gaia-x
Hub Germany

What Is a Data Space?

Definition of the Concept Data Space

Gefördert durch:

Bundesministerium
für Wirtschaft
und Klimaschutz

aufgrund eines Beschlusses
des Deutschen Bundestages

White Paper 1/2022
September 2022

**Dr. Abel Reiberg, Dr. Crispin Niebel
und Peter Kraemer**

acatech
DEUTSCHE AKADEMIE DER
TECHNIKWISSENSCHAFTEN

About the Series

White Papers of the Gaia-X Hub Germany are intended for discourse and the exchange of ideas. They reflect the opinion of the authors and not necessarily that of the Gaia-X Association or any other institution of the Gaia-X ecosystem.

Authors

acatech – National Academy of Science and Engineering
Peter Kraemer, Head of Technological Sovereignty and Industrial Value Creation, Head of Gaia-X Hub Germany
Dr. Crispin Niebel, Scientific Advisor
Dr. Abel Reiberg, Scientific Advisor

Publisher

Gaia-X Hub Germany
c/o acatech – Deutsche Akademie der Technikwissenschaften e.V.
Karolinenplatz 4
80333 München

Recommended Citation

Reiberg, Niebel, Kraemer (2022), 'What is a Data Space?', Gaia-X Hub Germany, White Paper 1/2022.

Table of Contents

Summary

Although the term data space has been used increasingly in recent years, there is often a lack of clarity or disagreement about its meaning. This white paper therefore aims to contribute to a common understanding of the data space phenomenon. To this end, among other things, a **definition** is developed according to which data spaces can be understood as: **A federated, open infrastructure for sovereign data sharing based on common policies, rules, and standards.** This definition and its foundations will be further explained using the example of the Mobility Data Space (MDS).

The white paper is divided into **four sections**:

Section one first discusses the measures that are currently being taken to realize data spaces. These include both regulatory measures, such as the *Data Governance Act*, and distributive measures, such as the Gaia-X funding competition of the German Federal Ministry of Economic Affairs and Climate Action (BMWK). It also discusses the many goals that can be pursued with data spaces. These include cross-sectoral goals - such as economic prosperity, informational self-determination, political sovereignty, technical security, etc. - as well as sector-specific goals, such as increased security in the area of mobility, strengthening innovation in the area of medicine, or improved quality and accessibility of education.

Section two then looks at the perspectives from which data spaces are currently being viewed. As it turns out, important aspects of data spaces are coming to the fore from all relevant perspectives (such as technical, legal or economic perspectives). These include the potential of data spaces to realize efficiencies in data exchange, to promote data sovereignty, and to enable new business models. Despite the prevalence of different perspectives, relevant definitions of the term data space share many commonalities. Section two of this white paper identifies these commonalities and then brings them together to form the above definition of the term data space.

In section three, the concept of data space is then further concretised on the basis of the definition developed. For this purpose, the components, function, properties, limits and genesis of data spaces are discussed. Data exchange is considered to be the most important function. The components that are considered include the participants - providers, recipients and intermediaries of the data exchange. Among other things, the decentralisation, transparency and interoperability of data spaces are outlined as important properties.

Finally**, in section four,** the concept of data space is distinguished from other common terms such as data ecosystem and data marketplace. The former is explained as a broad term that, in addition to data exchange, also includes upstream and downstream processes of data collection and processing and the actors involved. The latter one is understood as a term that encompasses the commercial exchange of data, as it can also be realised through or within data spaces.

1. Measures Taken and Objectives in the Area of Data Spaces

1.1. Measures Taken for the Realisation of Data Spaces

The term data space has become a battle cry of the digitalisation movement in recent years. This is also reflected in a multitude of measures with which the creation of data spaces is being promoted. These include, in particular, numerous funding programmes and projects with which actors from business, science and politics promote the practical implementation of data spaces.

Gaia-X was initiated in 2019 and is an international initiative to create the framework conditions for interoperable, open and sovereign European data spaces. Initially launched as a Franco-German partner project, the project now unites over 350 organisations from all parts of the world (Gaia-X Association, 2022a). In Germany, supported by a funding competition of the Federal Ministry of Economic Affairs and Climate Action, the implementation of concrete application examples and lighthouse projects is currently underway.

