even competitors need to be coordinated towards common aims and objectives).

[64] DIN SPEC 27070:2020-03
https://www.din.de/en/wdc-beuth:din21:319111044, German version]. English version facilitated by IDSA.

3.2.3 Embryonic data spaces

The first call of the Digital Europe Programme[65] will address two typologies of data spaces for the manufacturing sector: "The main objective is to build and deploy two operational data spaces for specific value chains in the manufacturing sector, which enable companies in different user roles (supplier, client, service provider, …) to interact with large amounts of industrial data across their organisational borders. The first data space will address agile supply chain management and execution, and the second one dynamic asset management and predictive/prescriptive maintenance". There are already examples of successful implementations of data spaces in the manufacturing sector related to both typologies.

In the supply chain management subdomain (MARKET 4.0 H2020 project), the Smart Connected Supplier Network (SCSN)[66] is an initiative of manufacturing companies and their IT suppliers in the high-tech manufacturing supply chain. The goal is to facilitate cross-factory communication and thus ensuring supply chain transparency and interoperability. Manufacturing companies can use a single SCSN connection to exchange purchase-to-pay information with all their suppliers and customers. This reduces both administrative effort and human errors, while increasing supply chain agility.

In the subdomain of dynamic asset management and predictive/prescriptive maintenance, the QU4LITY project demonstrates data-driven ZDM (zero-defect manufacturing) solutions and related services in a combination of five strategic ZDM plug & control lighthouse equipment pilots as well as nine production lighthouse facility pilots. QU4LITY has developed a set of sovereign data spaces supporting the creation of multiple digital infrastructures for ZDM, enhancing the typical manufacturing scenario with a bouquet of data-driven value-added services provided on top of these infrastructures. In such a scenario, Georg Fischer AG, a manufacturing SME, is currently about to implement "a digital machine and part twins for zero-defect manufacturing" data space. Current barriers to high accuracy in manufacturing in multi-technology and automated cells are related to limitations in data aggregation to either machining processes or machine health scopes. Zero-defect manufacturing in these systems will therefore be possible by taking into account already during the planning stage how machine mechanics evolve towards states where deviations are more likely to occur, where failures might damage the machine, or where uncertainties are introduced by maintenance, repair, or any other uncontrolled factor in the value chain. This requires new approaches to production, promoting better and innovative defect management and production control methods that are consistent with the integration of ZDM processes (namely in-line inspection technologies) and integration of tools for autonomous/automatic smart-system decision-making across the entire supply/value chain.

[65] https://digital-strategy.ec.europa.eu/en/activities/digital-programme
[66] https://smart-connected.nl/

3.3 Agri-food

The Farm-to-Fork Strategy[67] is at the heart of the European Green Deal aiming to make agri-food systems fair, healthy and environmentally-friendly. The agri-food sector is also crucial for the European economy and overall resilience of the EU member states. The sector involves long and complex supply chains. About 70 percent of agricultural production is manufactured by the food and beverage industry, with products arriving at supermarkets after being highly processed and transformed. The number of farmers is high, while profit margins are low. Farming is a volatile fixed-location business that is dependent on local conditions (e.g. the weather), and labour force used in farms is highly seasonal and mobile. Although the food and beverage industry is dominated by multinational companies, it is a highly diversified sector with more than 290,000 SMEs generating almost 50 percent of turnover and providing two thirds of employment in the sector.[68]

3.3.1 Data space scenarios, opportunities, and challenges

The agri-food sector may have a reputation of being slow in adopting digital technologies, but in reality, it is rapidly moving towards gradual digital transformation. When we refer to agri-food, we need to refer to a broad ecosystem, which starts upstream along the supply chain (and here we refer to agriculture) and arrives up to the shelfs and the consumer along a complex data-sharing process that includes players in a number of diversified areas (e.g. food processing, distribution platforms, wholesalers, retailers). Agriculture is facing an era of digitally enhanced farming, where data is generated during various stages of agricultural production and related operations, and where IoT solutions, robotics, and sophisticated decision-support systems are being deployed. The same applies to the food and beverage industry. Larger companies are progressing more rapidly than SMEs, but investments are growing and overall awareness of the benefits of digital transformation is increasing.

Data-driven approaches open up unprecedented opportunities to 1) improve resource efficiency, productivity, and ecological sustainability; 2) dynamically adapt business plans to changing markets and consumer expectations; 3) decrease administrative costs and enable science-based policies; 4) provide more prosperous living conditions for rural communities; and 5) improve the relationship between the consumer and the different actors across the value chain.

Today, if we refer to the agri-food sector, we need to speak about the "food system". The fact that more and more data is being shared and exchanged poses a major challenge for the EU agri-food sector. The nature of agricultural data is very specific and diverse (e.g. livestock and fishstock data, land and agronomic data, climate data, machine data, financial data, compliance data etc.). If we look at the overall supply chain, the situation becomes even more complicated,

[67] EU Farm to Fork Strategy, https://ec.europa.eu/food/farm2fork_en
[68] FoodDrinkEurope Data & Trends 2020

since food origin, quality, and safety information, including nutritional information, mix with aspects related to manufacturing, energy, logistics, and waste management. To tap into all of the potential benefits, data sharing between different stakeholders must be conducted under fair and transparent rules.

There are important questions about privacy, data protection, intellectual property, data usage control, interoperability, and trust to be resolved. As farmers and other small businesses in the sector are often family businesses (99 percent of Europe's food and beverage companies are SMEs[69]), it is important to apply human-centric data principles to privacy and data usage control, as data shared can be personal, sensitive and/or confidential. Agri-food companies should feel confident that they are actually benefiting fairly from data sharing and exchange.

The following priorities need to be faced when speaking about data in the agri-food sector: 1) the agri-food sector needs human skills training and development of technical capabilities to make full use of opportunities provided by data spaces (i.e. data is of no use if it cannot be interpreted correctly and transformed into meaningful business decisions; to do so, state-of-the-art data analytics tools are needed to reap the benefits of vast amounts of collected data); 2) interoperability of different systems and devices is still in its infancy, limiting the success of digital transformation; 3) significant differences can be identified regarding digital readiness of EU member states, which make requirements for skills training and technical interoperability at EU level an even more pertinent issue when it comes to designing European data spaces; 4) data quality is fundamentally important for the sector (and that the right data is available at the right moment); and 5) the business models behind the implementation of European agri-food data spaces require synergies and still need to be clarified.

As we follow the EU's Farm-to-Fork rationale[70], it is important to consider the connections of the agri-food sector with other domains, especially manufacturing (Industry 4.0), logistics, energy, and health. Cross-domain data-sharing requirements naturally arise from such interconnection and overlap. For example, to identify the accurate carbon footprint of an agricultural product, a large amount of data is needed about the specific conditions of the farming process in question, the practices of the food producer, the logistical operations conducted, and finally the practices of the retailer.

[69] FoodDrinkEurope Data & Trends 2020
[70] EU Farm to Fork Strategy, https://ec.europa.eu/food/farm2fork_en

3.3.2 Instantiation of data space design principles

Decentralised soft infrastructure: The code of conduct for data sharing in the agricultural sector specifies the conditions for defining the soft infrastructure based on contractual agreements and guidance on fair and transparent use of data. This code of conduct is the first attempt to define a framework for data sharing within an industry sector.[71] Interoperability in digital agriculture will become an essential requirement, as the system landscape is getting more and more heterogeneous with an increasing number of different machines and entities that must exchange information.[72] To reach a truly European solution, an agreement on a set of data and system interoperability mechanisms and standards is needed. This will help avoid being locked into existing platform architectures. One key issue is whether the suppliers of farm management systems (FMS) are ready to allow for interoperability and federation of their data platforms with other systems. Furthermore, interoperability is important with regard to high-value datasets to be collected and made available as open data, for example related to R&I initiatives or policy monitoring needs.

Data sovereignty: To achieve data sovereignty, functional reference architectures need to be in place to support controlled data sharing and exchange between parties across all applications and the entire agri-food value chain. A joint approach to build a trusted infrastructure for a common agri-food data space should start from a technical basis broadly agreed upon across sectors beyond the agri-food industry.

Level playing field for data sharing and exchange: Ensuring a level playing field for agri-food data spaces implies that new entrants will not face insurmountable barriers when seeking admission to a data space. A level playing field will empower users, farmers and other small businesses (most agri-food companies are SMEs) likewise. Agri-food companies should feel confident that they can really benefit fairly from data sharing and exchange, avoiding monopolistic situations and abusive power of large companies and big platforms. As data shared by farms and other small businesses is often sensitive or confidential data, it will be important to apply human-centric principles to privacy and data usage control.

Public-private governance: Should be based on generally agreed principles for sharing and exchanging agricultural data within the agri-food value chain. Multiple stakeholders have been active in defining transparent and trusted practices for the sector. The EU Code of Conduct on agricultural data sharing[73] can be seen as a best practice to be followed by other domains wishing to create thriving and balanced data spaces. It constitutes a joint effort from inside the sector to shed more light on contractual relations and provide guidance on the use of agricultural data.

[71] EU Code of conduct on agricultural data sharing by contractual agreement, https://copa-cogeca.eu/Publications
[72] Excerpt from the presentation by Stefan Rilling. Fraunhofer IAIS. ATLAS project coordinator. https://aioti.eu/wp-content/uploads/2020/07/Report_Data_Sharing_in_Agriculture_Online_Webinar_10.06.2020_Final.pdf
[73] EU Code of conduct on agricultural data sharing by contractual agreement, https://copa-cogeca.eu/Publications

3.3.3 Embryonic data spaces

There are several examples of national or regional alliances working on data sharing and exchange in the agri-food sector, including API-AGRO in France,[74] DjustConnect in Belgium,[75] JoinData in the Netherlands,[76] and DKE-agrirouter in Germany[77].

Being a concrete manifestation of the EU Code of Conduct for the agricultural sector, DjustConnect is a data sharing and reuse platform originating from the Flemish region of Belgium, aiming to provide a trusted data and services ecosystem. DjustConnect is collaborating with other European platforms with similar objectives to establish a European agri-food data space. Participants can make use of DjustConnect to publish various data resources, including IoT data, and to authorize access and provide consent on data transactions. DjustConnect provides a dashboard to farmers providing practical means for data usage control according to data sovereignty principles. The goal is to guarantee that farmers are in full control of their data and know about all transactions done with it. Based on early experiences made with the platform, FMS suppliers can benefit in several ways: 1) reaching new markets and scaling-up business; 2) sharing infrastructure investment costs; and 3) getting access to consumers and engaging with third parties.

To make sure a common European agri-food data space will effectively arise from these initiatives, proactive thinking and comprehensive planning is necessary.[78] Our recommendation is to closely monitor embryonic initiatives in the agri-food sector and learn and benefit from their practices and solutions.

