

Chapter 1

The Evolution of Data Spaces



Boris Otto

Abstract The role data plays in enterprises is changing as the digital transformation in many sectors gains speed. New business opportunities through data-driven innovation emerge from data sharing in ecosystems. In ecosystems, the interest of the individual must be brought into alignment with the interest of the ecosystem. Trust between participants, data interoperability, and data sovereignty are key requirements which can be met by data spaces. Data spaces are a distributed data integration concept which is taken up by consortia aiming at supporting ecosystem. GAIA-X and IDS specify reference architectures for distributed data infrastructures and data spaces, respectively. While the benefits of data spaces for a fair data economy are recognized by business and policy makers, a deeper understanding is required about the design and evolution of data spaces. This chapter introduces fundamental concepts, identifies design tasks and options, and, thus, provides guidance for the establishment of data spaces.

1.1 Data Sharing in Data Ecosystems

1.1.1 *The Role of Data for Enterprises*

The role of data for businesses has continuously changed over the last decades. The proliferation of the platform economy and the continuing consumerization of many areas of business, in particular industrial activities, are only two examples of developments which require enterprises to rethink the way they manage data—both internally and together with external partners.

When it comes to the role of data for enterprises today, four types of roles can be identified. First, data is still—as it has been over the last decades—an enabler of operational excellence within a company. Integration and automation of business processes requires an effective and efficient data resource management. Second, data

B. Otto (✉)

Fraunhofer Institute for Software and Systems Engineering ISST, Dortmund, Germany

e-mail: boris.otto@isst.fraunhofer.de

© The Author(s) 2022

B. Otto et al. (eds.), *Designing Data Spaces*,

https://doi.org/10.1007/978-3-030-93975-5_1

has become a product which is sold in the market. For example, mobile telecommunication providers are used to sell anonymized data on the mobility and movement of their customers. The information generated from this data can be used, for example, by transportation and traffic agencies for maintenance plans of highway infrastructures or for better traffic management. Third, data is a source of business innovation. Data-driven services in various domains require access and combination of data from different sources provided by different members in a business ecosystem. Original equipment manufacturers (OEM) of industrial machinery, for example, collaborate with their customers, with service operators, component suppliers, and big data analytics firms to be able to offer better end-to-end maintenance services to their customers. And fourth, data is considered a strategic resource for the long-term sustainability of the economy. The European Union, for example, estimates the value of the data economy at 550 billion euros at minimum [1].

The value of data can only unfold when data is used [2]. Therefore, both policy makers and the private sector have an interest in increasing the use, and the re-use, of data. In particular, industrial enterprises possess a “treasure of data” which stems from manufacturing processes, from condition monitoring of installed base products and from the customer experiences. However, holders of this data have also an interest in not getting exploited when it comes to data. Making data available—in particular data of good quality—is not for free but comes at a cost. Thus, data providers must have a stake in the question on how their data should be used by others.

1.1.2 Data Sharing and Data Sovereignty

Data sharing goes beyond exchanging data. Data exchange has been happening between companies for more than 40 years. Electronic Data Interchange (EDI) is a good example. Message standards such as ANSI X.12 or EDIFACT define data exchange messages for trade and commerce and standardize purchase orders, invoices, dispatch notifications, etc. However, collaboration between EDI partners does not go beyond the pure exchange of data. The use of the data is limited for internal purposes of the involved parties. Data sharing, in contrast, means a collaborative use of the data for a shared goal [3]. Examples are collaborative customer innovation or efficiency gains for an entire group of companies.

In this context, ecosystems are a multilateral form of organizing to achieve a shared goal. In contrast to networks, for example, ecosystems are more centered around a customer innovation, more dynamic when it comes to their composition and characterized by a balance of prosperity of the individual members on the one hand and the ecosystem as a whole on the other hand [4].

Customer innovation is often realized through end-to-end support of a customer process. Good examples can be found in the mobility domain. Intermodal mobility services which support mobile individuals to plan, orchestrate, and perform their trips from the start of their trip to their destination are only possible when different

members of the mobility ecosystem team up. Examples are local and regional public transport providers, railway companies and airlines, car rental and taxi services, hotel chains, etc.

To ensure fairness in the data sharing and, thus, sustainability of ecosystems, data sovereignty is a key prerequisite. Data sovereignty refers to the capability of a legal entity or natural person to determine and execute usage rights when it comes to their data.