Within the framework of the *Digital Europe Programme of the Connecting Europe facility* and *Horizon 2020*, investments are also being made in the development of European data spaces. For example, data spaces are to be developed for the areas of: Mobility, Manufacturing, Agriculture, Cultural Heritage, Health, Media, Public Administration, Public Procurement, Law Enforcement and Finance (European Commission, 2022c).

These initiatives are just a few of many examples of measures of a **distributive nature** that are being taken to create data spaces. These are flanked by numerous measures of a **regulatory nature** that are currently being implemented at all levels of political systems, but especially at the European level. These are intended to contain negative developments in the data economy, especially in the area of large platforms, and at the same time promote positive developments in the area of data spaces.

For example, the *Data Governance Act* created new Europe-wide legal foundations for data exchange and defined uniform requirements for data intermediaries.

The *Data Act* is also intended to promote the European exchange of data by guaranteeing users access to data obtained by and through them.

The *Digital Markets Act* and the *Digital Services Act*, in turn, contain requirements for platform providers with particular market power and ensure that negative effects of market concentrations in the data economy are curbed.

What these measures have in common is that they create a favourable environment for the development of data spaces. These in turn allow for the pursuit of a variety of economic, technological and societal goals.

1.2. Goals of the Realisation of Data Spaces

The development of data spaces promises to take the data economy to a higher level of development. In this way, existing problems of a data economy shaped by platforms can be solved and opportunities of a data economy shaped by data spaces can be exploited. The number of goals that can be pursued is long and reflects the many forms of implementation, areas of application and sectors in which data spaces can be realised. However, many of the **cross-sectoral goals** mentioned in the literature can be simplified into three areas:

Value Creation

The current value of the data economy in the overall economy is already large and still growing (Micheletti et al., 2020, p. 13). There is already a large amount of data distributed between a large number of companies. This represents an enormous potential for profitable exchange. To realise this, however, a trustworthy and transparent data exchange infrastructure is needed. Data spaces enable consumers and companies to exchange data sovereignly. This means that they allow the participants to determine the cornerstones of the exchange themselves, such as partners, duration, purpose, etc. At the same time, their open and decentralised structure ensures that open competition is guaranteed and market concentration is avoided. Negative consequences such as information asymmetries and loss of control are prevented, while value creation is made possible. In this way, it can be ensured that companies, including in particular small and medium-sized enterprises, participate in the data economy and remain competitive. Europe's economic independence is thus strengthened.

Sovereignty and Self-Determination

Users of data - both companies and consumers - benefit greatly from today's possibilities of communication and data transfer. However, they often do not have the control they would like to have. Data collected about them or generated by them is often beyond their direct control. This can pose existential dangers for organisations and individuals alike. Data spaces promise a remedy here as well. The decentralised structure allows consumers and businesses to determine their own contractual partners and conditions - so that data is transferred according to the users and their preferences regarding the scope, duration, subject matter and partners of the exchange. A situation in which conditions for data exchange that do not correspond to the preferences of the users are unavoidable due to a lack of alternatives is excluded. The sovereignty of the market participants is promoted through the control gained and the accompanying build-up of competence. They have both the ability and the possibility to determine with whom they share which data and under which conditions. Ultimately, this also strengthens the European digital economy and with it Europe's geopolitical sovereignty.

Efficient Use of Resources

New information technologies enable newer productive exchange processes. However, it is becoming apparent that exchange on the basis of centralised infrastructures is not always the most efficient option. Nowadays, data is often generated in a decentral manner. Data spaces make it possible to deal with this initial situation efficiently by leaving data at the point of origin and only merging it when and to the extent that it is actually necessary. They thus provide a framework for a range of technological developments to efficiently use existing resources - be they the resources of consumers, companies, entire national economies or the natural resources of our planet.

The above-mentioned goals are only a few of the many cross-sectoral goals that are currently being pursued with data spaces. In addition to these, there are a multitude of other **sector-specific goals**, which are particularly evident with regard to the respective implementation examples.

For example, increased value creation, more efficient use of resources and strengthened sovereignty in almost every sector can lead to greater acceptance and demand on the user side and greater diversity and quality on the provider side. This will in turn have specific positive consequences according to the core functions of the application or economic sector concerned: in the health sector, an effective and fast fight against health problems; in the education sector, for example, better access to learning opportunities; or in the construction sector, a fast and efficient implementation of projects.