[74] https://api-agro.eu/en/
[75] https://www.djustconnect.be/nl
[76] https://join-data.nl/en/
[77] https://my-agrirouter.com/en/

3.4 Healthcare

The Covid-19 crisis has significantly raised the urgency for efficient use of health data across borders. It has also highlighted the importance of joint European health initiatives and data-sharing scenarios. Data can improve patient outcomes, foster research, and accelerate the development of new health services only if it is shared between stakeholders and reused by them. In this process, privacy must be respected, data usage control enforced and transparency ensured. Trust in the use of health data between member states requires transparency, functional frameworks, and joint regulation.

The European Data Strategy describes the current state as follows: "Health is an area where the EU can benefit from the data revolution, increasing the quality of healthcare, while decreasing costs. Progress will often depend on the willingness of Member States and healthcare providers to join forces and find ways to use and combine data, in a manner compliant with the GDPR, under which health data merit specific protection. While the GDPR has created a level playing field for the use of health personal data, fragmentation remains within and between Member States and the governance models for accessing data are diverse".

Establishing a European health data space is an integral part of building a European Health Union, a process launched by the Commission in November 2020. Investments to support the European health data space will be made under the EU4 Health Programme, as well as under the Horizon Europe and the Digital Europe Programme.

3.4.1 Data space scenarios, opportunities, and challenges

Digital transformation in the healthcare sector has evolved relatively slowly over the last thirty years, mostly due to the complexity of healthcare systems, the nature of the relationship of patients and healthcare professionals, and legal and ethical issues such as privacy and sensitivity of healthcare data. Furthermore, the healthcare domain is focused mainly on services and processes rather than products. On the other hand, there continues to be enormous opportunities regarding 1) increasing quality and patient outcomes of healthcare; 2) reducing healthcare cost; and 3) improving patient and caregiver experience by leveraging the potential of digital services.

The EU initiative eHealth[79] defines three major pillars of digital healthcare in Europe: 1) secure data access and sharing; 2) connecting and sharing health data for research, faster diagnosis, and improved health; and 3) strengthening citizen empowerment and individual care through digital services. There is a growing need to tap into the huge potential of health data to support medical research with the aim of improving prevention, diagnosis, treatments, drugs, and medical devices. Data resources such as EHRs already provide abundant opportunities within individual care institutes. However, as patients generate an ever-increasing stream of data by themselves, data should be better integrated and labelled with proper quality assurance for large scale AI applications making use of the data.[80] In comparison with other sectors, the very context of production of health data is extremely important and needs to be captured aside from the data themselves in order to make those data truly reusable. Furthermore, most health data today are still available in a free-text format, which prevents it from being directly machine processable. To make data created in a clinical context reusable for other purposes, adapted use-friendly natural-language processing tools together with healthcare providers' incentives packages are needed.

The concept of health data should be considered broadly to cover the whole lifespan of an individual. This means that self-monitoring data from wearables and non-medical sources should be increasingly considered in addition to clinical data. Digital services can empower citizens, making it easier for them to take a greater role in the management of their own health from following prevention guidelines and being motivated to lead a healthier life, manage chronic conditions and provide feedback to healthcare providers.

In a wider perspective, any data affecting humans (e.g. food-related retail data, environmental data, or work-related information) can be regarded as cross-sectorial health data. Innovators would benefit from more cross-sectoral sharing of health data. This data is also increasingly available and used by the individuals themselves for health awareness and self-management, but also by policy makers involved in decision-making for population health, care process improvement, or insurance policy development.

The European health data space should benefit a broad range of stakeholders – from the individual citizen to public authorities and researchers to companies developing and/or using cutting-edge technologies. It should promote the development of healthcare systems as well as availability, efficiency, and sustainability of services. Research and innovation will also benefit from easier access to data. The target state should be a human-driven healthcare system that uses data extensively. This requires to set shared targets, define rules and create a clear and shared plan guiding all actions. All this should be based on data sovereignty principles ensuring privacy and transparency as key elements of trust.[81]

[79] https://digital-strategy.ec.europa.eu/en/library/staff-working-document-enabling-digital-transformation-health-and-care-digital-single-market
[80] AI in Healthcare, BDVA White Paper, November 2020 https://www.bdva.eu/sites/default/files/AI%20in%20Healthcare%20Whitepaper_November%202020_0.pdf
[81] Towards trustworthy health data ecosystems, Sitra, https://media.sitra.fi/2020/10/08101601/towards-trustworthy-health-data-ecosystems-2.pdf

3.4.2 Instantiation of data space design principles

Decentralised soft infrastructure: The SITRA fair data economy rulebook is a useful toolkit for creating a decentralised soft infrastructure based on common agreed rules. It provides templates and a checklist for business, legal, technology, data, and ethical aspects. The Portuguese National Health Service (SPMS) has successfully implemented the SITRA rulebook, providing to the Portuguese healthcare ecosystem a contractual framework, general terms and conditions for data sharing, a governance model and a guide how to use it.[82] Again, one of the main goals when setting up sector-specific data spaces is interoperability. What is needed is a broad data quality framework that encompasses semantic interoperability and FAIR principles. In the research domain, several major initiatives, mainly supported by the IMI framework[83], have tested the concept of a federated architecture, where data remains stored where it is produced and only the result of a request is made available.

Data sovereignty: To enable European health data spaces, the infrastructure architecture and the guidelines for technical interoperability must allow for establishing appropriate trust mechanisms (e.g. methods for anonymisation and pseudonymisation). Novel privacy-preserving technologies (e.g. differential privacy) can be used for health data to create synthetic datasets that sufficiently resemble source data while at the same time avoiding privacy issues. Data providers should also be encouraged to make available synthetic health datasets that can be used for the purposes of research and innovation.

Level playing field for data sharing and exchange: A level playing field will be created by granting access to data to create new services for the benefit of everybody. This will strengthen the innovation capacity of European companies as well as of research and development organisations. Clarifying fair practices and providing necessary regulation for the secondary use of health data for research and innovation purposes will be of paramount importance. A new model of a person-centred European data economy will improve the health and well-being of individuals and societies, strengthen health systems, and foster economic growth by enhancing co-operation for data processing, leveraging data for well-being, delivering new services, and making Europe a leader in data-based health and well-being innovation.[84]

[82] Europe rules – making the fair data economy flourish https://www.sitra.fi/en/events/europe-rules-making-the-fair-data-economy-flourish/#programme

[83] See for example: https://www.ehden.eu/

[84] https://media.sitra.fi/2020/10/08101601/towards-trustworthy-health-data-ecosystems-2.pdf

Public-private governance for European healthcare data spaces should complement the horizontal European data spaces framework, including functions and responsibilities of the relevant actors in the healthcare sector. Furthermore, it is vitally important to ensure data sovereignty of citizens regarding their health data. In parallel to initiatives driven by EU member states, European citizens/patients can contribute directly to the creation of European data spaces.

MyData[85] is an initiative that defines sector-independent data sovereignty principles for the benefit of individuals. The goal is to enable individuals to make informed decisions regarding the sharing and exchange of personal data, and to make each individual the central connection point for mobility of data between sectors. This self-determination can be achieved not only by legal protection, but also by proactive actions to share the power of data with individuals. Overall, the aim is to maximise the collective benefits of personal data by fairly sharing them between organisations and individuals.

3.4.3 Embryonic data spaces

A number of EU member states are currently developing national infrastructures for health data reuse. Nevertheless, the road to truly open (as opposed to ad-hoc member states supported) European data spaces will require some time considering the sensitivity of the issue and the number of stakeholders to be involved. Furthermore, important data flows such as G2B are still not common practice and will need some time to be adopted.

eHealth Digital Service Infrastructure (eHDSI) is an example of a major initiative to establish a European Electronic Health Record (EHR) exchange format that is accessible by all EU citizens. Piloted since 2012[86] and launched in 2019, this infrastructure allows electronic prescriptions and patient summaries to be exchanged between healthcare providers across national borders. The goal of eHDSI is threefold: 1) enable data exchange between 22 member states by 2022; 2) expand to medical images, laboratory results, and discharge reports; and 3) enhance the virtual consultation model and registries of European Reference Networks (especially regarding rare diseases as priority use cases).

Towards European Health Data Spaces (TEHDAS) is a Joint Action including 26 European nations. It plays an important role on the way towards a truly European approach under the third EU Health Programme[87], with the overall goal of helping EU member states and the EU Commission in developing and promoting concepts for sharing of data for citizens' health, public health, and health research & innovation in Europe. Furthermore, the Joint Action aims

[85] MyData Global http://mydata.org
[86] Towards trustworthy health data ecosystems, Sitra, https://media.sitra.fi/2020/10/08101601/towards-trustworthy-health-data-ecosystems-2.pdf
[87] TEHDAS Joint Action for the European Health Data Space: https://projectsites.vtt.fi/sites/premed/files/workshop2020/Premed_workshop_Kalliola_Sitra.pdf

at establishing an operational framework and governance model for the exchange and secondary use of health (care) data between (European) countries, respecting the principles of transparency, trust, fairness, and citizen empowerment. It also aims to provide solutions for improving quality, interoperability, and fairness, and to create a catalogue of services to support the secondary use of health data across European health data spaces. In general, the Joint Action provides a European perspective towards improving citizens' capacity to engage with data and citizens' trust in data sharing.

3.5 Energy

Energy systems are basic pillars of the welfare society. Oil, gas, thermal and electricity networks are considered critical infrastructures. In this field, there are utility actors in charge of maintaining, modernising and guaranteeing sustainability, and quality of supply in a competitive way. For example, electricity grids and the information systems used to operate them are continuously improved, optimised, and enhanced in order to achieve sustainability and network evolution. Furthermore, we are seeing the transformation towards renewable energy sources and distributed energy production – a trend providing tremendous opportunity to rethink data sharing and digital services.

3.5.1 Data spaces scenarios, opportunities, and challenges

The European Data Strategy delineates a common European energy data space to promote improved availability and cross-sector sharing of data in a customer-centric, secure and trustworthy manner. This will facilitate innovative solutions and support decarbonisation of the energy system.

Current systems are evolving through centralised datahubs, which are now emerging to collect data from smart meters, electricity retailers, and grid operators. The draft amendment to the Electricity Market Act (which took effect on February 1, 2019) obliges electricity retailers and distribution system operators to use datahub services for data management in their business processes related to information exchange and electricity trading on retail markets. Among the relevant stakeholders are energy service companies, demand side management providers and aggregators, energy communities, and equipment manufacturers. In general, smart grids start to be increasingly data centric. Utilities, renewables plant owners, and retailers rely on vast data lakes. Knowledge extraction from data is possible with the help of data analytics and artificial intelligence.

Utilities, electric and gas networks, and infrastructure owners and managers need to modernise infrastructures via open innovation ecosystems, for which digitalisation and data sharing are

key elements. Data-centric storage paradigms are evolving to federated data management. At the same time, new data governance tools are developed to support the global energy transition that will transform the energy consumption in key sectors such as heating, mobility, manufacturing, or transportation.

