bitkom

Data Spaces and Data Ecosystems – First Explainer and Current Status

September 2022

Content

1 What is in this paper

1.1 Content

- **We offer a first classification and outline of conceivable details for a design of data spaces and data ecosystems.**
- **In this consequence, we discuss to what extent the terms data space and data ecosystem could be distinguished from each other.**
- **There are already numerous frameworks and initiatives for decentralised, sovereign data spaces and data ecosystems, which are at different stages of implementation, some of which we elaborate on.**
- **Currently, we observe numerous open sets of questions that pop up again and again in discussions on data spaces and data ecosystems. These include definition of terms, (de)centralisation of roles and processes, as well as questions of economic efficiency and sustainability.**

1.2 Context

- Public and private stakeholders worldwide are increasingly recognising the opportunities and potential of Big Data, IoT or smart data systems. However, many stakeholders still find it difficult to transfer these concepts and thus realise concreate added business value in their own organisations.
- This has also been taken up by policymakers who are increasing incentives for companies and society through initiatives such as the German government's data strategy or the EU data strategy. There are also concrete obligations to share data, for example within the framework of the proposed Data Act, which further suggests a discussion of data spaces as a possibly suitable instrument.
- In the discourse on data spaces and data ecosystems in general, there is a rather immense proliferation in terms of definitions, architecture, incentives, and roles. This partly leads to hesitancy and lack of trust on the part of users, confusion for policymakers and regulators, and the fact that some (small and medium-sized) companies do not yet see concrete added value and/or do not know where they can participate.
- Many stakeholders are not yet completely certain about the added value of data spaces in achieving efficiency gains and developing new business models. Whether and to what extent which data spaces will be convincing as enablers for the (European) data economy will have to be shown on the basis of acceptance in the market, scalability, as well as specific problem fit per sector and across sectors.
- It seems useful to distinguish between political and economic goals for the creation of data spaces. These do not have to contradict each other, but they are not automatically congruent in terms of scope and functionality.

63%

of enterprises in Germany do not share data with other enterprises. (Bitkom Research, 2022)

2 Why is this relevant

2.1 History

Data has already been shared for a long time, for example on the basis of existing supplier relationships, to fulfil information claims, for research projects, or to develop new business models between interested market participants.

2.2 Policy

Also due to the ongoing legislative and funding policy initiatives on data sharing - including the Data Act Proposal[1] or Common European Dataspaces[2] - there is a great interest in data spaces as enablers for data sharing. In the case of the Data Act Proposal, data sharing obligations vis-à-vis users and third parties in certain circumstances as well as data sharing with public sector bodies are particularly worth mentioning. With Common European Dataspaces, the EU Commission has launched a programme of legislative and other measures for many sectors to increase the availability of data. Also relevant is the Data Governance Act (DGA), which introduces obligations of conduct for data intermediation services, which can also be applied to data spaces and data ecosystems.

Only 3%

of enterprises in Germany are active in data spaces. (Bitkom Research, 2022)

2.3 Status Quo

Data platforms, marketplaces, and other business models for data sharing and joint, secure interaction already exist on the market. Despite these perspectives, it is true in many sectors that the potential of data sharing has not yet been exhausted. Although initiatives for data spaces and data ecosystems also exist in some sectors, they hardly do so in others. Where they do exist, state of affairs and dissemination are quite different.

2.4 Why is this the case

Today, data sharing is often not as systematic and efficient as it could be. For example, data formats have to be adapted, interfaces validated, or contracts negotiated, which stands in the way of broad and scalable use. Data spaces and data ecosystems for data exchange often have difficulties in winning over a significant/critical mass of participants or in establishing a fair negotiating position between operators and participants which would help to gain trust.
Other reasons are legal uncertainty in dealing with data, a lack of standards, trust between the actors involved that needs to be built up and, at the same time, the pending question of incentives for data sharing ("Is it worthwhile sharing data?").

1 COM (2022) 68, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2022%3A0068%3AFIN
2 In its 2020 Data Strategy (COM(2020) 66 final), the EU Commission announced, among other things, to promote common European data spaces in strategic sectors and areas of public interest, https://ec.europa.eu/commission/presscorner/api/files/attachment/862109/European_data_strategy_en.pdf.pdf

2.5 Takeaway

In this respect, the challenge is to **realise network effects** on the one hand, to **decentralise bargaining power** on the other, while at the same time to create an **incentive to operate and participate** in the necessary structures.