Infobox 1

Mobility Data Space Example

The MDS is an example of a data space that contributes to the achievement of both fundamental and sector-specific goals in the field of mobility.

It creates the basis for exchange between different organisations in the mobility sector - including, for example, car manufacturers, providers of map services and weather services, municipalities, road authorities and mobility service providers. For example, data on local hazards collected by car manufacturers will be shared with map service providers in order to be able to display appropriate warnings in vehicles. Similarly, weather data collected by weather service providers is shared with mobility service providers in order to provide weather-dependent recommendations for mobility offers. Such applications in turn make it possible to achieve various goals:

For example, by providing the vehicles they produce with effective warning systems or the mobility services they offer with weather-specific recommendations, the respective companies increase the quality of their products or services. This, in turn, is reflected in customer satisfaction, turnover and ultimately the companies' profits. This strengthens the competitiveness of the companies as well as that of the entire business location.

The sovereignty of those involved is also strengthened. Providers of data, in the above example weather services that provide weather data or car manufacturers that provide sensor data, are put in a position to negotiate and agree on all essential contractual details for the exchange of data. Conformity with central legal requirements such as the General Data Protection Regulation is ensured. On a technical level, the decentralised structure of the MDS in particular ensures that intransparency, information asymmetries and an imbalance of negotiating power are avoided. Since this enables each participant in the MDS to participate in the data exchange in a sovereign manner, corresponding competences for the use and trade of data, which are of increasing importance, can also be built up.

The data exchange itself is efficient. The MDS relies on the system architecture of the International Data Spaces Association (ISDA) and thus on a decentralised structure. Thanks to this decentralised structure, capacities for the transmission, storage and processing of data are used in a particularly demand-oriented manner, especially for the participants. Avoidable costs for setting up and using the corresponding capacities are saved. In addition, the data exchange also leads to a more efficient use of resources - both on the part of the providers and on the part of the users. For example, thanks to corresponding use cases, data on the availability of parking spaces or on road damage is not collected multiple times by different companies and organisations, but is exchanged and combined - to the extent that it generally makes sense. Users can save resources with the realised services (such as information services on parking spaces and road damage), for example by reducing travel times and energy consumption accordingly. Finally, there are also benefits for society as a whole, for example because the use of raw materials and environmental pollution are minimised.

Furthermore, specific goals of the mobility sector are also achieved. Danger warnings ensure greater safety in traffic. Precisely tailored mobility offers increase comfort and strengthen mobility as a whole. And information about available parking spaces reduces unwanted traffic in search of a parking space.

Overall, a look at concrete implementation projects shows that a variety of goals can be pursued with data spaces. In order to achieve these, however, it is important to become aware of the different perspectives on data spaces and to base a common understanding on them.

2. A Common Understanding of Data Spaces

2.1. Different Perspectives of the Phenomenon Data Space

From a **technical perspective,** data spaces are often described as a means of data integration. Data that is available in one place and needed elsewhere does not necessarily have to be transferred completely and permanently. Rather, it is sufficient to integrate distributed data to the extent and for the duration that is necessary. The technical work on data spaces raises awareness of this fact and shows how - thanks to a federated architecture - data can be used as efficiently as possible.

From an **economic perspective**, data spaces are seen in particular as a means of economic exchange. The fact that data is distributed between entrepreneurs, companies and consumers represents an economic potential. The economic perspective highlights the fact that incentives must be given so that companies can participate in the exchange of data and realise economic and geopolitical sovereignty as well as social welfare.

With a view to the self-determination of the individual, which comes to the fore from a **legal perspective**, data spaces can ultimately be understood as a means of sovereignty. Until now, users - companies as well as consumers - have often had little control over the data that is collected by, through and about them. Data is often scattered across different service providers and is largely beyond the control of users. The concept of data space also serves to develop approaches to give users more control. It shows, for example, that opportunities for participation and obligations of neutrality are of great importance for certain intermediaries. Corresponding regulations and requirements are currently becoming part of various legal systems. In the case of Gaia-X, it is in particular the requirements of the EU and the EU member states that must be taken into account, including, for example, the *General Data Protection Regulation* and the *Data Governance Act*. These requirements play a major role in determining the form that the respective data spaces can take.