The most important challenge of the energy sector in the next years will be to achieve the European Green Deal objectives. The energy sector will have to address the transformation of the energy system into one which is not only carbon-neutral but also more cost-effective, energy-efficient and secure. Facilitating secure data sharing in the ecosystem and across sectors will bring more innovation and greater benefits in the energy domain to achieve societal and business goals.

The Renewable Energy Directive (2009/28/EC) establishes an overall policy for the production and promotion of energy from renewable sources in the EU. The increasing thrust towards renewable energy sources, the need of managing renewable primary resources, and bidirectional energy flows, along with the need to provide information access to end users, makes digitalisation and data sharing necessary and an important tool for future data spaces. The transformation towards renewables and distributed energy production creates new opportunities for data markets and digital services.

Cross-sectoral data spaces will add a new dimension to these highly distributed systems altogether. Data-driven business models will be a moving target when more and more data is available. Data analytics and AI models will support business goals and business models towards exploiting P2X (i.e. power transformation to various forms of energy). Examples of opportunities for renewables-based energy systems are peer-to-peer markets, AI consumption forecasts, new market models, data operator business, and AI automated decision-making, all of which require data sharing in order to achieve scalability and interoperability.

In a data-driven economy, data brings insight into the operational behaviour of organisations, societies, and machines. Data from energy infrastructures are essential to better understand the consumption patterns of energy users, acquire knowledge about the bottlenecks of the transmission and distribution networks, target reinforcement needs, and enable operation of networks in real time.

3.5.2 Instantiation of data spaces design principles

Decentralised soft infrastructure: The soft infrastructure will facilitate the sharing and exchange of energy related data between organisations based on a framework of agreements. This framework will be defined on four levels: legal, organisational, semantic and technical interoperability (based on an integrated governance approach). Interoperability of energy data systems has been very much a matter of defining application programming interfaces (APIs) for energy systems, providing services such as order submission, reporting, clearing of trade, or market data. The main challenge towards interoperable data spaces is to agree on a single approach that is accepted by all participants. Thus, interoperability requires use of widely accepted standards and solutions.

Data sovereignty: Trust, being a major component of data sovereignty, should be a target for data exchange in a scalable manner to facilitate the development of cross-sectoral data space opportunities. Trust manifests itself on many levels of systems and applications to be used for data spaces in the energy domain and across other sectors (e.g. components in the system architecture and their communication must be trusted, partners must be trusted, and the data must be valid). In cross-sectoral environments, data harmonisation means that different computer systems can exchange data so that the data makes sense between systems. While the IDSA reference architecture provides an approach for trusted data exchange, actors in the energy ecosystem, using existing standard semantic frameworks, can provide data harmonisation solutions.

Level playing field for data sharing and exchange: A level playing field for data sharing and exchange will allow players in the energy sector to cooperate on the design and maintenance of the soft infrastructure underlying data spaces. The data sharing in the level playing fild will allow them to compete on providing better equipment operation and value-added services to end-users. This will involve a large variety of actors generating a huge amount of data to increase energy efficiency and optimised energy asset management. In energy data spaces, organisations can capitalise on the data from energy infrastructures in order to develop new services and business models.

Public-private governance in the energy domain concerns both personal data and non-personal (i.e. industrial) data. Currently, the most well-known example are smart meters, which are used to share energy consumption information. In some cases, smart-meter data is made available to other actors than energy companies (e.g. loi Numerique in France proposed by Axelle Lemaire), and when combined with personal information, the data falls under GDPR regulations and can be considered personal data. While regulation is catching up, also implementation of cross-sectoral data spaces of energy and other domains is possible. Consent management for data flows, identity management, and access management are means that are available for information system developers. Commercial readiness in these areas is increasing. Standardized data-sharing components and data access via federation services provide means for governance of data spaces.

3.5.3 Embryonic data spaces

Wind aerogenerators

Only windfarm developers and wind turbine OEMs have access to the data collected from the wind energy turbines in operation. Therefore, they are the only players that are presently extracting value out of data at the top of the value chain. Most European component suppliers and ICT companies do not have access to data produced by the different systems in wind turbines in real-life operation, which means that they are missing the opportunity to extract value out of data and improve their competitiveness through digitalisation of products and services.

The value proposition here would be to design and develop an offshore-wind digital platform based on the IDSA reference architecture and components as core technologies, and to enable data sharing between data owners and data users. The business model would be based on data monetisation by providing data and associated services through the platform. Also windfarm plant owners and OEMs would indirectly benefit from the data-sharing activities because their technology providers will design and develop better products based on data and information about the operational performance and status during the lifecycle.

Data owners in this scenario are windfarm owners/developers, who own and operate wind farms, and wind turbine OEMs. In most of cases, the wind turbine OEM signs contracts with their customers (i.e. the wind farm owners) with exclusive rights to collect, use and manage the data obtained from the wind turbines for different purposes (energy efficiency, O&M, life extension etc.). Data users are companies across the value chain (i.e. tier-2 and tier-3 suppliers of components for the wind turbines, who have access to data), engineering firms designing the equipment and using data to improve the design and lifetime of components, TIC companies demanding access to data in order to offer developments and services to companies in all segments of the value chain (using big data, data analytics, artificial intelligence etc.). Furthermore, technology centres and universities are potential data users supporting the manufacturers in improving the design of the equipment they supply.

Within the "Interregional Innovation Projects" framework, DG REGIO of the European Commission funded the "Sensing & Remote Monitoring" (S&RM) pilot action, which was developed by the Marine Renewable Energy (MRE) interregional partnership during 2018 and 2019. The partnership was led by the Basque government (through the Economic Development Agency SPRI) and the initiative was coordinated by the Basque Energy Cluster. As a main result of this interregional cooperation, a sustainable business case was defined for the development and exploitation of a digital platform, as a marketplace where data generators (windfarm owners and OEMs) share relevant data from wind turbines to make them accessible for data users along the wind energy value chain (component suppliers, ICT companies, researchers etc.). The business case stated that "data privacy and security will be guaranteed through data usage control and data provenance, implementing technological solutions to cope with data sovereignty challenges based on the International Data Spaces (IDS) Reference Architecture Model".

Mobility and energy – a cross-sectoral data-sharing space

Data sharing between sector-specific systems can provide a totally new perspective for smart-city concepts. Especially city planning is experiencing a transformation of tools and instruments by leveraging information systems for data sharing. One issue in city planning is the growing number of electric vehicles that require charging. To facilitate this, smart grids must adapt to continuous changes on grid load balancing. Data sharing from energy grid to city planners gives constant updates and enables simulations of future energy use. Subsequently, simulations aid in route planning to optimise traffic in altering situations. Energy use of public transportation can be optimised as well using information coming from the vehicles.

Trust is a necessary feature of such a data-sharing environment, since the members of the community are business entities. As this example is based on pursuing a common good, agreements must be in place for entities describing the terms of data use. In principle, energy charging poles are only generating the volumes of used energy, not vehicle data. When route optimization is done, vehicle location data is necessary, but vehicle identification or information about who is driving is not needed. Interoperability between mobility data from vehicles, and smart grid data to city planning systems is a requirement. Therefore, harmonisation of data is needed when data exchange is done.

If the city planning information system is designed and implemented using generic interoperable architecture design, it will enable scalability. This in turn allows extensions with more data sources and theoretically endless exploitation of new data sets. Eventually, the system can become dynamic, where datasets are treated as components having parameters such as rules, agreements, levels of certification, and authorisation. Harmonisation will be automatically due to metadata descriptions and artificial intelligence coupling the components together.

Data Space Governance and Business Models

4

4 Data space Governance and Business Models

4.1 Introduction

Today's lack of a harmonised approach to establishing data spaces is more of a coordination and scaling problem than a technology problem. To set up data spaces that give users control over their data and interoperate with each other across sectors, adequate technology exists alongside with process knowledge to leverage it. What is required now is coordinated engineering and continuous maintenance, driven by sound European governance.

A data space is the total set interoperable data-sharing applications by actors in a specific sector or domain, either by their own development or through a certified software vendor, data broker, or marketplace. It is adamant that from the onset the aim is that data spaces over time, will systematically harmonise parts of their technical, operational, functional and legal aspects, leading to the emergence of a uniform, de-facto 'soft infrastructure' ensuring cross-sectoral data space interoperability. This harmonisation of common aspects in every data space into a soft infrastructure will enable users (citizens, businesses, governments) to stay in control of their data even across different sectors and applications (i.e. across different data spaces). This can be compared to the evolution of electronic payments in Europe (another special form of data sharing, unified by SEPA), which can be regarded as a soft infrastructure as well. It is a combination of rules and design decisions on top of an existing physical infrastructure of cables, services, and software stacks.

This soft infrastructure can only be achieved with good coordination – and good coordination comes with good governance. Good governance is about balancing the interest, input, and energy of private and public actors in order to ensure innovation and continuity in the long run. In this light, we must see the recently published Data Governance Act (DGA)[88] as the enabling governance framework for European data spaces to be established.

The draft DGA is proposing a two-tier governance structure: a governance entity required for each data space and an overall governance organisation concerned with all common aspects of data space interoperability and data sovereignty, thereby creating the de-facto 'soft infrastructure'. These aspects will be common for all data spaces, including (but not limited) to identity, metadata, consent, legal terms, and security. Such a soft infrastructure will also define the roles required for a data space to work (data provider, data consumer, etc.) as well as the rules all actors have to abide by.

In the DGA, a data intermediary, is the general term for a party (e.g. a broker, marketplace operator, or facilitator) that organises the sharing and exchange of data between all actors (both organisations and individuals).

[88] European Commission (November 2020) *Proposal for a regulation of the European parliament and of the council on European data governance (Data Governance Act).* Available at: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=71222

The Commission envisages the development of a general authorisation framework for 'data intermediaries' (such as data marketplaces, brokers, or facilitators). This framework will ensure that these data intermediaries know what rules they have to comply with. This also means that any party that meets the criteria specified by the framework can become an 'authorised data intermediary'. EU member states will oversee the rightful application of the framework through the appointment of a National Competent Authority (NCA).