Data spaces support data sharing and data sovereignty in ecosystems as they are based on a distributed software infrastructure which provides the required software functionality.

1.1.3 Example Mobility Data Space

Figure 1.1 shows the architecture model of a mobility data space which aims at enabling data-driven mobility services, for example, intermodal mobility as a service (MaaS) as mentioned above.

The architecture consists of three layers. The mobility ecosystem represents the first layer on which services are offered to the mobile citizen. These services improve the mobility experience of the individual, allow for better management of traffic flows, and increase the utilization of means of transportation, among others.

The services on the first layer require a mobility data space in the narrower sense, i.e., a “shared digital twin” of the various constituents of mobility ecosystems. The digital twin consists of the data of the individual digital twins and comprises, for example, timetables, charging statuses of electric scooters, utilization of buses and trams, the travel preferences and plans for individuals, etc. The shared digital twin represents the second architecture layer.

Data sovereignty [5] must always be ensured for all participants in this data space. Data providers must have control and transparency of what happens to their shared data, and data consumers must be able to trust both data providers and data sources. A federated software infrastructure is needed as the third architecture layer. Relevant software services ensure data interoperability, data sovereignty, and trust among participants.

1.1.4 Need for Action and Research Goal

The example of the mobility domain shows the important role data spaces play both for companies to innovate and also for governments and policy makers to ensure data sovereignty on an economic level.

Both the European Data Strategy [6] and the Data Strategy of the Federal Government of Germany emphasize this important role. The European Data Strategy

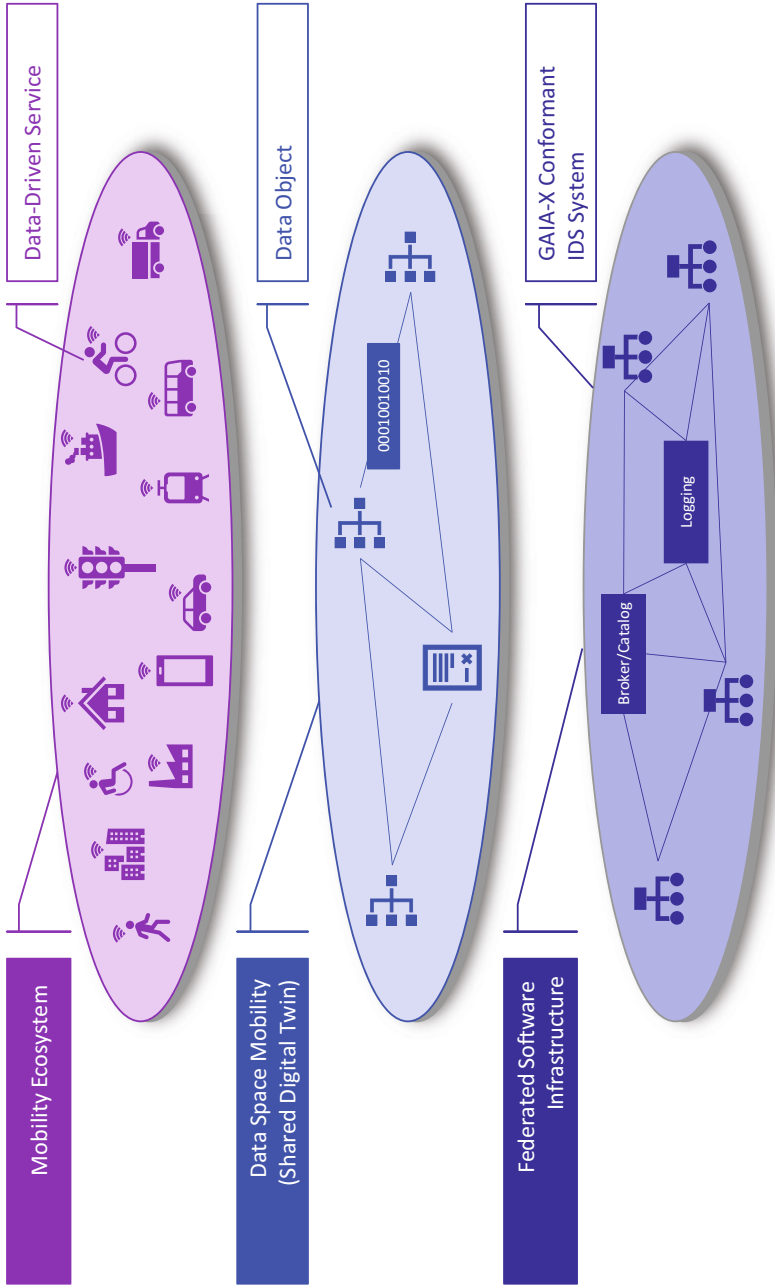


Fig. 1.1 Mobility data space (©2021, Fraunhofer ISST)

calls for the establishment of domain-specific data spaces in the Single European Market.