Each of the specific perspectives mentioned is therefore important or even necessary to contribute specific insights into the data space phenomenon. However, it is equally important to establish a common basic understanding of data spaces. This is the only way to ensure that the individual goals can be pursued equally. It therefore makes sense to work on common theoretical foundations that can guide the practical realisation of data spaces. One contribution to this is the development of a definition that can be widely agreed upon.

2.2. A Definition of the Term Data Space

So far, there is still a lack of a generally accepted definition of the term data space. Even in central texts (Regulation (EU) 2021/694), a definition is often avoided. If a definition is given, it usually remains either vague or differs from previous approaches. This leaves the task of developing a concrete definition that can be widely agreed upon.

One way to do this is to identify promising approaches and bring them together. A comparison of the definitions of central actors such as the EU Commission, the Federal Ministry for Economic Affairs and Climate Action, the Big Data Value Association, IDS, etc. shows that certain concept clusters repeatedly appear. These reflect basic principles of data spaces and therefore lend themselves as the core of a widely agreeable and concrete definition of the term data space:

Sharing

Almost all definitions describe the exchange of data - referred to as "sharing" (Gaia-X Association, 2022b; Nagel et al., 2021), "access" (European Commission, 2022a, p. 2), "integration" (Curry, 2020) and "exchange" (Nagel et al., 2021; Otto, 2022, p. 2) - as the core function of a data space. It is important to note that the exchange of data can take different forms. On the one hand, it can take the form of a transfer, after which data is available to both exchange partners. On the other hand, it can take the form of access to data that is limited in time, content or scope. Data spaces enable their participants to provide and receive data with a high degree of control, or to shape and contribute to the exchange that comes about in this way.

Governance

As the majority of definitions, such as those of Gaia-X, IDS, dataspaces4.eu and the EU Commission, emphasise, there is no anarchy in data spaces. Rather, there are commonly agreed rules, described as "standards and rules" (Gaia-X Association, 2022b), "shared principles and standards" (Nagel, 2022), "shared policies and rules" (Nagel et al., 2021, p. 7), "shared policies, standards, rules" (Joining Forces in 'Team Data Spaces', 2022, p. 1) and "rules of administrative and contractual nature" (Bodenkamp, 2022, p. 4). In other words, there are always governance mechanisms in data spaces that guarantee the rights of the participants and a smooth exchange. These governance mechanisms are not imposed on the participants, but are based on democratic legitimacy - international norms, national and supranational laws or rules and standards agreed between the participants of the data space.

Control/Data Sovereignty

The idea of retaining control over one's own data and its use is an important basis for giving participants trust and is in line with certain rules and values - especially in the EU - regarding data sovereignty. For example, the Gaia-X website states "Data sovereignty and trust are essential for the working of data spaces" (Gaia-X Association, 2022b). Open DEI, for example, points out how important it is for data spaces to have "building blocks ensuring data sovereignty" (Nagel et al., 2021, p. 23).

Openness

With another cluster of terms, central definitions of the term data space explain that they are to be understood as "open" (European Commission, 2022a, p. 3; Nagel et al., 2021, p. 5) in the

sense of "easy access" (European Commission, 2020, p. 4; Nagel et al., 2021, p. 7) and an "open number of organisations/individuals" (European Commission, 2022a, p. 3). It is precisely the openness of data spaces that ensures that undesirable developments that are evident in the data economy, particularly in the area of large platforms, do not continue. Data spaces enable a large number of participants in different roles (as providers, recipients or intermediaries) to enter into exchange and, if necessary, also into competition and where lock-in effects are also avoided.

Federation/Interoperability

Data spaces are consistently described in existing definitions as "federated", (Joining Forces in 'Team Data Spaces', 2022, p. 1; Nagel et al., 2021, p. 7), "loosely integrated" (Curry, 2020, p. 10) and "interoperable" (Gaia-X Association, 2022b; Nagel, 2022, p. 1). The IDS, for example, summarises: "data spaces are a federated network" (Nagel, 2022). Data spaces therefore do not form a centralised, but a decentralised whole. A data space forms a unit that can, however, be linked to or embedded in other data spaces. This is made possible by using open standards for the most part. This enables cross-data-space exchange and continuous expansion of data spaces.

In summary, exchange, openness, federation and governance are important basic principles of data spaces in the eyes of central actors. They should therefore not be missing from a general definition of the term data space. A definition that takes up all principles and should therefore be suitable for consensus would be the following.