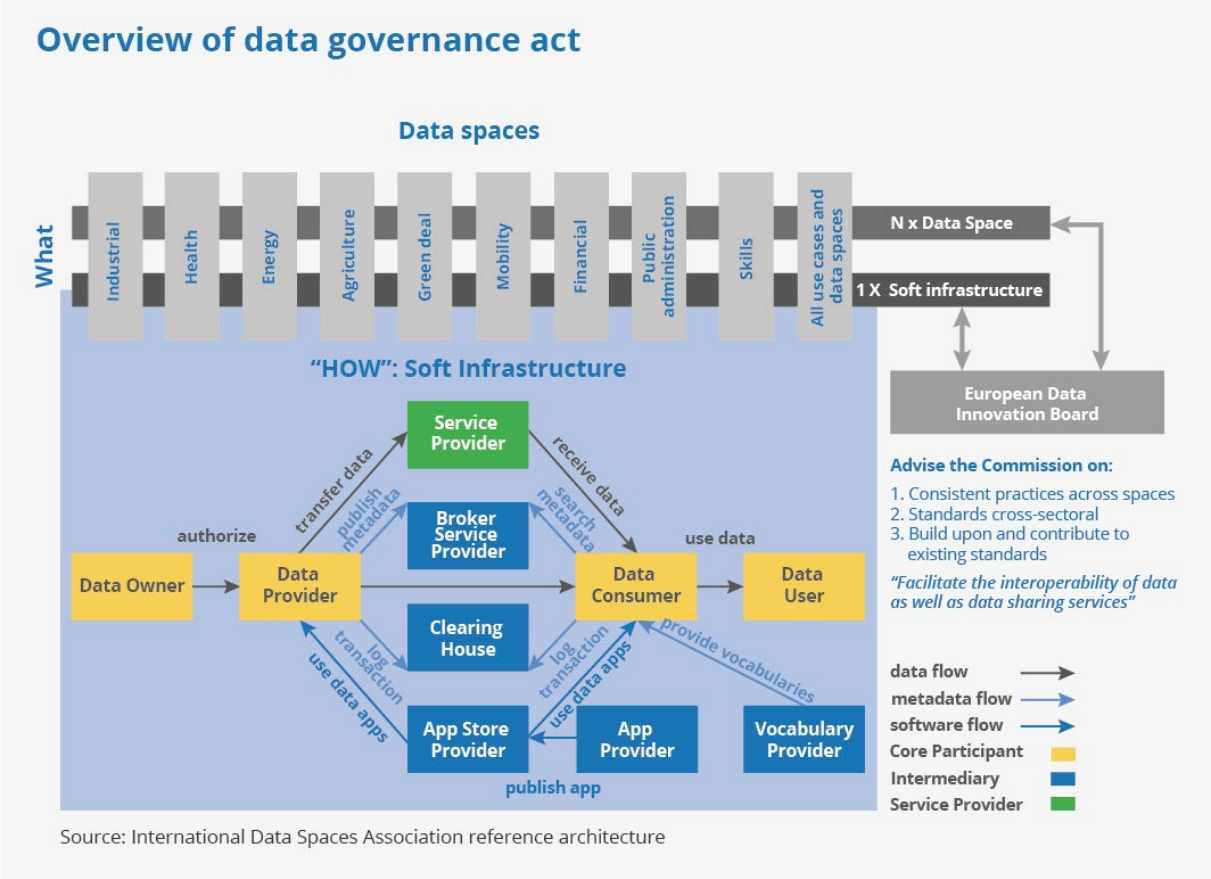


Figure 11 Overview of data governance act

The DGA stipulates the establishment of an expert group chaired by the Commission, called 'Data Innovation Board' (DIB). These experts, who will be selected by the Commission, will be representatives from European data spaces and from relevant economic sectors and interest domains, respectively. The DIB will advise the Commission with regard to establishing consistent practices of data sharing and exchange, maintaining the general authorisation framework, ensuring cross-sectoral interoperability of data spaces, and facilitating collaboration with National Competent Authorities (NCAs).

The Commission has not been explicit yet on how the general authorisation framework is to be created and how the totality of data-sharing applications will have to converge into the de-facto 'soft infrastructure'. What *is* clear, though, is that there will only be one authorisation framework for Europe, and that the NCAs (to be appointed) will take responsibility for overseeing the rightful application of the framework. This means that EU member states will not be allowed to create their own, national authorisation frameworks.

The big question now is: How do we get to this general European authorisation framework that will determine how data sharing and exchange in Europe takes place in practice?

The Commission and the DIB will play the strategic part in this endeavour. The DIB will set the direction and specify the terms and conditions for the general authorisation framework to meet. Therefore, the DIB must be mandated in a way that it can serve the goals of all EU Directorates.

In addition to the strategic level, Europe will need to address the tactical and operational level, i.e. the coordination activities that need to follow the strategic directives regarding aspects such as data space interoperability/federation and data sovereignty. The authors of the position paper therefore propose the establishment of a 'Data Exchange Board' (DEB), which will be responsible for detailing, maintaining and adopting the general authorisation framework, that will ultimately define the soft infrastructure. This means that the DEB will execute on the tactical and operational level what the DIB has set as goals on the strategic level (i.e. the DEB will be responsible for the 'How'-dimension indicated in Figure 11). To do so, the DEB will leverage all existing technological developments and collect the input from all stakeholders (data space users, technology providers, public sector, etc.).

A lot of experience with data spaces is at hand, and a lot of research on the topic has been conducted in the past years (the previous chapters of this position paper have elaborated on this already). From a technology and process viewpoint, there is no doubt that data space interoperability and data sovereignty can be achieved. This means: it is now a matter of coordination, collaboration, co-creation, agreement, and adoption. The regulatory direction set out by the DGA will certainly help achieve this goal.

The first task of the DEB will be to pave the way for getting to a single European authorisation framework by converging specialist knowledge and experience with regard to data space use cases. The DEB will consist of actors that have a direct interest in the establishment, governance, and adoption of a standardised approach to data sharing and exchange with the help of European data spaces. These actors will be researchers, practitioners, and public and private initiatives for data sharing and exchange. Specialists and researchers will be given an ongoing consultative and supportive role (some permanently, others on a project basis). The DEB will be an operational, permanently staffed organisation delivering against the strategic goals as set by the DIB and the Commission. The first years (decade) the DEB will focus on establishing the soft infrastructure that will ensure a unified user experience across data spaces. After this moonshot is realised, the DEB will change its focus to maintaining this soft infrastructure and keeping it up to date (change management).

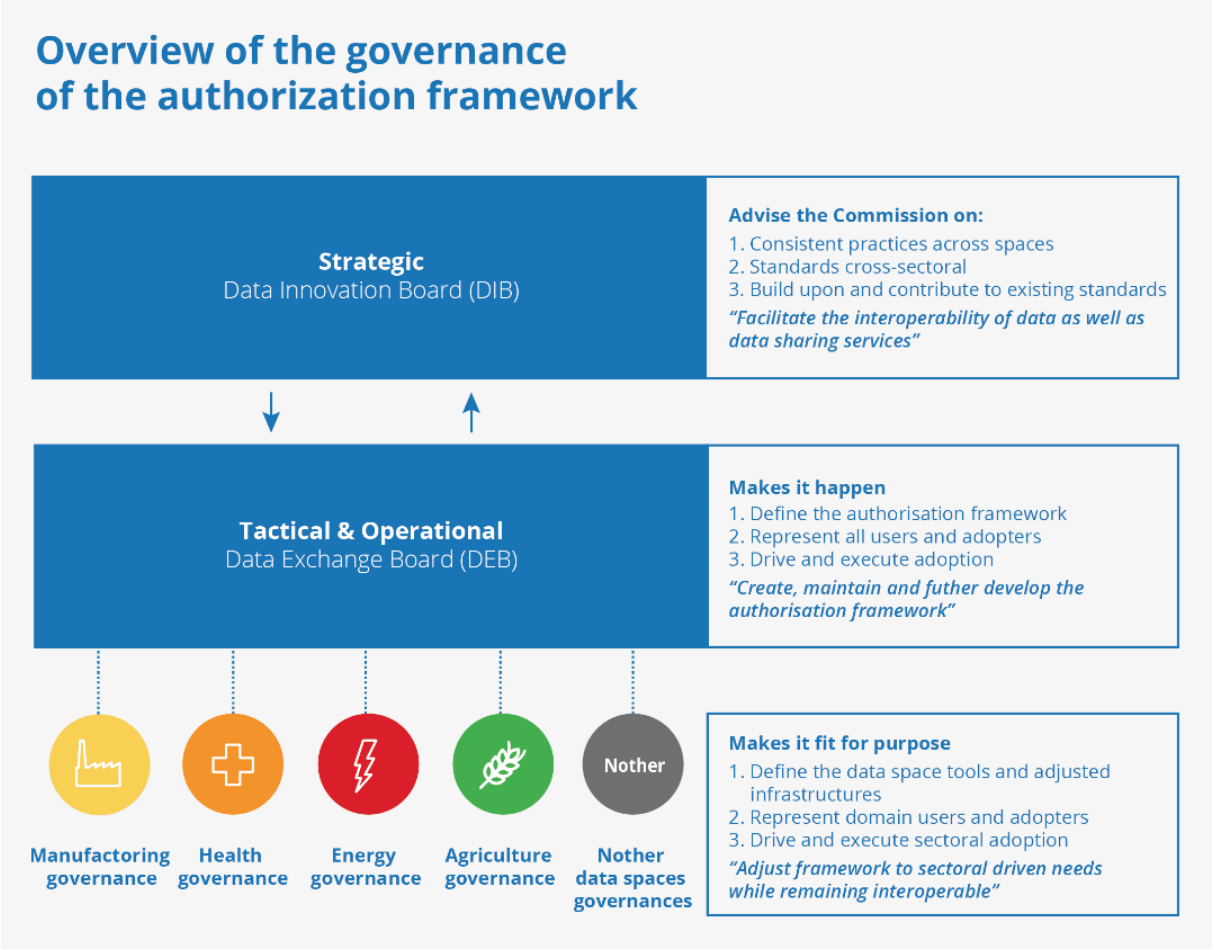


Figure 12 Overview of data spaces and their governance

The individual data space governance entities will have two focus areas: First, they are, together with the DEB, in charge of defining and facilitating the general authorisation framework; second, they are responsible for agreeing on the data-sharing aspects not covered by the general authorisation framework, but needed in order for their specific use case to work. As a result, we will see N x data spaces, each with its own governance function and connected to the overall DEB governance structure (under the auspices of the DIB). It can be expected that over time data spaces will converge with regard to how they are governed.

A first release of the general authorisation framework, and live implementations of data spaces adhering to it, can be expected within 18 months from project start. This will be possible because a broad foundation of research findings and practical experience regarding interoperable data spaces is already available today in Europe.

Once the first version of the framework is available, maintenance and further development of the framework will be done under a formal governance process. An essential part of this governance process will be the participant admission, certification, and monitoring process. If compliant with the framework, every actor could become an 'authorised data intermediary', no matter if it entered the data ecosystem only recently or has been a long-standing member of it. Existing platform businesses can be data intermediary as well as long as they abide the the soft infrastructure rule book and become authorised.

In some sectors, central governance organisations managing harmonisation initiatives regarding data spaces, have already been set up. Whenever possible, the effort of establishing European data spaces should build upon existing solutions and initiatives, including collaboration with players that have a proven track record in managing such an organisation. As a result, an optimal governance structure will evolve, where EU stakeholders work together and agree on the specifications and functions of European data spaces. In this process, the knowledge, products, and practices of actors already operating in data spaces need to be mobilised. Furthermore, lessons-learned should be collected from other areas in which collaborative governance processes are relevant (such as the internet, electronic payment, or telecommunications).

