

Schweihoff, Julia; Lipovetskaja, Anzelika; Jussen-Lengersdorf, Ilka; Möller, Frederik

**Article — Published Version**

## Stuck in the middle with you: Conceptualizing data intermediaries and data intermediation services

Electronic Markets

*Suggested Citation:* Schweihoff, Julia; Lipovetskaja, Anzelika; Jussen-Lengersdorf, Ilka; Möller, Frederik (2024) : Stuck in the middle with you: Conceptualizing data intermediaries and data intermediation services, *Electronic Markets*, ISSN 1422-8890, Springer Berlin Heidelberg, Berlin/Heidelberg, Vol. 34, Iss. 1, <https://doi.org/10.1007/s12525-024-00729-9>

This Version is available at:

<https://hdl.handle.net/10419/315795>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<http://creativecommons.org/licenses/by/4.0/>



# Stuck in the middle with you: Conceptualizing data intermediaries and data intermediation services

Julia Schweihoff<sup>1,2</sup> · Anzelika Lipovetskaja<sup>2,3</sup> · Ilka Jussen-Lengersdorf<sup>1,3</sup> · Frederik Möller<sup>2,3</sup>

Received: 22 March 2024 / Accepted: 21 August 2024 / Published online: 27 September 2024  
© The Author(s) 2024

## Abstract

Data intermediation services are a novel class of services mitigating various aspects of inter-organizational data sharing. Facilitating data intermediation services is a pivotal activity that leverages the benefits of the data economy. These services are usually provided by so-called data intermediaries, who organize data sharing between one or multiple data provider(s) and data consumer(s). Against the background of prominent legislation, these data intermediaries (e.g., data marketplaces or data trusts) receive increasing attention in research, practice, and politics. For instance, the Data Governance Act regulates data intermediation services, prompting organizations sharing data to comply with various regulations, including registration with government bodies or acquiring a trusted label. However, to date, little research has been done about the nature of data intermediation services, motivating us to tackle this issue in a two-fold strategy. First, we derive a visualization of a data intermediary and concept subsuming its potential functions. Second, we explore archetypical configurations of co-existing services of data intermediaries. With this, we contribute to understanding the data intermediaries as a concept and enrich this understanding with more detailed service configurations.

**Keywords** Data intermediaries · Archetypes · Data sharing · Taxonomy

**JEL Classification** A1

## Introduction

“Data intermediaries represent a new policy lever to navigate the challenges of the growing data ecosystem.” — Flanagan and Warren (2022)

Sharing data is increasingly relevant for companies, even though they are typically hesitant to share them due to a range of obstacles (Fassnacht et al., 2023; Lefebvre et al., 2023). For instance, those obstacles are a lack of available technological infrastructure and the challenge of complying with applicable regulations (Candelon et al., 2024; Ebel

et al., 2021). Concerns about sharing competitively relevant information, violating applicable law, or unclear costs also stand in contrast to the advantages, such as better economic efficiency through the use of shared data (Wernick et al., 2020). Companies benefit from new sources of revenue or data, access to new data sources and customers, cost reductions (Bitkom (Ed.), 2023), or the “discovery of new insights,” “faster, more accurate decision making,” “increased prediction accuracy,” “optimized process efficiency and coordination,” and “increased innovation” (Hoffman et al., 2019). In addition to benefits for companies, data sharing also offers advantages on a social level (Hoffman et al., 2019). In particular, the COVID-19 pandemic has shown how relevant it can be to share data (World Economic Forum (Ed.), 2021) to achieve the speed necessary to fight a globally spreading disease. In Germany, for example, there are already several data trustees who collect health data and make it available for research while complying with privacy-preserving measures (German Centre for Cancer Registry Data, 2013; Trusted Third Party of the University Medicine Greifswald, 2024; Westdeutsche

---

✉ Julia Schweihoff  
julia.schweihoff@tu-dortmund.de

<sup>1</sup> TU Dortmund University, Joseph-Von-Fraunhofer-Str. 2-4, 44227 Dortmund, Germany

<sup>2</sup> TU Braunschweig, Mühlenpfordtstr. 23, 38106 Brunswick, Germany

<sup>3</sup> Fraunhofer ISST, Speicherstr. 6, 44147 Dortmund, Germany

Biobank Essen [WBE], 2019). Expediting data-sharing activities is widely perceived as a value-creating mechanism, which shifted from simply solving ad hoc existing problems to a strategic activity for value creation (e.g., Wixom et al., 2023). These potentials stretch numerous industries, such as agriculture (e.g., Wysel et al., 2021), manufacturing (e.g., Tang et al., 2024), or supply chain management (e.g., Culot et al., 2024) and are usually associated with, but not limited to, saving costs and saving time through optimization. There are use cases in the energy sector where data intermediaries (DIs) use data to share consumers' energy data with financial service providers (Data Sharing Coalition (Ed.), 2020).

The core of successful data sharing is the presence of *trust* in those who receive or mediate the data (Capgemini Research Institute (Ed.), 2021; Ebel et al., 2021; Flanagan & Warren, 2022). One approach to tackling these issues is to use DIs that offer data providers a range of data management services pertaining to data sharing, such as *supporting legal compliance*, *anonymization*, or *data security* (Jussen et al., 2023b). These services offered by DIs are data intermediation services (DISs) that facilitate the many facets of inter-organizational data sharing on behalf of data providers and data consumers (Carovano & Finck, 2023; Schweihoff et al., 2023). DIs provide these DISs and receive increasing attention as they are at the heart of the (developing) European Data Economy (Joint Research Centre, 2023), prompting a “data intermediary hype” (Richter, 2023 p. 458). They promise a host of benefits, such as reducing transaction costs and enabling novel data-sharing transactions that were previously impossible (Martens et al., 2020). In practice, DIs are instantiated in a plethora of variants (e.g., Ditfurth & Lienemann, 2022; Micheli et al., 2023), and two prominent examples are *data marketplaces* and *data trusts* providing alternating complementary services and functions for data provider(s) and data consumer(s). The *data marketplace* aims for data monetization, giving data providers, typically, a forum to offer their data for financial compensation (e.g., Bergman et al., 2022; Jussen et al., 2024a). *Data trusts* also organize data sharing but explicitly offer services around *data pseudonymization* and *anonymization* and facilitating legally compliant data provision and consumption (e.g., Lipovetskaja et al., 2024; Radosevic et al., 2023; Stachon et al., 2023).