Data Spaces are defined as: A federated, open infrastructure for sovereign data sharing, based on common policies, rules and standards.

Datenräume sind definiert als: Eine föderierte, offene Infrastruktur für souveränen Datenaustausch, die auf gemeinsamen Vereinbarungen, Regeln und Standards beruht.

Infobox 2

Mobility Data Space Example

One example of the implementation of the above-mentioned principles of data spaces is the Mobility Data Space. As a lighthouse project, it shows how sharing, governance, sovereignty, openness and interoperability can be realised in practice:

In the case of the Mobility Data Space, the exchange of data - the sharing - takes place, as explained above, between actors in the mobility sector, including, for example, mobility service providers, car manufacturers, suppliers, etc. The MDS also allows for a sectoral exchange of data. In addition, the MDS enables cross-sectoral participation of all actors who offer or want to use mobility-relevant data, including for example weather services or insurance companies. The participants can exchange a wide variety of data, such as data on road damage, storms, charging infrastructures, etc. This data is not stored in the MDS, but rather the MDS allows the participants to exchange data directly.

In this exchange, the participants retain full control / sovereignty over their data. This means that they can decide at any time which data they want to exchange and under which conditions. For this purpose, data providers and data recipients define the corresponding details in a peer-to-peer contract. Beforehand, all participants must undergo a standardised certification process. This ensures that only the designated participants receive access to the data and that they are trustworthy. Furthermore, secure data exchange is ensured in particular via the connectors of the MDS, the software component through which participants gain access to the MDS. This ensures that the exchange cannot be interrupted, viewed or manipulated.

The MDS is generally open to all interested organisations - it is thus characterised by openness. So far, about 200 organisations are participating in the MDS, including large as well as medium and small enterprises. Until 2024, participation in the MDS is free of charge. This makes it easier for smaller actors in particular to become part of the data sharing community without incurring extensive costs and thus to generate benefits from value creation potential.

The same conditions of participation apply to all participants - they can participate equally in the further development of the MDS. For effective governance, the neutral supporting organisation was set up as a non-profit. All revenues of *Datenraum Mobilität GmbH* will be invested in the further development of the data space.

It is particularly relevant for the further development of the MDS that it is characterised by interoperability. The architecture of the MDS is based on the IDS reference architecture. It is therefore interoperable with all data space initiatives developed under the IDS umbrella. In addition, the MDS will be designed to be Gaia-X compatible. This means that the data spaces currently being developed on the basis of the Gaia-X architecture will also be compatible with the MDS. This means that participants in the MDS will be able to participate in a large number of other data spaces without any further implementation effort. In this way, easy scalability and great growth potential of the MDS will be achieved.

3. The Concept “Data Space” in Detail

A common definition can form the core of a common understanding about data spaces. Ideally, however, the common understanding should go beyond a common definition and, for example, provide answers to the questions of what the function, what the form and what the properties of a data space are.

In the following section, an increasingly dense description of the concept of data space will be created on the basis of the core questions addressed and thus the above definition will be expanded step by step. In doing so, a large number of those characteristics and properties of data spaces described in the literature will be taken up and systematically related.

3.1. What Are the Components of a Data Space?

As explained above, data spaces can be regarded as spaces for sharing data. Due to the decentralised structure of data spaces, this exchange often does not take the form of transmission, but rather (however limited) access to data, which in turn can take place directly as well as indirectly, for example in the case of compute-to-data.

From the principle that data spaces are spaces of data exchange, it follows that necessary components of a data space are the participants and objects of data exchange. Participants are basically the providers and recipients of the exchanged data. Participants are also intermediaries between providers and recipients (so-called intermediaries). The roles mentioned can be assumed by organisations as well as individuals. In principle, the same persons/organisations can take on different roles (providers, recipients or intermediaries) in different exchange processes. For example, a school that is a participant in an educational data space could receive data - such as data on available instructional content - on the one hand, and provide data - such as data on the effectiveness of the instructional content ultimately used - on the other.

In a broader understanding, those practices, arrangements and institutions that enable, facilitate or control the exchange of data can also be considered part of a data space. Examples include software protocols, standards and contracts. For example, connectors can also be considered part of a data space. These are client-side applications that allow participation in data exchange with a maximum of control over the data concerned.