3 Classification attempt

We believe that **incentive-compatible data ecosystems are a key factor for the success of data spaces**. This means that data sharing must be complemented by analysis, combination, and exploitation of data to create socially and economically optimal added value. To structure this, the two terms **data space** and **data ecosystem** could be helpful.

A quarter (26%) of enterprises in Germany have never heard about data spaces, around the same amount (28%) have heard about it but are unsure what it means. 27% know roughly what it's about, and only 16% feel able to explain what a data space is. (Bitkom Research, 2022)

3.1 Data Space

In its Data Strategy, the EU Commission introduced a „**Common European Dataspace**" to the discussion, which is to be complemented by **sectoral data spaces**.[3] The „Common European Dataspace" is meant to have the following key characteristics:[4]

> - A secure and privacy-preserving infrastructure to pool, access, share, process and use data.
> - A clear and practical structure for access to and use of data in a fair, transparent, proportionate and/non-discriminatory manner and clear and trustworthy data governance mechanisms.
> - European rules and values, in particular personal data protection, consumer protection legislation and competition law, are fully respected.

The Gaia-X Architecture Document offers the following thoughts:[5]

> A Data Space is a virtual data integration concept defined as a set of participants and a set of relationships among them, where participants provide their data resources and computing services.
>
> Data Spaces have the following design principles:
> - data resides in its sources;
> - only semantic integration of data and no common data schema;
> - nesting and overlaps are possible;
> - spontaneous networking of data, data visiting and coexistence of data are enabled.
>
> Within one Data Ecosystem, several Data Spaces can emerge.

Even on the basis of two conceptual definitions of two key players in the European data economy - without taking any details into account - no clear picture emerges of what constitutes a data space and what does not.

Anecdotally, the impression nowadays tends to be that ***n* actors have at least *n+1* data space definitions**. This is not a problem per se, but it does pose communication challenges in terms of getting new actors excited or developing initiatives.

We as Bitkom, with experience in finding compromises in the membership, have also not succeeded in finding a sufficiently precise, conclusive definition of terms in the short term. In the following, we therefore want to outline certain functions that could be helpful without venturing a definition.

3 COM(2020) 66 final, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52020DC0066
4 SWD(2022) 45 final, p. 2, available via https://digital-strategy.ec.europa.eu/en/library/staff-working-document-data-spaces
5 Gaia-X Architecture Document 21.12 Release, p. 86, available via https://docs.gaia-x.eu/technical-committee/architecture-document/. This definition has not made it into the 22.04 version of the document.

In contrast to licensing of data or commissioned data processing between often only two or a few actors that is common in the market, a data space could enable data sharing between any number of actors on an ad-hoc basis. The reduction of search and transaction costs with simultaneous easy scalability is sometimes the most important differentiating factor compared to many previous models for the exchange of data (1:1 or 1:n).

Schematically, a data space could have the following functions:

- Exchange: compatible technical interfaces for the exchange of (meta-)data.
- Openness & trust: Basic openness for all interested parties, identification and security of participants and transactions
- Contract logic: initiation, conclusion, and management of contracts
- Self-governance: Documented governance of actors and processes (how decisions are made)
- Enforcement: reliable governance of data with appropriate agreements
- Security and resilience: state of the art

Schematically, a data space could enable public bodies, companies and users to exchange data. Both can be providers (= suppliers) and recipients (= consumers) of data.

As can be seen in Figure 1, when distinguishing between metadata and data, a data space could A) serve as a coordination point and exchange platform for metadata, while the actual data is shared decentrally, or B) in addition to being a coordination point and exchange platform for metadata, it could also accept and forward the actual data in a centralised manner.

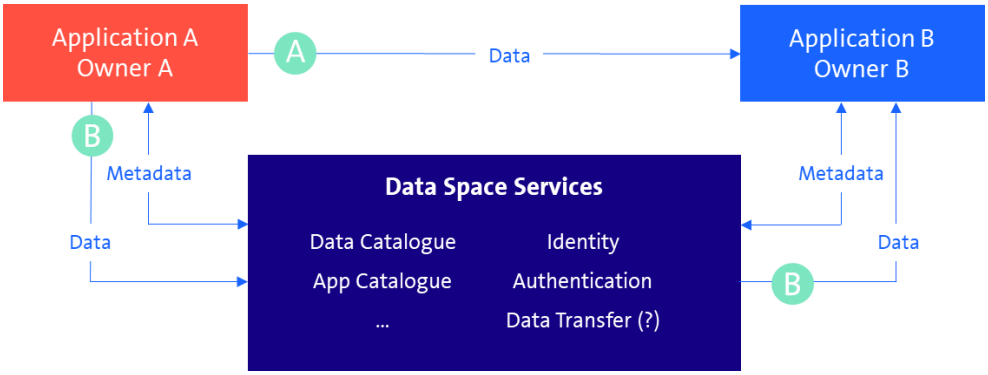

Figure 1: Data space Flowchart data and metadata. Version decentralised (A), centralised (B).