4.2 Overall Governance Structure

Figure 13 illustrates the tasks and functions of the DEB under the strategic guidance and responsibility of the DIB.

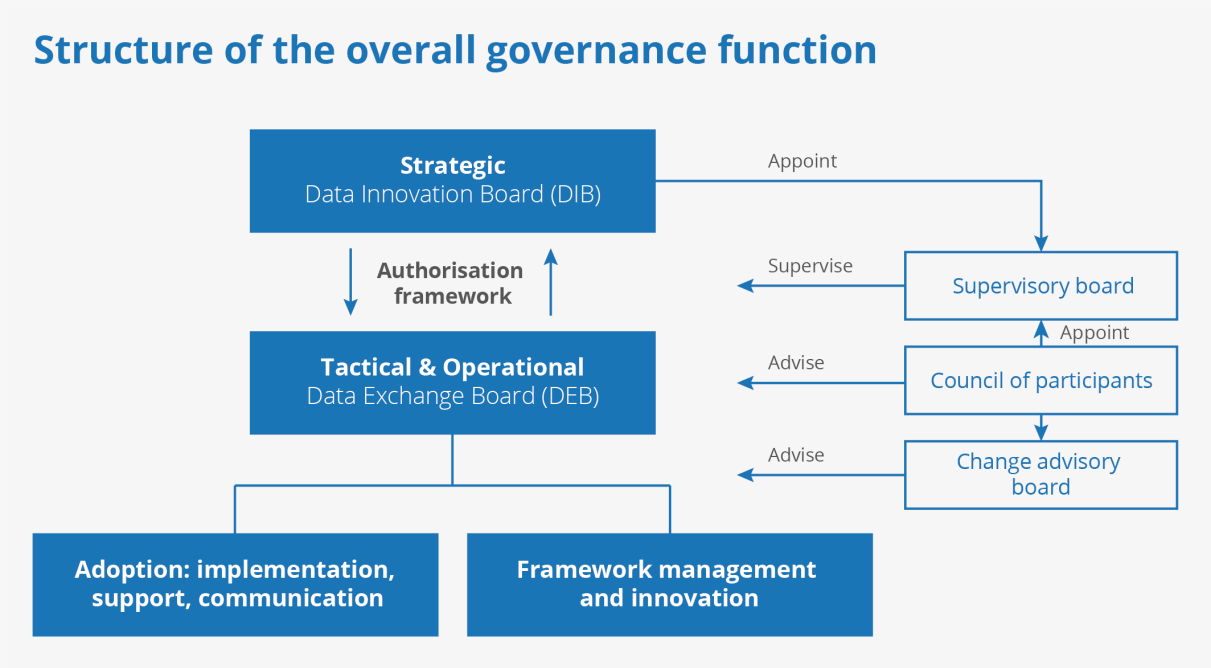


Figure 13 Overall governance structure for soft infrastructure and the data spaces

The DEB will have two main areas of activities: 1) taking care of the authorisation framework, and 2) dealing with everything concerning European data spaces being live and adopted (i.e. implementation, support, communication).

The DEB itself will have a two-tier structure. A central role will be given to the 'Council of Participants. This council could exist of representatives of Certified and Adhering Parties. The council of participants will be entitled to appoint the members of the Supervisory Board (together with the DIB).

The DEB will have its own staff. Part of this staff should be representatives from the governance organisations of the individual data spaces. Each data space's governance organisation will appoint a fair share of representatives to staff the DEB. Additionally, each data space's participants could send representatives to the Supervisory Board, the Council of Participants, and the Change Advisory Board.

The organisational structure of the individual data space's governance entities should be similar to the overall governance structure. The authors of this paper recommend defining a strategic, a tactical and an operational level for individual data space's governance as well, with

each level being supervised by organisational bodies consisting of representatives coming from participating organisations.

4.3 Functions of governance

The previous paragraph outlined how the DEB is responsible for governance on the tactical and operational level. This will ensure a level playing field and democratic decision-making regarding all aspects of European data spaces to establish. As explained above, a data space consists of various roles, most of them being non-exclusive and interoperable. A key distinction must be made between 'adhering' and 'certified' parties:

» Adhering parties (in short) are 'users' of the data space; adhering parties will be businesses, governments, or individuals. In most real life cases these adhering parties will adhere through the fact that they will subject to the terms and conditions of their IT providers, which in their turn implement the adhering requirements in their solutions to end users as defined by the authorisation framework

» Certified parties (in short) are those parties that play a facilitating role in the data(sharing) process enabling adhering parties to 'use' the data space. This means that any party that assumes a role such as data broker, data market player, e-commerce platform, software vendor, or data collective will need to be certified. This is a wider understanding than is currently mentioned in the DGA draft. In other words: all parties acting as service providers that hold data of businesses, governments, or individuals. Services to be rendered from such roles include identity provisioning, authorisation registry, functional testing, participant certification, security testing, etc. (all of which to be defined and agreed upon by the DEB).

Both certified and adhering parties will be supported by the activities of the governance entity. These activities will facilitate adoption of the framework and will support framework management and further development based on the input of stakeholders.

The term 'certified party' as used here and the term 'data intermediary' as used in the DGA interrelate as follows: all data intermediaries as meant by the DGA will (by definition) also be certified parties. However, the depicted soft infrastructure allows for possibly more types of certified parties than the data intermediary as now defined in the DGA. The soft infrastructure, by definition, will bring forward the rules and requirements required from each of the -to be defined- certified roles needed to run the ecosystem of the soft infrastructure.

Topics of governance of adhering and certified parties can typically be categorised into four main areas:

1. *Maintenance and further development* of the set of agreements and standards defining the 'soft infrastructure' (i.e. of the authorisation framework);
2. *Admission and certification* of the members of the network (i.e. of all data intermediaries);
3. *Technical and implementation support* for certified and adhering parties;
4. *Communication and education*, aiming both at end users and IT vendors/professionals.

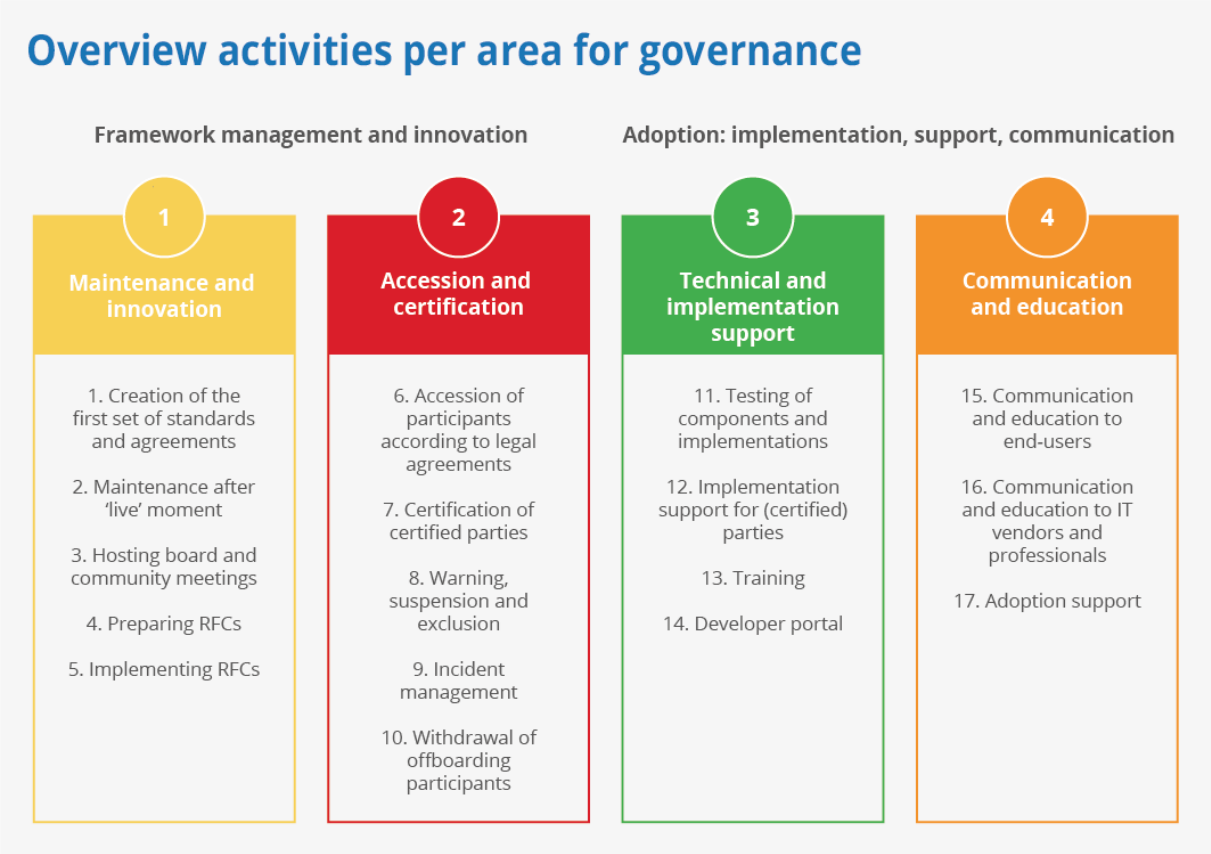


Figure 14 Activities in four areas of governance.

All these tasks and activities of the governance organisation need to be thoroughly developed and matched against each other. As the market evolves and more and more actors start using the authorisation framework and become part of the soft infrastructure, the level of detail, professionalism, and scale of these tasks and activities will increase.

4.4 Data space Business Models

Relevant business models in connection with data spaces will exist on two levels:

Level 1: For the individual actors (both adhering and certified parties)

Level 2: With regard to creating and maintaining the data space and the shared soft infrastructure.

4.4.1 Business model for individual actors

A data space is made up of multiple actors, which together form a data ecosystem. Data will only be shared and exchanged between actors if they decide to do so. One reason for doing so may be to create business value. Such business value does not necessarily need to be of monetary nature, but can also manifest itself in a better quality of a product or service.

Business models for monetisation of data exchange of adhering parties have been analysed in a number of studies. A common classification of business models is based on the given perspective, with a distinction being made between data exploitation and monetisation models from an internal perspective and from an external perspective. From an internal perspective, these models try to create value from the benefit gained from the data itself (e.g. gaining a competitive advantage in the market or minimising risks in the business). From an external perspective, data exploitation and monetisation models try to generate greater value for customers (either through raw data itself or through value-added services based on insights resulting from raw data). In the latter case, monetisation can either be direct (e.g. by selling data or a service) or indirect (e.g. by increasing customer loyalty). All these models have in common that all stages of data exploitation can be monetised (i.e. raw data, correlation of raw data with other data sources, descriptive/predictive analytics, etc.).