3.2. What Is the Function of a Data Space?

A data space enables the exchange of data - this is its core function. Sub-functions of a data space are functions that enable, facilitate or control the provision, receipt and transmission of data (the logical components of the exchange). These include, for example, functions to catalogue or identify participants and objects of the exchange.

In the case of Gaia-X, for example, such functions are performed by Gaia-X Federation Services. These include services for identification management, cataloguing of data and services, secure data exchange, ensuring rule compliance and portal design. Identification management, for example, is based on the concept of Self-Sovereign-Identity and ensures that only those participants who are supposed to take part in the exchange actually do so.

3.3. What Are the Properties of a Data Space?

On the one hand, data spaces have constituent properties - properties that make them a data space. On the other hand, they have desirable properties - properties that data spaces must have in order to qualify as a "European data space" or Gaia-X compliant data space, for example. In the following section, we will first discuss the constituent and then the desirable properties of a data space.

Scope: A data space necessarily has a spatial and temporal scope. It should encompass a certain amount of data and a certain number of participants, each over a certain period of time. For example, an exchange between only two clinics does not constitute a data space for medical data, even if the exchange is permanent. Similarly, an exchange between all clinics in any country does not constitute a data space if the exchange occurs only once.

Decentralisation: A data space necessarily has a degree of decentralisation. This is measured by the actual number of participants compared to the possible number. Absolute decentralisation exists, for example, if all exchange processes take place via different participants in the data space. Minimal decentralisation exists if all exchange processes take place via a single entity. If, for example, the cataloguing of databases is carried out by a single provider, minimal decentralisation is given for this partial service. The degree of decentralisation can vary from data space to data space. The widespread view is that a strong decentralisation of data spaces is desirable. Positive effects that can result from this include more competition and (consequently) more innovation.

Federation/interoperability: In order to enable the decentralisation of data spaces on the one hand and to ensure that as many data space participants as possible can exchange information with each other on the other, the most important initiatives for data spaces rely on a federated structure. Data spaces are designed to be interoperable. Exchange processes can thus be realised beyond the boundaries of a data space and participants who have already fulfilled the prerequisite for participation in a data space can participate in other data spaces with a minimum of effort. This is usually realised through harmonised processes, formats and semantics/ontologies that enable an effective and smooth exchange of information.

Transparency: A data space must have a certain degree of transparency. This transparency is measured by the amount of information available about what is happening within the data space, to actual and potential participants, and to affected and unaffected third parties. In order for the data space to function, it must generally be recognisable to participants what data (in what quantity and quality) is provided or received, for what purpose and for what duration. A certain degree of transparency is therefore absolutely necessary for the functioning of a data space. When a desirable or acceptable level is reached is a normative question that will be the subject of future opinion-forming processes.

Sovereignty: Insofar as a data space presupposes a (voluntary) exchange of data, a data space must have a certain degree of sovereignty for the participants. Participants should be able to decide whether to participate in the exchange according to their preferences regarding content, scope, purpose, duration and participants. This also requires that actual or potential participants are not discriminated against, i.e. unreasonably hindered in their ability to exchange, by the actions of other participants or by the design of the relevant practices, arrangements and facilities of the data space.

Trust: Linked to transparency and sovereignty is the notion of trust. The more transparency there is, be it in terms of data quality or evidence of data, and the greater the sovereign control over the data, the more participants can trust the data space. In addition, clear compliance with legal rules and the regulatory environment further strengthens trust in the data space. Accordingly, one obstacle to building trust among European companies is the concern about unlawful access to data by authorities not subject to EU jurisdiction (European Commission, 2022b, p. 21). In the case of Gaia-X, this concern is addressed in particular by the labelling system. Another decisive factor for trust in data spaces is their technical security and resilience. Advantages in data security can be realised, for example, by choosing a decentralised, federated architecture (such as in Gaia-X) instead of a centralised architecture. When data is stored in centralised systems, there is always a risk that a security breach will be far more serious than in a decentralised, federated infrastructure. The more trust participants have in the data space, the more willing they will be to make their data available. A minimum level of trust among participants is necessary for a data space to function.

3.4. What Are the Boundaries of a Data Space?

Participants in a data space will usually define boundaries of the space in question. This may be explicit or implicit and is usually reflected in practices, agreements and/or facilities of the data space. For example, if a data space is described as an education data space, it can be assumed that it should include relevant educational institutions. However, sharing across data space boundaries can or should be possible.