3.2 Data Ecosystem

The Gaia-X Architecture Document offers the following thoughts:[6]

> Data Spaces form the foundation of Data Ecosystems. In general, Data Ecosystems enable Participants to leverage data as a strategic resource in an inter-organizational network without restrictions of a fixed defined partner or central keystone companies. For data to realize its full potential, it must be made available in cross-company, cross-industry Ecosystems. Therefore, Data Ecosystems not only enable significant data value chain improvements, but provide the technical means to enable Data Sovereignty. Such sovereign data sharing addresses different layers and enables a broad range of business models that would otherwise be impossible. Trust and control mechanisms encourage the acceleration of data sharing and proliferate the growth of Ecosystems.

Bitkom is a member of the Gaia-X European Association for Data and Cloud AISBL and emphasises, not only but also against this background, that a **data ecosystem** could and should go beyond the concept of data sharing in the data space and complement it.

We consider a data space as the "inner core" of a data ecosystem.

A data ecosystem should include various services, such as:

- services as a preliminary step (e.g. generation and preparation) for sharing data in a data space, and
- services as a processing step after receiving data from a data space in order to generate concrete added value from it.

As a result, such services could be **key incentives** for economic participation in a data space and help to go **beyond data sharing as a primary purpose**, supporting data space participants in working with data, increasing efficiency through new insights, or

6 Gaia-X Architecture Document, Release 21.12, p. 69, available via https://docs.gaia-x.eu/technical-committee/architecture-document/. A similar but enhanced definition can be found in the 22.04 version of the document.

developing data-driven business models. **This can play a significant role in the acceptance and growth of data spaces.**

Building on the distinction in Figure 1 between decentralised and centralised data exchange (not metadata), data could furthermore be prepared or post-processed during both processes by services of a data ecosystem, whereby both A) a decentralised data exchange or B) a centralised data exchange via data space is conceivable, as shown in Figure 2.

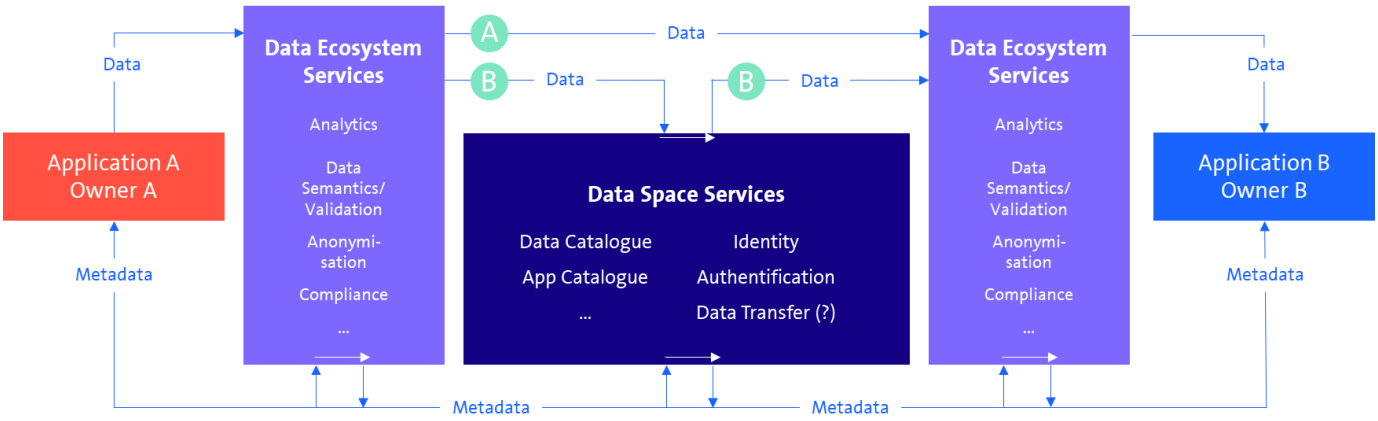

Figure 2: Data Space & Data Ecosystem Flowchart Data and metadata. Version decentralised (A) / centralised (B).