Regarding the external perspective, numerous studies about data monetisation are available, dealing with diverse use cases and different economic sectors. In the e-commerce sector, for instance, Amazon and AliExpress are two prominent examples of online sales and purchase, in which sellers show their products on the digital platform with the aim of generating direct return in the form of sales. A similar business model is the one of TripAdvisor in the tourism sector. These companies, which are data owners and at the same time marketplace operators and service providers, generate direct monetary return by receiving and aggregating raw data from data providers. In contrast, companies in the finance sector have credit data and customer data, which is anonymized, processed and sold (e.g. in the form of reports for the consumer sector).

Business models of certified parties and therefore data intermediairies will be more of a service provider nature. Reaping profits from the data itself is not allowed, as the DGA draft states.

4.4.2 Business model with regard to data space creation and maintenance

A data space consists of a set of agreements on legal, technical, functional and operational aspects as specified by the general authorisation framework. Based on this framework, actors providing and/or consuming data, as well as software vendors, can implement their own solutions. This set of agreements needs to be initially created and then maintained over time. The business model in this respect is one of collectively funding the creation and maintenance of data spaces, without a profit objective. The higher the adoption rate regarding European data spaces, and the higher the relevance of European data spaces for participants, the easier it will be to fund all efforts by the participants themselves.

As always, the beginning is the hardest thing. The 'chicken and egg' problem of adoption needs to be overcome in order to secure public interest in developing a common 'soft infrastructure' for the proliferation of data spaces. Once a critical mass of usage of the authorisation framework and the soft infrastructure is achieved, the recovery of the costs will be a trivial charge to the users of data spaces.

Therefore, it is crucial that sufficient public funding is guaranteed for the first several years, so that there will not be a single doubt for markets and participants about the commitment and sustainability regarding the establishment of European data spaces and the associated 'soft infrastructure'. It is just like the costs of the legal system (the costs of courts etc) or the costs of democracy.

In the first years, some common attitudes of prospective participants will need to be overcome, like e.g.:

» 'Everyone wants world peace, but nobody feels responsible for it.'

» 'I will only join a common standard or framework if it has turned out to be a success.'

» 'What is the business case for me?'

It is the responsibility of the public sector now to mitigate 'coordination failure' and cater for 'scalability to infinity'. By stepping in now with direction, regulation, and proper means, the European breeding ground for the next phase of the digital economy can be laid, where European businesses can thrive and new initiatives can compete on a level playing field. This can be the starting point for everything that is being done with data in Europe – be it data of citizens, businesses, governments, or things.

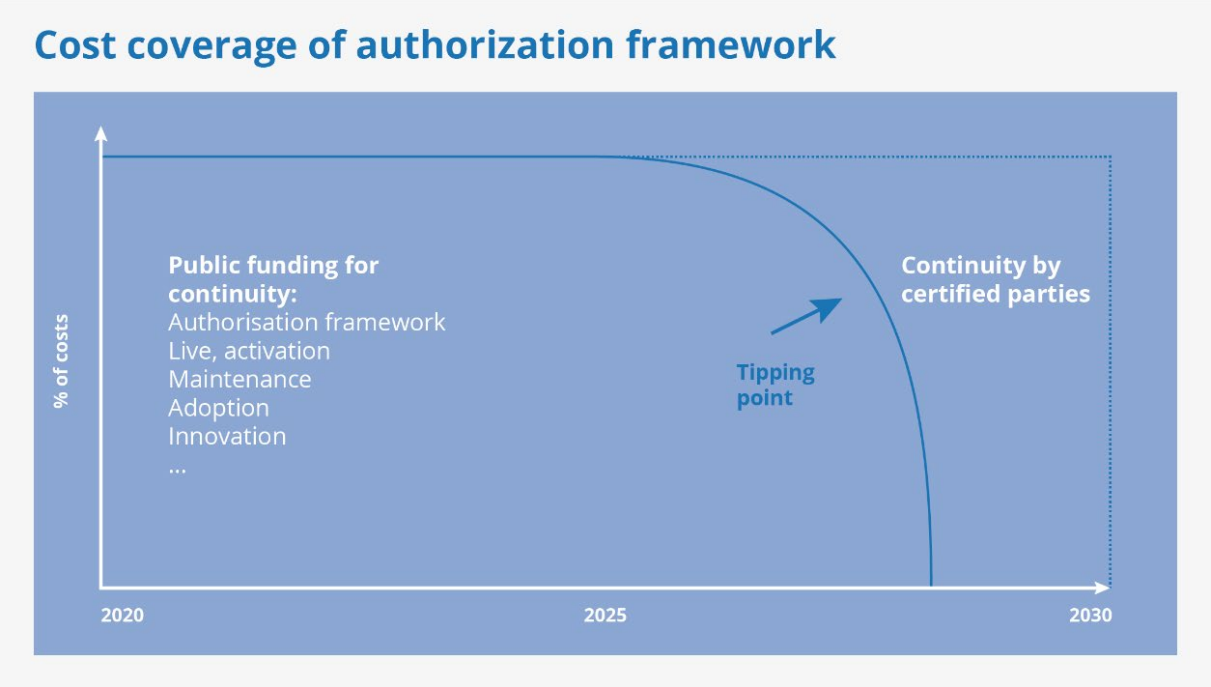


Figure 15 Cost coverage of authorization framework.

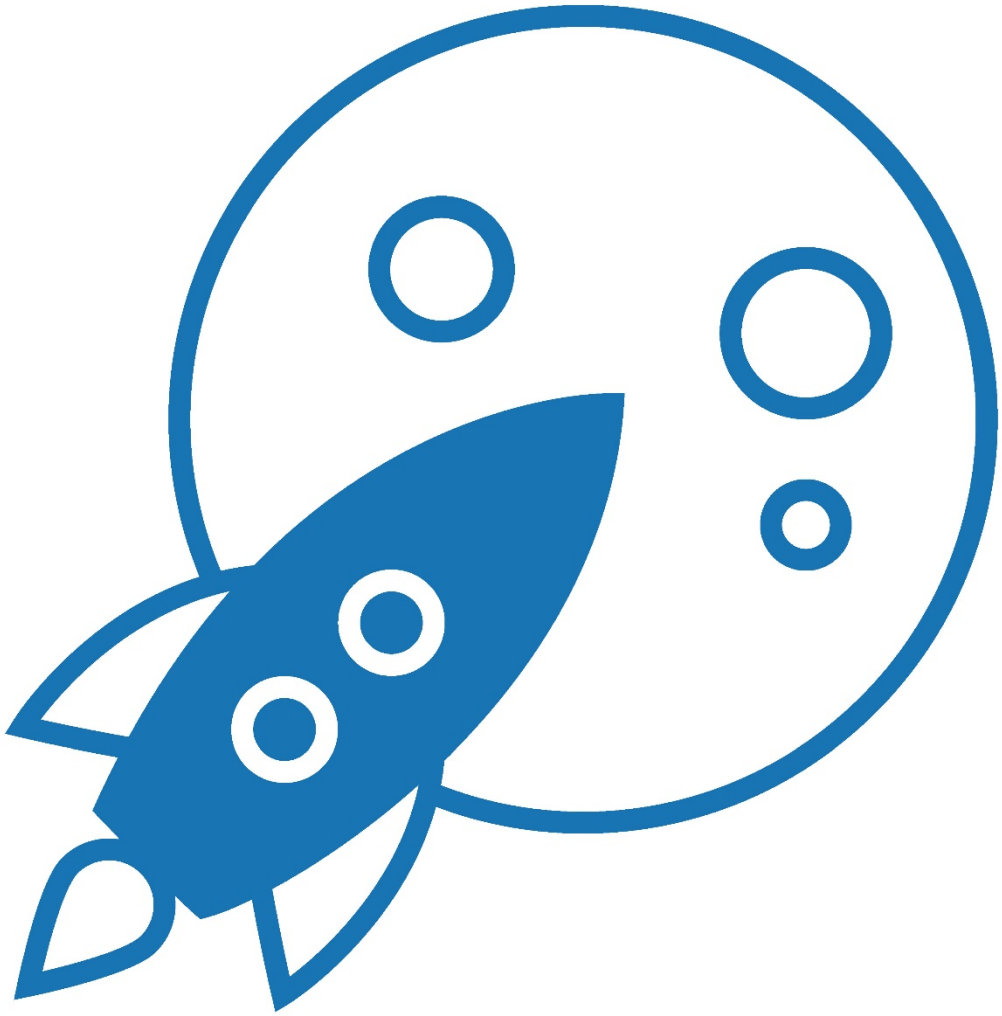
As Figure 15 suggests, initial funding must be provided by public money in order to accelerate the 'flywheel'. Over time, certified data space participants will contribute to funding. Once the break-even point is reached, public funding can be reduced to zero. If there is a proper setup and proper funding, the tipping point could be reached within six to eight years after project start.

The vision of European data spaces and the associated 'soft infrastructure' can be turned into reality at an estimated 1.5 to 2.0 billion euros within a decade. This must be done in combination with adequate market guidance through well-balanced regulation, for which the Digital Governance Act provides a sound basis. A major part of the required sum will be spent on the adoption of data spaces by SMEs, consumers, and governments, raising the awareness and communicating both the 'why' as well as the down-to-earth-benefits.

The main realisation is that the amount of investment in the next incarnation of the digital economy is inversely related to the economic and societal impact of data spaces. This is native for data-driven investments in general, as we can see from the capital investments in global big-tech platforms, which were relatively small compared to the economic and societal impact these platforms have had. This impact is comparable to the realisation of other societal infrastructures such as electricity, sewage and roads. Infrastructures build health and wealth for people and their nation. Now we have the opportunity to innovate our data infrastructure towards the same spirit where public and private interests are balanced.

Action: we call for a Moonshot

5



5 Action: we call for a Moonshot

The previous chapters of this position paper described the fundamentals of data spaces, and how to establish data spaces sustainably to leverage their economic potential in the long run. This last chapter outlines what steps European policy makers and politicians should take now in order to launch the 'data space rocket' to the moon. When President Kennedy called for the moonshot in 1961, this sparked a decade of innovations at all levels (new fabrics, technology, physics etc), but all in great coherence with the shared objective: man on the moon. All efforts converged towards the end goal. The same should happen here: it should be clear to all that Europe is creating a soft infrastructure with data sovereignty as key design principle, empowering citizens and organisations through their data.

The investment called for in Chapter 4 is needed to converge previously scattered initiatives throughout Europe and ultimately deploy European data spaces on the basis of a common, general 'soft infrastructure'. In line with the BDVA[89], we recommend the similar goals to be achieved within a time window of 10 to 15 years.