Projects such as Gaia-X aim to create a federation of data spaces. In order to realise such a federation, the practices, agreements and facilities of a data space should be compatible with those of other data spaces to such an extent that basic functionalities for an exchange across spatial boundaries are given.

3.5. When Is a Data Space Fully Developed?

In individual publications, the distinction between development stages of data spaces is suggested (European Commission, 2022a, p. 12). The question arises as to how such development stages are to be delineated and how a transition is to be conceptualised. Conceptually, for example, two paths of data space development can be distinguished: on the one hand, a data space can develop without its boundaries being redefined. For example, a data space on patient data could gradually expand as an increasing number of patients decide to share data. On the other hand, an expansion is conceivable that goes hand in hand with the redefinition of boundaries. For example, it is possible that a data space in which patient data is exchanged will change into a data space for health information in general. This can happen by gradually offering other data (possibly from new participants) in the data space on patient data. This can also happen by merging the data space for patient data with a data space for other health data by realising a complete merging of the data spaces.

4. Demarcation of the Data Space Concept from Other Concepts

Knowing what a data space is also means knowing what a data space is not. Concepts such as "data ecosystem", "data market", "data intermediary" are used again and again in discourses about data spaces. In some cases, they are used synonymously. In the following section - in order to further contribute to a common understanding of data spaces - the concept of data space will be distinguished from other central concepts:

4.1. Data Ecosystems

There is no generally accepted definition for the concept of data ecosystem, nor for the concept of data space and all other concepts mentioned in this section. In some cases, the term data ecosystem is used synonymously with data space. Relevant texts on the term often refer to a "data life cycle" (S. Oliveira et al., 2019, pp. 15-16) of data in the sense of a sequence of different process steps from the collection of data to integration, analysis and the like. In line with the Bitkom proposal (2022), a data ecosystem can be regarded as a higher-level unit compared to the data space. In addition to the exchange processes that are the content of the data space, this unit includes upstream and downstream processes of obtaining and processing data.

4.2. Platforms

It is widely believed that the existence of a platform is closely linked to the existence of multi-sided markets (Rochet & Tirole, 2003). Such multi-sided markets are also to be expected in the context of data spaces. For example, within a data space for medical data, a two-sided market is conceivable that is served by an intermediary, which on the one hand receives personal health data and passes it on to service providers for medical recommendations and on the other hand receives medical recommendations and passes them on to the providers of health data. As the literature on platforms shows, multi-sided markets have certain effects (network effects) that promote a concentration of relevant markets (Evans & Schmalensee, 2005). As explained above, an absolute concentration or a minimal decentralisation of the relevant exchange processes contradicts the concept of data space. However, the extent to which a gradual concentration/centralisation of the data space to a few participants can be desired or accepted is an open question. The relevant documents on Gaia-X and similar projects make it clear that a low level of concentration is desired in these and that measures that can counteract concentration (e.g. measures for interoperability, transparency, non-discrimination, etc.) are to be taken by participants in the corresponding data spaces (Gaia-X Association, 2022a).

4.3. Data Intermediation Services

As data intermediary services stand between providers and receivers of data. Intermediation does not necessarily require that data intermediaries themselves receive data or access to it. Organisations and individuals whose services enable, facilitate or control the exchange of data

can also be regarded as data intermediaries. In this sense, data intermediary service providers can be considered a possible or even a necessary component of a data space. The EU's *Data Governance Act* contains comprehensive requirements for providers of such services. These include neutrality and transparency obligations. For example, providers are not allowed to process the data themselves, but only the corresponding metadata, and this is only permitted for the purpose of improving their own services.

4.4. Data Marketplace

In a broader sense, the term data marketplace refers to any permanent location for commercial data trading (Abbas et al., 2021). In a narrower sense, the term is used for online platforms that enable data trading (Spiekermann, 2019). In this sense, data marketplaces are to be regarded as a form of data brokerage services.

4.5. Data Lake

The concept of data lake was developed in particular to distinguish it from the concept of data warehouse (Woods, 2011). The latter term usually refers to a data collection in which the data structure or schema is defined in advance. The former term, however, refers to a data collection in which data of different forms (structured to unstructured) are included without prior adaptation. The advantage of storing data in data lakes is, for example, that a large part of the data of an organisation can be included without great effort, in order to process it further if necessary and to use it for new purposes.