3.3 Why is gets more complicated

These questions become even more important against the backdrop of the DGA: to what extent is cost-covering or even profitable operation of a data ecosystem compatible with the DGA regulations for data intermediation services? A key question here will be which actors in the context of a data space are subject to which types of neutrality obligations, remuneration restrictions, and other behavioural obligations. In other words, the challenge in this context is which models of economic action might be possible and potentially promising in the context of a DGA-regulated data space.

The EU Commission's Data Act proposal also provides for certain documentation obligations for operators of data spaces.

In addition, when sharing data - also in data spaces - further legal review requirements apply in individual cases, for example related to:

- competition and antitrust law (incl. protection of trade and business secrets, pricing)
- copyright and other IP rights
- data protection law,

which have not yet been (conclusively) answered.

4 Current Status

Numerous frameworks and initiatives for decentralised sovereign data spaces and data ecosystems already exist and are at various stages of implementation. Most of them tend to be in the spectrum between preliminary study, conceptualisation, and prototype, but to our knowledge rarely in (enterprise) production operation.

When discussing and participating - as far as possible - in such initiatives, the "push/pull discussion" can be particularly relevant. By this we mean whether there is primarily a push effect from research projects or primarily concept-based initiatives, or instead there is primarily a pull effect from market participants who see added value in the project and are not afraid of concrete efforts. In the long term, it will be important to translate use cases into business cases in order to operate data spaces at least on a cost-covering basis after start-up funding, which is absolutely desirable from a funding policy perspective, for which the existence of frameworks is probably helpful but only conditionally necessary or even sufficient.

In the following, we have begun to compile a non-exhaustive collection of different horizontal initiatives and multi-sectoral approaches. There are many others and we do not make any recommendation or assessment of them. It will be true here and elsewhere that the frameworks, use cases, implementations, initiatives operate at extremely different levels and hierarchies and have different levels of maturity. What they all have in common is that they are moving forward dynamically.

4.1 Selected Acteurs/Documents/Projects

Acteurs

- Data Spaces Business Alliance is a consortium of Gaia-X AISIBL, BDVA, Fiware and IDSA, with the goal of developing a common framework for data spaces.
- Gaia-X AISBL, and respective hubs
- Big Data Value Association (BDVA)
- FIWARE Foundation (Fiware), and respective hubs
- International Data Spaces Association (IDSA), and respective hubs
- Mydata.org, and respective hubs
- Various other (commercial) providers of concepts/platforms for trading/sharing data.
- The Data Spaces Support Centre is meant to support data spaces in creating an interoperable environment for data sharing and to enable re-use and further use of data between different sectors. As of August 2022, it has not yet started to operate. Overview Documents
- Staff Working Document of the EU Commission Common European Dataspaces depicts important elements of the status quo in policy and funding policy regarding common European data spaces.

- Gaia-X Data Spaces Position Papier depicts various relevant use cases from Gaia-X industry verticals under the umbrella of the Gaia-X Data Spaces Business Committee.
- SITRA Fair Data Economy Rulebook in particular contains various blueprints for legal, technical, and other types of agreements.
- IDS RAM 4.0 is a reference architecture of IDSA, which is meant to enable data sharing between organisations.
- Open Dei Design Principles for Data Spaces by IDSA in particular contains design principles, high level building blocks and a governance proposal for data spaces. Horizontal Projects
- Eclipse Data Space Connector Projekt is a project under the umbrella oft he Eclipse Foundation that develops a connector technology, which is being / will be used in particular by Catena-X initiative and which is/will be compliant with the Gaia-X as well as IDS standards.
- Different Connectors, based on IDS standard, are in particular used in the context of the Mobility Data Space.
- Simpl is an announced publicly funded project of the EU Commission under the Digital Europe Programme, within which a *Smart Middleware Platform* is meant to be developed. This again is meant to enable *Cloud-to-Edge-Federations* and to support *Common European Dataspaces* in particular.

Sectoral Projects

- Next to projects / initiatives with horizontal character, there are various sectoral data space initiatives in different EU member states. By example and non exhaustively, these include such, which assemble in the Gaia-X Hub Germany, the initiative Catena-X[7], or the Mobility Data Space[8] (in which Bitkom members are involved in bodies) as well as similar public funding projects, and projects without public funding.