Data spaces are distributed by design and address not only individual companies but ecosystems or even entire domains. Their underlying software infrastructure must be in place before business innovation can be realized. It is an important cornerstone of the data economy mentioned above.

As with all infrastructures, many stakeholders are concerned with a set of questions when it comes to planning, designing, implementing, and maintaining this software infrastructures for data spaces. Typical questions are related to the functionality of the infrastructure services, to their openness, to their funding and financing, and to their governance. At present, many of these questions are still unanswered.

In this context, this chapter aims at laying some fundamental foundations for data spaces, to elaborate on the evolution of data spaces and to analyze the most important design tasks to be addressed. It helps both individual businesses to position themselves in existing and starting data space activities and it supports policy makers in their endeavor to pave the way for a fair data economy.

1.2 Conceptual and Technological Foundations

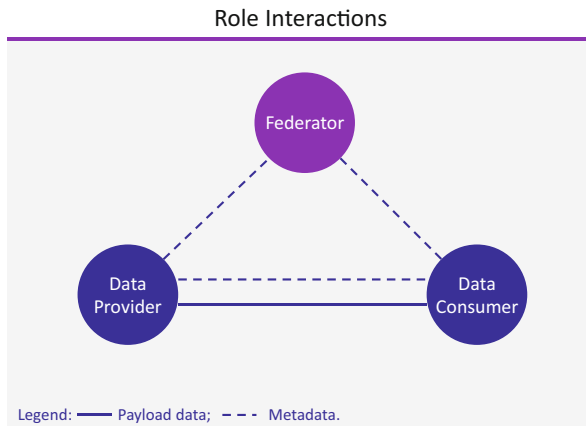
1.2.1 Data Spaces Defined

The notion of data spaces was coined some 15 years ago in computer science [7, 8]. Data spaces were introduced by that time as a data integration concept. In contrast to central data integration approaches (e.g., data consolidation hubs), data spaces do not require a physical integration of the data, but leave the data stored at the source. In addition to that, they do not require a common database schema that data from various sources must adhere to. Integration is rather achieved on semantic level using shared vocabularies, for example. Because of that, data spaces allow for data redundancies and “co-existence” of data. Furthermore, data spaces can be nested and overlapping so that individual participants can be part of multiple data spaces.

Besides this original technological definition of data spaces, the increased use of the term in the business community has led to an understanding of the data space notion as a form of collaboration on data. Practitioners in various industrial domains interpret data spaces as a business collaboration format driven by the desire to achieve shared goals. An example is Catena-X, the initiative launched by parts of the German automotive industry, which aims at a data space allowing for integrated trusted data chains in the automotive supply and production network. The business definition of data spaces refers to goals and decision-making rights and processes between the consortiums of participants.

In addition to that, the infrastructure nature of data spaces requires a common understanding of the concept from a legal point of view. Data spaces and their

Fig. 1.2 Data space roles
(©2021, Fraunhofer ISST)



underlying software infrastructures must support trust, interoperability, and portability of data and data sovereignty and must be nondiscriminatory. Thus, data spaces can be understood as intermediaries and data sharing service providers to which the EU Data Governance Act¹ applies which is currently under review.

1.2.2 Roles and Responsibilities in Data Spaces

Figure 1.2 shows the fundamental roles in a data space in their interactions between each other. As mentioned above, a data space is a distributed data integration concept. Thus, there is no central data store or data vault into which data providers deliver their data and from where it can be accessed and retrieved by data consumers. In contrast, the exchange of the data happens directly between the two participants.

However, to meet the fundamental requirements of data spaces in terms of trust among participants, data security, and interoperability, intermediary services are required. The role of the federator is to provide these intermediary services which include cataloging and brokering of data sources, ensuring trust between participants, and offering data sovereignty services.

Table 1.1 shows the responsibilities of the three roles when it comes to data sharing and exchange in data spaces.

¹See Regulation of the European Parliament and of the Council on European data governance of 2020 [9].