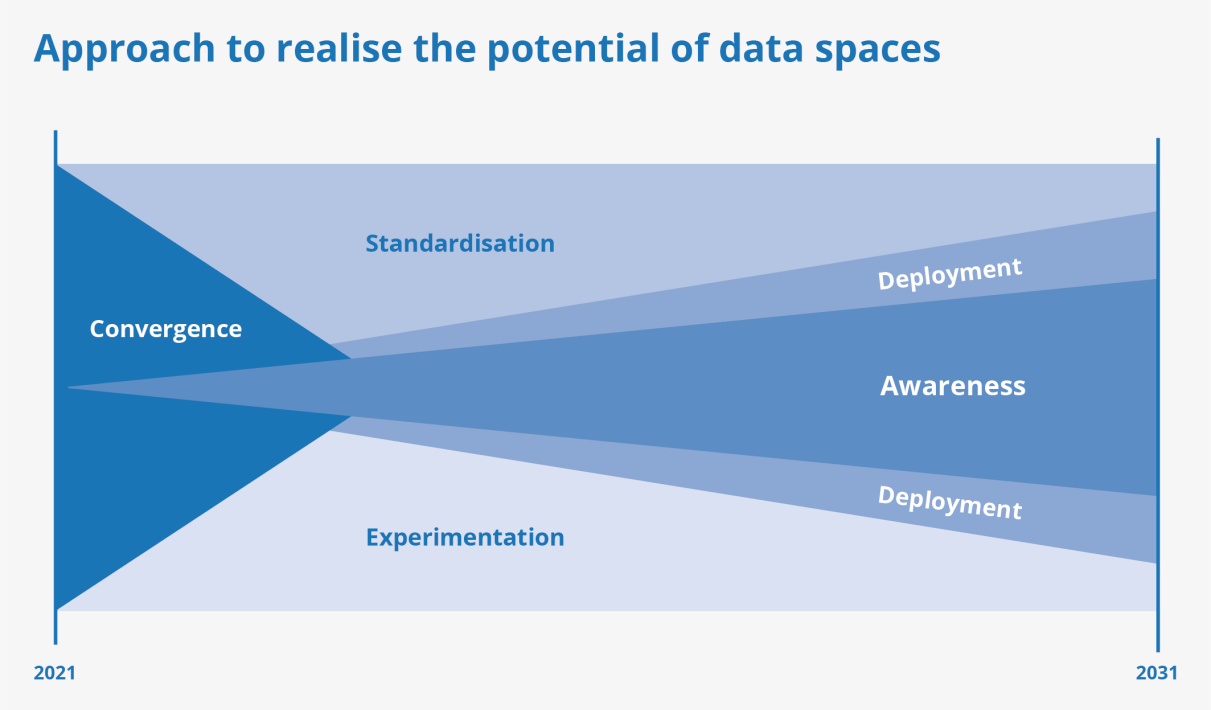


Figure 16 Schematic activity streams for the coming decade towards data spaces and their soft infrastructure (source: BDVA).

[89] BDVA White Paper, Towards a European-governed data sharing space, November 2020

Phase 1: Convergence

The first phase for establishing data spaces is about converging current European initiatives (e.g. IDSA, Data Sharing Coalition, MyData, BDVA, IHAN, FIWARE or Gaia-X) in order to co-create a single result, which will be well accepted for adoption by a critical mass of stakeholders: the first version of the soft infrastructure. This phase will take about 2 to 3 years. Three aspects will be mission-critical in this endeavour:

» Create awareness: Before the first version of the soft infrastructure is published, the concept, rationale, and functional range of the soft infrastructure needs to be communicated and promoted on a large scale. Even though the coalition will represent the market as good as possible, not all potential stakeholders can be involved in the co-creation process. Therefore, they should have the option to raise their voice during and after the creation process. This is all the more important as after the convergence phase adoption will start immediately, and much more stakeholders than those directly involved in the coalition, should be familiar with, and support, the agreements and standards.

» Establish governance structure: To do so, three steps are necessary: first, the governance structure proposed in Chapter 4 (see Figure 13) must be shaped, and the right people must be appointed as members of the DIB and the DEB; second, under the leadership of DIB and DEB operational processes must be defined (including communication, decision, and escalation lines); third, a coalition of the willing must gather with their use cases to populate the individual working groups on business / operational / legal and functional / technical matters. What is particularly important here is to include representatives from existing initiatives and ensure fair representation of all member states involved, and all industries affected.

» Co-create a set of agreements for soft infrastructure: Co-creation of the soft infrastructure mainly is about establishing coherent functional, operational and legal agreements as well as agreeing technical standards, which together provide the foundation for interoperability across data spaces. These agreements and standards must be specified in a rule book.

Phase 2: Deployment

Once the first version of the soft infrastructure is available, it is mandatory that the governance structure is in place and up and running in order to ensure broad adoption and sufficient scalability of data spaces. This phase will take about 10 years to achieve full maturity. By then, data space governance will fulfil the tasks as indicated in Chapter 4.3. (see Figure 14).

» Maintenance and innovation: The soft infrastructure will be the dynamic set of agreements and standards that moves with market needs, technological developments, and regulatory requirements. The DEB will be in charge of facilitating that these agreements evolve in line with the market. New use cases will shine a light on new, previously unaddressed user requirements. In periodic intervals, agreements and standards will have to be updated in good coordination with the stakeholders. Maturity of agreements and standards will depend on the level of adoption.

» Governance of daily operations and processes: After the convergence phase, operational governance will be in charge of creating and maintaining daily operations and processes. One of the most important processes will be admission and certification of new participants. As the number of participants will grow significantly over time, maturity of data space governance bodies will be dependent on the level of adoption.

» Awareness, education, and behaviour: At the beginning, a majority of data space participants will not be familiar with how to use and leverage data spaces. Therefore, ample attention should be given on creating awareness, providing education, and schooling behaviour of data space participants. This will not just be a means to increase adoption, but also to ensure that new participants understand the value and consequences of data spaces. While a push with regard to awareness, education, and behaviour should be subsidized by the Commission during the first ten years, communication and collaboration between data space participants themselves will take over this role in the end.

» Implementation support to speed up adoption: Once the rule book specifying the agreements and standards is available, it will be crucial to promote adoption. Implementation support via the DEB will help early adopters. They will be helping to generate the lighthouse effect on their initial application of data spaces in order to accelerate a snowball effect for further adoption across industries and interest domains. Once Adhering Roles start with adoption, the business case for Certified Roles will become stronger. In addition, Certified Roles will facilitate adoption and provision of value-added services for Adhering Roles. Over time, these network effects will ultimately result in an exponential increase in adoption, as knowledge, best practices and service offerings become more apparent.

In parallel with the convergence and the deployment phase, standardisation, experimentation, and awareness activities will have to take place during the next 10 years:

Awareness:
Creating awareness will be key during both the convergence and the deployment phase for three reasons:

0 First, it will ensure interested parties get to know the concept and functional range of data spaces;

1 second, it will scale up adoption from experimenting (local) initiatives to large-scale and interoperable usage. Specific attention needs to go to the IT service and solution sectors. There lies the amplifier for adoption, once these parties include data space functionality into their offering. They do this by adoption of the soft infrastructure rule book into their technical solutions.

2 Third and last, but certainly not least: it will establish trust among users; The importance of this cannot be overestimated. It should be clear to all involved that this is happening and why it is happening. Not just to those interested from the beginning, but especially to the larger audience.

Standardisation:
Alongside with the creation of the first version of the soft infrastructure, standardisation activities need to continue, steered by the market developments which will be voiced and prioritised by the market participants. Once the first version of the soft infrastructure's rule book is available, data spaces will develop in line with current trends and new developments on a global level regarding technological standards.

Experimentation:
Data space use cases will have two functions: during the convergence phase, they will provide a framework for key data space functionality; afterwards, during the deployment phase, they will serve to experiment on the technical and organisational conditions and demonstrate data space functionality to interested parties.

Conclusion and call-to-action
This position paper provides a clear path forward towards the establishment of European data spaces.

The Commission is well positioned to take the lead in the coming decade in supporting the co-creation process of developing the data space soft infrastructure in a coordinated and collaborative manner, focusing primarily on governance. Continuous financing for a decade is crucial. Probably an IPCEI[90] type of funding structure should be considered.

The Data Governance Act has confirmed the importance of governance in such an endeavour. Each data space will have its own governance entity, while there will be an overall governance structure referring to all aspects that lead to interoperability of data spaces.

The Data Exchange Board, which is proposed by the authors of this position paper, will complement the DIB (as specified by the Commission) on a tactical and operational level. With this governance structure as the foundation, the Commission should invest in a two-phase co-creation (converge and deploy) project ultimately leading to the development of the soft infrastructure underlying European data spaces, which in turn will lead to broad adoption and sufficient scalability of a renewed, thriving data economy.

With a critical mass of European member states collaborating in this endeavour, Europe will be able to stand its ground in the data economy by creating the data level playing for a thriving digital and sustainable economy for the decades to come.

[90] IPCEI: acronym for Important Project of Common European Interest

Appendix

6

6 Appendix

6.1 Glossary

Term	Definition
Accountability	Having accountability means that someone can be described as being liable or answerable for the completion of a certain task. Responsibility can be delegated, but accountability cannot.
Actor	An organisation or an individual performing one or more roles.
Application Programming Interface (API)	A technical interface consisting of a set of protocols and data structuring (API specifications) which enables computer systems to directly communicate with each other. Data or services can be directly requested from a server by adhering to the protocols.
Attribute	Any distinctive feature, characteristic or property of a data object that can be identified or isolated quantitatively or qualitatively by either human or automated means.
Authentication	A process that is used to confirm that a claimed attribute of an entity is actually correct.
Authenticity	In the context of information security, authenticity refers to the truthfulness of information and whether it has been transmitted or created by an authentic sender. Authenticity can be achieved, e.g. by digitally signing a message with the sender's private key. The recipient can verify the digital signature with the matching public key.
Authorisation	The process of giving someone or something permission to do something, for example to gain access to services, data or other functionalities.
Authorisation Registry (AR)	An authorisation registry manages Records of Authorisation (and, if relevant, Records of Delegation) so that Participants in the Collaborative Solution can verify whether a Data Consumer is authorised to access a specific Data Asset.
Authorization framework	Overarching framework on how to participate in data spaces, named by the Data Governance Act. See also soft infrastructure

Term	Definition
Bilateral Agreement	Covers agreements between two data-sharing actors, ranging from legal obligations to non-binding agreements of principle allowing them to share data.
Business and Policy Components	(e.g., SLAs, business models): These components specify and implement the policies that regulate the exchange and sharing of data between the different actors, on a business level
Certificate Authority	A trusted third-party entity issuing digital certificates (e.g. X509-certificates) or host services to validate certificates issued.
Collaborative Solution	A solution in which multiple stakeholders work together to facilitate many-to-many data sharing. The solution can make use of multiple models (i.e. platform and scheme).
Confidentiality	In the context of information security, confidentiality refers to the protection of information from disclosure to unauthorised parties.
Consent	Any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her.
Credentials	In the context of information security, credentials are used to control access of someone or something to something, for example to services, data or other functionalities. The right credentials validate (i.e. Authentication) the identity claimed during Identification.
CRUD	CRUD (acronym for Create, Read, Update, Delete) are considered to be basic functions regarding stored data. In computer programming, possible actions are often mapped to these standard CRUD functions in order to clarify the actions. For example, standard HTTP(S) actions GET and POST refer to Read and Create functions regarding stored data.
Data application providers	Providers of applications that transform, process or visualise data
Data Asset	A data resource, controlled by an organisation to generate revenue, e.g.: a system, application output file, document, database, web page.