Both the term data warehouse and the term data lake usually stand for data storage within individual organisations. The concept of the data space, on the other hand, refers to the exchange of data, especially between organisations, as well as the components, prerequisites, etc. of this exchange.

Bibliography

Abbas, A. E., Agahari, W., van de Ven, M., Zuiderwijk, A., & de Reuver, M. (2021). *Business Data Sharing through Data Marketplaces:* A Systematic Literature Review. Journal of Theoretical and Applied Electronic Commerce Research, 16(7), 3321–3339. https://doi.org/10.3390/jtaer16070180

Bodenkamp, J. (2022). *Digital Europe & Data Spaces* [Vortrag]. Digital Europe Data Spaces Calls.

Curry, E. (2020). *Real-time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems.* Springer International Publishing. https://doi.org/10.1007/978-3-030-29665-0

European Commission. (2020). *Mitteilung der Kommission an das Europäische Parlament, den Rat, den Europäischen Wirtschafts- und Sozialausschuss und den Ausschuss der Regionen—Eine Europäische Dantenstrategie* (COM(2020)66). https://eur-lex.europa.eu/legal-content/de/ALL/?uri=CELEX:52020DC0066

European Commission. (2022a). *Commission Staff Working Document on Common European Data Spaces* (SWD(2022) 45 final). https://ec.europa.eu/newsroom/dae/redirection/document/83562

European Commission. (2022b). *Impact Assessment Report Accompanying the document Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act)* (SWD(2022) 34 final). https://ec.europa.eu/newsroom/dae/redirection/document/83524

European Commission. (2022c). *The Digital Europe Program*. digital-strategy.ec.europa.eu. https://digital-strategy.ec.europa.eu/en

Evans, D. S., & Schmalensee, R. (2005). *The Industrial Organization of Markets with Two-Sided Platforms* (Working Paper Nr. 11603). National Bureau of Economic Research. https://doi.org/10.3386/w11603

Gaia-X Association. (2022a). *About Gaia-X*. Gaia-X.eu. https://gaia-x.eu/what-is-gaia-x/about-gaia-x/

Gaia-X Association. (2022b). *Data Spaces. Gaia-X.eu.* https://www.gaia-x.eu/what-is-gaia-x/data-spaces

Joining Forces in 'Team Data Spaces'. (2022). *dataspaces4.eu.* https://dataspaces4.eu/wp-content/uploads/Joining-Forces-in-Team-Data-Spaces.pdf

Micheletti, G., Cattaneo, G., Glennon, M., La Croce, C., & Mitta, C. (2020). *The European Data Market Monitoring Tool* (d2.9 final study report; S. 101). Europäische Kommission.

Nagel, L. (2022). *The Magic of Data Spaces Now.* internationaldataspaces.org. https://internationaldataspaces.org/the-magic-of-data-spaces-now/

Nagel, Lars, & Lycklama, D. (2021). Design Principles for Data Spaces—Position Paper. International Dataspaces Association. https://design-principles-for-data-spaces.org/

Otto, B. (2022). A federated infrastructure for European data spaces. *Communications of the ACM*, 65, 44–45. https://doi.org/10.1145/3512341

Regulation (EU) 2021/694 of the European Parliament and of the Council of 29 April 2021 establishing the Digital Europe Programme and repealing Decision (EU) 2015/2240, 166 OJ L (2021). http://data.europa.eu/eli/reg/2021/694/oj/eng

Rochet, J.-C., & Tirole, J. (2003). Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, 1(4), 990–1029. https://doi.org/10.1162/154247603322493212

S. Oliveira, M. I., Barros Lima, G. de F., & Farias Lóscio, B. (2019). Investigations into Data Ecosystems: A systematic mapping study. *Knowledge and Information Systems*, 61(2), 589–630. https://doi.org/10.1007/s10115-018-1323-6

Spiekermann, M. (2019). *Datenmarktplätze—Plattformen für Datenaustausch und Datenmonetarisierung in der Data Economy.*

Woods, D. (2011). *Big Data Requires a Big, New Architecture.* Forbes. https://www.forbes.com/sites/ciocentral/2011/07/21/big-data-requires-a-big-new-architecture/?sh=5df0b1f81157