Table 1.1 Data space roles and responsibilities

	Data provider	Data consumer	Federator
Publish data source	R		S
Search for data source		R	S
Register participant	R	R	S
Identify participant	R	R	S
Authorize participant	R	R	
Request data exchange	S	R	
Perform data exchange	R	S	
Log data exchange	S	S	R
Constrain data use	R		
Perform data use		R	

Legend: R, responsible; S, supportive

1.2.3 GAIA-X and IDS

The International Data Spaces (IDS) initiative was launched in 2015 with Fraunhofer research project funded by the German Federal Ministry for Education and Research. It aimed at the design and prototyping of a distributed software architecture for data sovereignty. In parallel, the IDS Association (IDSA)² formed as a not-for-profit industry association and took up the work of Fraunhofer research to further develop it into the IDS Reference Architecture Model (IDS RAM). The IDS RAM is a technology-agnostic architecture description for a data space software architecture. Central software components described in the IDS RAM (e.g., IDS Connector, IDS Broker, IDS Clearing House) found their way into DIN SPEC 27070 which provides a blueprint for a secure gateway for trusted data exchange.

Today, IDSA has more than 130 members from more than 20 countries. Major IDSA activities are the maintenance of the IDS RAM, the definition of the certification process and the role of the Certification Body, and the implementation of an open-source software (OSS) strategy.

The GAIA-X initiative³ formed as a response to a call for a data infrastructure articulated in the German strategy on artificial intelligence. GAIA-X aims at data sovereignty in a broader context as IDSA does, because within the GAIA-X are not only data sharing and exchange but also the storing and handling of data on cloud platforms.

The Federation Services form the core of the GAIA-X architecture. They comprise a federated catalogue of distributed services, sovereign data exchange, identity and trust management, and compliance services.

The IDS and GAIA-X initiative are closely aligned in order to allow for seamless integration of the architectures and support processes [10].

²See <https://internationaldataspaces.org/>.

³See <https://www.gaia-x.eu/>.

1.3 Evolutionary Stages of Data Space Ecosystems

Data ecosystems are complex multilateral forms of organization which involve multiple different members. Therefore, ecosystems develop along evolutionary stages (see Fig. 1.3).

In a first stage, ecosystems are closed setups between a limited number of members. In this case, a separate entity of a federator is not necessarily needed, but the federator responsibilities are taken over by one of the members. An example of a closed ecosystem is the production and supply network of a single OEM in the automotive industry.

The second evolutionary stage is characterized by the openness of the ecosystem regarding their members. Participants are not always known but come and leave in a dynamic fashion. This leads to increased requirements when it comes to trust and interoperability, for example. The Catena-X initiative with its present scope is a good example for an open ecosystem because Catena-X is explicitly directed at the entire automotive supply chain with its multiple tens of thousands of companies.

A third evolutionary stage reflects the fact that individual members do not belong to one ecosystem only but are members of multiple ones. In this case, data sharing across the boundaries of an ecosystem must be possible which leads to additional requirements when it comes to trust among participants, interoperability of data and metadata, etc.

Table 1.2 shows different data space implementation options depending on the three evolutionary stages of ecosystems. Complexity in terms of interoperability, sovereignty, and trust and security increases along the evolutionary path.

To achieve interoperability of data space activities across ecosystem boundaries, for example, it must be made sure that federators agree on a unique system of identifiers and description schemes of data sources. Similarly, data usage control policies as a means enabling data sovereignty must be unambiguously understood across different ecosystems as well. The same holds true for digital identities which are used to identify and authenticate participants.

Because of that, an “ecosystem of federators” must be implemented—on top of the ecosystems of participants.

1.4 Designing Data Spaces

1.4.1 *Ecosystem Perspective*

As mentioned above, data spaces must be seen in the context of the ecosystem they support and the underlying software infrastructure. Taking an ecosystem view, design activities address the three architecture layers introduced in Fig. 1.1.