Term	Definition
Data Anonymization	It is the process of removing personally identifiable information from data sets.
Data Consumers	An individual, group, or application that receives data in the form of a collection. The data is used for query, analysis, and reporting.
Data Governance	A system that employs interoperability components (standards and poli-cies) to ensure the acceptable use and high quality of data within a specific ecosystem. Manages the availability, usability, consistency, integrity, and security of the data used.
Data Exchange Board (DEB)	A proposed addition to the Data Innovation Board, that operates on a tactical and operational level
Data Innovation Board (DIB)	An expert group chaired by the commision. This governance body is foreseen by the Data Governance Act, that operate at a strategic level and advises the Commision
Data Level Playing field	Competitive landscape has an equal level playing field for all organisations. This means that large monopolistic players can no longer position themselves as exclusive data owners. Data owners regain control and can move between providers. Owning data is no longer the competitive dynamic that decides the winner, and organisations need to offer value to the users to convince them
Data marketplace	A place where data providers and data consumers can find each other to stimulate data exchange or access
Data marketplace provider	Provide capabilities that will allow for the operation of data marketplaces
Data Model	Description of how data can be stored, processed and accessed.
Data Owners	Entity which has ownership of the data and that has rights to grant or revoke terms and conditions for access and use of data
Data Pollution	The abundance of data in the digital environment and the damage this can cause to citizens and businesses. It arises from the fact that people and organisations have been giving away massive amounts of data for decades.
Data Producers	Any person, organisation or machine that produces data

Term	Definition
Data Portability	The ability of data to be easily moved across interoperable applications and domains. The legal right to data portability, granted in some jurisdictions to individuals, can be delivered through a range of technical mechanisms and varies in scope according to the jurisdiction. Our principle of data portability encompasses the ease of both access to and reuse of data.
Data Providers	Any person or organisation that makes data available.
Data routing and pre-processing (DR&P)	Routing of a data asset to the data consumer
Data Self-determination	The capacity of an individual or organisation to control who has access to their (personal) data and under what conditions (see also: Data Sovereignty).
Data Sharing economy (design principle)	Creating the conditions for data trading, which requires a) non financial incentive mechanisms, b) financial incentive mechanisms, including models to monetize data and methods to determine the value of data, and c) agreements mechanisms,
Data Sharing empowerment (design principle)	Ensuring that decisions can be made by appropriate stakeholders. This means that tools and organisational practices are available for a) governance in data spaces b) citizen engagement support c) data sovereignty support and d) federation
Data Sharing interoperability (design principle)	Providing the ability for all applications in data spaces to create, use, transfer and effectively exchange data. This requires among other definition of data exchange APIs and data models supporting a) semantic interoperability, b) behavioural interoperability and c) policy interoperability
Data Sharing publication (design principle)	Enabling data to be published so it can be easily located by data consumers.
Data Sharing trustworthiness (design principle)	Ensuring that data spaces operate according to expected requirements. This means that the development of data sharing applications must support a) security-by-design b) privacy-by-design c) assurance-by-design
Data Source	A source of data assets that is being exposed to data consumers by data providers. The role responsible for collecting, storing, and controlling personal data which persons, operators, and data using services may wish to access and use.

Term	Definition
Data Sources and Services Directory/Catalogue	This directory facilitates dynamic expansion of the data space. Is support dynamic discovery of data processing and data analytic services.
Data Sovereignty	The capability of an individual or organisation to be entirely self-determining with regard to their data (see also: Data Self-determination).
Data space	A data ecosystem, defined by a sector or application, whereby decentralised infrastructure enables trustworthy data sharing with commonly agreed capabilities (data sovereignty and roles).
Data space community (design principle)	Fostering maximum reuse of data space solutions. This includes a) open solutions b) reusability c) open source and d) sustainability of solutions
Data space engineering flexibility (design principle)	Providing the ability for engineers to add customised features in data processing applications and data platforms to enable a) interoperability flexibility b) trustworthiness flexibility and c) data processing flexibility
Data Using Service	The role responsible for processing personal data from one or more data sources to deliver a service.
Delegation	The act of designating someone or something to act for another or to represent others. In a data sharing scheme, this means that one party designates another party to share or consume data or to issue authorisations on their behalf.
Ecosystem	The overall system created by the activities and connections of a set of actors and infrastructure interacting according to a common set of rules. Multiple ecosystems can exist, overlap, and collaborate.
eIDAS	An EU regulation on electronic identification and trust services for electronic transactions in the European Single Market. This regulation covers important aspects related to electronic transactions, such as qualified electronic certificates. eIDAS provides a safe way for users to conduct business online.

Term	Definition
Encryption	Encryption is the process of converting data from plaintext to ciphertext. Plaintext (also called cleartext) represents data in its original (readable) format, whereas ciphertext (also called cryptogram) represents data in encrypted (unreadable) format. Decryption is the process of converting data from ciphertext to plaintext. The algorithm represents the mathematical or non-mathematical function used in the encryption and decryption process. A cryptographic key represents the input that controls the operation of the cryptographic algorithm. With symmetric encryption the same key is use for encryption and decryption, whereas with asymmetric encryption two different, but mathematically related keys are used for either encryption or decryption, a so-called public key and a private key.
Federated assurance management	It consists of having individual data spaces assurance management associated with a federated collaboration on a global security and privacy assurance management
Federated privacy management	It consists of having individual data spaces privacy management associated with a federated collaboration on a global privacy management
Federated security management	Having individual data spaces security management associated with a federated collaboration on a global security management. Such a framework includes five concepts: identify, protect, detect, respond and recover. Issues to be addressed include access control, usage control, trust and identity management.
Governance	A system of rules, practices, and processes used to direct and manage an ecosystem. A good organised governence needs 3 layers to work in close cohesion: Strategic, Tactical and Operational
GDPR	Global Data Protection Regulation (GDPR) is an EU regulation on data protection and privacy, applied from May 25th, 2018. The GDPR's primary aim is to give control to individuals over their personal data and to simplify the regulatory environment for international business
Hard infrastructure	The "tangible" part of the infrastructure, such as roads, rails, cables, but also software components. Elements that everybody who participates in the infrastructure can use
Identity Provider	An intermediary party offering services to create, maintain, manage and validate identity information for parties that share data within a collaborative solution (See also: Collaborative Solution).

Term	Definition
Individual	A natural, living human being.
Infrastructure	build the foundation on which all providers can provide individual services, while still interacting with each other. Infrastructures are sector agnostic. Infrastructures consits of hard and soft infrastructure (see hard and soft infrastructure)
Interoperability	The ability of different systems to work in conjunction with each other and for devices, applications or products to connect and communicate in a coordinated way, without effort from the person.
Levels of Assurance	Within online authentication, depending on the authentication protocol used, different levels of assurance give the server different degrees of certainty about the client's identity. Depending on parameters such as the quality of the registration process, quality of credentials, use of biometrics or multiple authentication factors and information security, an authentication protocol can provide a server with high or low confidence in the claimed identity of the client. For low-interest products, a low level of assurance might be sufficient, while for sensitive data it is essential that a server is confident that the client's claimed identity is valid.
Metadata	Information about data that helps describe, structure or administer that data.
Non-repudiation	In the context of information security, non-repudiation refers to the fact that the sending (or transmission) and receipt of the message cannot be denied by either of the involved parties (sender and recipient).
Operator	The role responsible for operating infrastructure and providing tools for the person in a human-centric system of personal data exchange. Operators enable people securely to access, manage, and use personal data about themselves as well as to control the flow of personal data within and between data sources and data using services.
Operator Network	A group of operators with some degree of mutual interoperability.
Person	The role of data subject as represented digitally in the ecosystem. Persons manage the use of personal data about themselves, for their own purposes, and maintain relationships with other roles.

Term	Definition
Policy Administration Points (PAP)	Entity where policies are administered
Policy Decision Points (PDP)	Entity that evaluates access requests that are received from the policy enforcement point (PEP). Subsequently an answer is sent back to the PEP.
Policy Enforcement Point (PEP)	Entity that determines whether an action is permitted or not. It takes any access requests and forwards these to the policy decision point
Policy Information Points (PIP)	Entity that collects additional (mostly dynamic) information on the data sharing to make accurate policy decisions
Proto operator	A product, service, or organisation that is in one way or another performing the role of an operator in personal data ecosystems or offers related tools, services, or technologies.
Persistent Identifier	A sequence of characters that identifies an entity, usually in the context of digital objects that are accessible over the internet. Typically, such an identifier is not only persistent but also actionable, i.e. it is a Uniform Resource Identifier (URI), usually of the https type, that you can paste into a web browser to be taken directly to the identified source.
Platform	A platform facilitates the exchange of value between two or more parties, with the multiple parties interacting through the platform.
Platform providers	Provide capabilities that will allow for the operation of (data) platforms
Provenance	Data origin.
Role	A function or set of responsibilities for a particular purpose.
Scheme	A common set of multilateral agreements that facilitates standardised and decentralised data sharing directly amongst participants.
Self-sovereign Identity (SSI)	A model for managing digital identities in which an individual or organisation has sole ownership over the ability to control their accounts and personal data without the need for intervening administrative authorities. SSI allows people to interact in the digital world with the same freedom and capacity for trust as they do in the offline world.

Term	Definition
Separation of Concerns (SoC)	A principle by which a modular approach to the development of a system is adopted. This approach entails each section addressing a different aspect (concern) of the overarching system. In the context of SoC in the personal data ecosystem, processing, storing, aggregating, displaying, governing data are concerns that need to be managed in a modular, transparent manner. SoC enables more opportunities for module upgrade, reuse, and independent development.
Soft infrastructure	A soft infrastructure is invisible, made up of technology neutral agreements and standards, on how to participate in an ecosystem. As all participants implement the same minimal set of functional, legal, technical and operational agreements and standards, they can interact in the same manner independent of the sector or domain
Structured Data Assets	Data that adheres to a predefined data model which is primarily useful for interpretation by machines.
Technical and Technological Components	(e.g., Software, Hardware, Middleware): These components enable the development of the technical solution of a data space. For example, they include devices, network protocols, middleware components and APIs that that enable the exchange of data between different platforms in secure and trustworthy ways
Trust Framework	A structure that lets people and organisations do business securely and reliably online.
Unstructured Data Assets	Data that does not have a pre-defined data model or is not organised in a pre-defined way, making it primarily interpretable by humans.

Contact on behalf of the
OPEN DEI Project

Head Office

INTERNATIONAL DATA SPACES ASSOCIATION

Emil-Figge-Str. 80
44227 Dortmund | Germany

phone: +49 231 70096 501
mail: info@internationaldataspaces.org

WWW.INTERNATIONALDATASPACES.ORG

@ids_association

international-data-spaces-association