The used terminology does not necessarily align with our offered understanding in chapter 3.

5 Collection of open questions

Here, we document several questions, which come up again and again in discussions around data spaces and data ecosystems. Not exhaustively, these include:

5.1 Definitions

- What are the precise definitions for data spaces and data ecosystems? Which characteristics do they have?
- Where exactly is the tipping point at which “ordinary” data transfer becomes a data space or a data ecosystem? Do database systems with multiple actors (already) fall under the term data space or data ecosystem?

7 Catena-X is an initiative supported by German Ministry of Economics and Climate Protection as one of its projects in the program Digitalisierung der Fahrzeughersteller und Zuliefererindustrie.
8 The mobility data space is an initiative supported by the German Ministry for Digital and Transport as one of its AI projects.

- Dataspaces are sometimes discussed in the context of cloud initiatives – is there a necessary link (what is the role of on-prem, hybrid, cloud-native, edge-services) or are data spaces actually infrastructure-independent?
- When stakeholders talk about data spaces, do they actually mean data ecosystems?

5.2 Centrality vs. Decentrality

In many scenarios, it is considerable to assign certain tasks to specific actors. Given that, certain centralisation tendencies can be observed. In light of the following questions, it is to be evaluated to what extent one or several "central" actors could assume such tasks.

- Do we need a central certification authority for participants, services, infrastructure and/or data as a trusted entity? Which organisation(s) can exercise such role? Can such trust services be performed decentrally or via self-declaration?
- Do comprehensive (ex-ante) certification needs create entry barriers for small and medium enterprises, that is for such who are meant to profit the most from data spaces?
- Is the actual data sharing performed decentrally between provider and recipient or is there a "clearing house" / transfer body? What about metadata? This discussion, which is also about trust, takes place in a similar manner in the context of data trusts.
- Is there a central catalogue with available data? If yes, how is it curated?
- Are contract processes / services administered centrally or decentrally?

56% of enterprises in Germany expect that data spaces would lead to more data sharing, 54% expect that a more secure data exchange would be possible. Roughly half (49%) would use data spaces in addition to current offers. 45% wish for more support of data spaces by policymakers. (Bitkom Research, 2022)

5.3 Economics and Sustainability

- Which economic viability can individual actors/enterprises expect in data spaces or data ecosystems?
- Who runs a data space? One or several private actors, a public actor?
- Which use cases can be turned into business cases? Which problem exactly can be solved using a data space?
- Which (if any) challenges and opportunities have to be overcome by existing frameworks and initiatives before they enter into production?

6 Next Steps

Bitkom will continue to work on data spaces and data ecosystems in an open-minded, constructive, and critical manner to develop, strengthen, and demand the practical relevance of data spaces and data ecosystems.
Bitkom and/or several of its member companies are currently doing this, e.g., in the context of different working groups of the Gaia-X European Association for Data and Cloud AISBL, in discussions and consultations with the German Gaia-X Hub or various other projects and initiatives.
Next to these strategic workstreams, Bitkom monitors, advises, works on virtually all legislative data policy proposals on German and European level and equally provides expertise and knowledge exchange for technical profiles, such as data scientists, data engineers, data stewards, data strategists.

Bitkom represents more than 2,000 companies of the digital economy, including 1,900 direct members. Through IT- and communication services only, our members generate a domestic turnover of 190 billion Euros per year, including 50 billion Euros in exports. Members of Bitkom employ more than 2 million people in Germany. Among the members are 1,000 small and medium-sized businesses, over 500 startups and nearly all global players. They offer a wide range of software technologies, IT-services, and telecommunications or internet services, produce hardware and consumer electronics, operate in the sectors of digital media or are in other ways affiliated to the digital economy. 80 percent of the companies' headquarters are located in Germany with an additional 8 percent each in the EU and the USA, as well as 4 percent in other regions. Bitkom supports the digital transformation of the German economy and advocates a broad participation in the digital progression of society. The aim is to establish Germany as globally leading location of the digital economy.

Publisher:
Bitkom e.V.
Albrechtstr. 10 | 10117 Berlin
Germany

Contact Person:
David Schönwerth | Policy Officer Data Economy
T 030 27576-179 | d.schoenwerth@bitkom.org

Responsible Bitkom Working Group:
WG Data Policy & Data Spaces

This publication is a convenience translation from the original version in German language: Bitkom, Datenräume und Datenökosysteme – Erste Einordnung und aktueller Stand, September 2022.