The business layer comprises incentive systems for participants and funding and financing models, for example. Even though data spaces follow a distributed design,

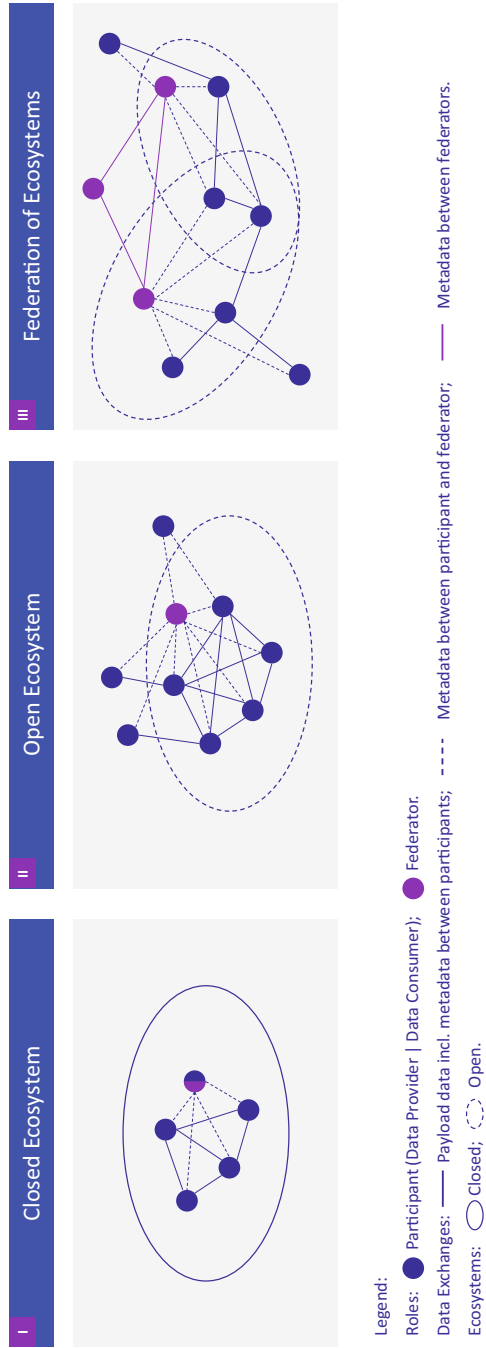


Fig. 1.3 Ecosystem evolutionary stages (©2021, Fraunhofer ISST)

Table 1.2 Data space implementations

	Closed ecosystem	Open ecosystem	Federation of ecosystems
Interoperability	<ul style="list-style-type: none"> • Proprietary schemas possible • Typical use of available domain-specific standards 	<ul style="list-style-type: none"> • Domain-specific, open standards required • Common vocabularies advisable 	<ul style="list-style-type: none"> • Cross-domain, open standards advisable • Mapping/translation between domain-specific standards • Uniqueness of identifiers (e.g. URIs) across domains needed
Sovereignty	<ul style="list-style-type: none"> • Traceability and transparency of data exchange not required if all participants are known 	<ul style="list-style-type: none"> • Increasing demands for policy enforcement because of unknown participants • Policies to be unanimously understood within ecosystem • Policies to allow for automated negotiation 	<ul style="list-style-type: none"> • Demand for policy enforcement because of unknown participants and cross-domain • Policies to be unanimously understood within ecosystem and to allow for automated negotiation
Trust and security	<ul style="list-style-type: none"> • Trust through the consortium • Digital certificates/tokens not necessarily required 	<ul style="list-style-type: none"> • Digital certificates/tokens required because of unknown participants • Dynamic technologies required in case of sensitive data use cases 	<ul style="list-style-type: none"> • Digital certificates/tokens to support cross-ecosystem application • Uniqueness of identities required

they form platforms for ecosystem. Therefore, fundamental economic characteristics of the platform economy such as network effects do also apply for data spaces. The data space services must be attractive for data providers, and data consumers and incentive systems must be put into place to achieve a critical mass of data providers. The more data sources are available through the data space, the more attractive it will be for data consumers.

Furthermore, funding and financing models must consider the infrastructural nature of data spaces. Investments in data spaces must be made before data-driven services can flourish. As there is both a public (and community) and a private (or individual) interest in the existence of data spaces, data space funding should come both from public and private sources.

The governance layer comprises questions of ecosystem governance and collaborative data governance. The first is related to the institutionalization of data space consortia. At present, one can observe the creation of joint ventures with or often without profit-oriented business purposes. The Mobility Data Space initiative in Germany, for example, plans to establish a not-for-profit limited liability company.

For data ecosystems which use data from individuals (e.g., consumers), calls are articulated to create so-called data trusts in which many individual data providers pool their interest in order to form a counterweight to large data platform providers.

Organizational governance and data governance are closely interrelated in data space organizations. Collaborative data governance mechanisms need to be put into place to determine data visibility, transparency, and sovereignty in data spaces. For example, a data provider may allow any data space participant to find their data source or limit it to a dedicated group of users. Central element of collaborative data governance are rules which are commonly understood when it comes to articulating usage conditions on the data. These rules can be understood as “terms and conditions” in the data economy.

On the technology layer, shared vocabularies must be established on multiple levels. First, a vocabulary must be in place unambiguously describing the key concepts of a data space (e.g., roles, responsibilities, characteristics of data resources, usage policies, etc.). Second, in order to achieve interoperability not only with regard to the exchange of data but also with regard to the shared use of data, vocabularies are needed to harmonize the understanding of the meaning of “payload” data itself. A payload data example in the mobility domain are timetables which need to be consistently understood between the different data space participants.

Furthermore, the software infrastructure must be built. The GAIA-X architecture in combination with the IDS RAM forms a “blueprint” for data space implementation.

1.4.2 Federator Perspective

A key role in data spaces is the federator. When designing a data space, the instantiation of the federator role is critical.

Design tasks comprise among others:

- **Portfolio of data space services:** In most cases, the portfolio comprises services described in architecture proposals such as the GAIA-X Architecture Document and the IDS RAM. However, on top of that, further services may be needed and desired. Examples of such additional data services are data traceability, data quality assurance, data trustee services, as well as mapping services.
- **Degree of decentralization:** Services provided by the federator enable data exchange and sharing of participants. Thus, they ensure the functioning of distributed data space designs. However, federator services can be designed and implemented in a decentralized way as well. Federator services can be:
 - Distributed, i.e., focusing on interoperability
 - Federated, i.e., focusing on being perceived “as one”
 - Shared, i.e., being implemented once for all participants.
- **Data Space Business Services:** The software services of the federator are the core of the service portfolio of a data space organization. However, there may be a

need for additional services. Examples are on-boarding services for participants, integration services, and value-added business services such as billing, etc.

The individual design of the federator role depends on the purpose of the data space, the scope of the shared goal of the participants, and its legal and regulatory environment.

1.5 Summary and Outlook

This chapter motivates data spaces as a suitable instrument to establish data ecosystems based on fair data sharing, i.e., on trust between participants, data sovereignty, and data interoperability. It introduces fundamental concepts necessary to understand data spaces from a technological, business, and legal point of view and outlines stages of data space evolution. Furthermore, it identifies important design tasks which can serve both business consortia and policy makers in their endeavors to establish data spaces in a certain domain.

As the establishment of data space is still an emerging topic, the scientific body of knowledge is still in its infancy when it comes to understanding design options in detail and the factors that influence the design. Moreover, knowledge does not exist about the growth and adoption of data spaces.

Consequently, there is a strong need for research in different disciplines such as computer science, information systems, and management science but also for interdisciplinary studies.

References

1. European Commission. (2020). *The European data market monitoring tool*. European Commission.
2. Otto, B. (2015). Quality and value of the data resource in large enterprises. *Information Systems Management*, 32(3), 234–251. <https://doi.org/10.1080/10580530.2015.1044344>
3. Capiello, C., Gal, A., Jarke, M., & Rehof, J. (2020). *Data ecosystems: Sovereign Data Exchange among organizations (Dagstuhl Seminar 19391)*. <https://doi.org/10.4230/DagRep.9.9.66>.
4. Gelhaar, J., Groß, T., & Otto, B. (2021). A taxonomy for data ecosystems. In: *Hawaii International Conference on System Sciences 2021* (pp. 6113–6122). University of Hawai'i at Manoa, Hamilton Library.
5. Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society*, 8(1). <https://doi.org/10.1177/2053951720982012>
6. European Commission. (2020). *A European strategy for data*. European Commission. Available online at <https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:52020DC0066&from=EN>. Accessed 4/25/2021.

7. Franklin, M., Halevy, A., & Maier, D. (2005). From databases to dataspace. *ACM SIGMOD Record*, 34(4), 27–33. <https://doi.org/10.1145/1107499.1107502>
8. Halevy, A., Franklin, M., & Maier, D. (2006, June 26–28). Principles of dataspace systems. In: G. Gottlob & J. van den Bussche (Eds.), *Proceedings of the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems—PODS '06, the twenty-fifth ACM SIGMOD-SIGACT-SIGART symposium, Chicago, IL, USA* (pp. 1–9). ACM Press.
9. Regulation of the European Parliament and of the Council on European data governance of 2020.
10. IDS Association & Fraunhofer. (2021). *GAIA-X and IDS*. International Data Spaces Association; Fraunhofer-Gesellschaft.

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

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