A key motivator for research and practice to explore DIs is the Data Governance Act (DGA), which entered into force in 2023 and regulates how DIs need to be operationalized. While the DGA is not the sole foundation for this article, it spurs the need for organizations to be aware of DIs. The spectrum of DISs provided by DIs is not easily demarcated since the DGA defines them as all services that “establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other, through technical, legal or other

means, including for the purpose of exercising the rights of data subjects in relation to personal data” (European Commission 2022, L152/19), which, seemingly, includes a wide range of potential DIs. Based on this spectrum of DISs and the potential variants of DIs, we can position them in the context of information systems (IS) research. DIs are (at least) two-sided *digital platforms* that organize and facilitate inter-organizational data sharing by providing DISs between data providers (supply side) and data consumers (demand side) (Ditfurth & Lienemann, 2022). In their nature as digital platforms, they generate and partially orchestrate platform ecosystems (e.g., Hein et al., 2020), which, in the context of inter-organizational data sharing, are seen as *data ecosystems* (Möller et al., 2024; Oliveira & Lóscio, 2018; Oliveira et al., 2019).

Given the external pressure from legislation and the potential of DIs and DISs for research and practice, we find it problematic that there is no “clear” understanding of what the umbrella terms “data intermediary” and “data intermediation services” include. For example, Micheli et al. (2020) identify six types of DIs — the *data marketplace*, *data unions*, *personal information management systems*, *data cooperatives*, *data trusts*, and *data sharing pools*. Ditfurth and Lienemann (2022) propose the *data marketplace*, *(industrial) data platforms*, and *data trusts* while highlighting, and we agree with that, that DIs are an evolving and emerging field and classifications are likely not robust at this point. To counteract the wide range of potential implementations and the accompanying conceptual blurriness, we propose to complement conceptual clarity and understanding of the nature of DIs. We do not aim to provide another classification of DIs but tackle the issue of clarifying DISs in a two-fold strategy. First, we consolidate the field of DIs by exploring DISs independently from the specific DI implementation. Second, we explore the configurations of non-exclusive (i.e., non-dependent on the specific DI implementation) DISs. To summarize, we pursue the two following research questions:

**Research question (RQ1):** What are the generic services of data intermediaries?

**Research question (RQ2):** What are archetypical configurational patterns for data intermediation services?

To achieve our goal, we follow the taxonomy-building method of Kundisch et al. (2021). Building a taxonomy is useful for our purpose since it accommodates the dual nature of our results. Taxonomies are frequently used to span the complete *playing field* of an object of interest (e.g., as design options, Möller et al., 2021). We combine this by employing cluster analysis to map out configurational patterns of co-existing DISs. This strategy is frequently used in IS research to explore reoccurring configurational patterns (e.g., Weking et al., 2020 or Fischer et al., 2020).

We structure the paper as follows. Following the “Introduction,” we outline the fundamentals of data intermediaries. The “Research design” section explains the research design, which consists of a taxonomy-building approach using both literature and empirical objects as the underlying data sources. In the “Findings: conceptualizing data intermediation services” section, we present both the generic and idealized understanding of DIs and present the clustered patterns of DISs. The “Contributions, limitations, and outlook” section concludes our paper with an overview of the contributions and limitations of our research as well as an outlook on further research.

## Data intermediaries

The term “intermediary” originates from the Latin words “inter” (engl. “between”) and “medius” (engl. “middle”), referring to an actor that connects multiple sides such as data providers and data consumers (Merriam-Webster, 2024b). The term was first mentioned in the late eighteenth century (Merriam-Webster, 2024b) and appeared in the literature in 1878, where it was quoted in a medical context (see Corrold, 1878). The concept of the “digital” intermediary became popular with the beginning of the digitalization of markets and so-called *electronic market intermediaries* (J. Bailey, 1996; Bhargava et al., 1999; Chrusciel & Zahedi, 1999; Sherer, 1995). For instance, these intermediaries provide support with online shopping, insurance, or the distribution of music (Bouwman et al., 2005; Kim & Talalayevsky, 2005; Moloney, 2005).

Engaging in data sharing poses significant barriers for organizations (e.g., Fassnacht et al., 2023; Legenvre & Hameri, 2023). For instance, Jussen et al. (2023b) report on a set of tensions in inter-organizational data sharing, contrasting *incentives* (e.g., data monetization) and *barriers* (e.g., data valuation) that organizations are tasked to find responses to. These barriers are manifold and span a problem space around organizations navigating the fear of data misappropriation, a lack of technical and cultural know-how, and legal and privacy concerns when sharing data (Cichy et al., 2021; Fassnacht et al., 2023). DIs are among the prominent solutions addressing some of these issues since they are tasked with taking over parts of the inter-organizational data-sharing process from organizations (Schweihoff et al., 2023). Initial approaches by the EU to overcome the barriers to the use of DIs include the offer of certification as a trustworthy DI (European Commission, 2024b).

Given the range of potential DI instances, we discuss the conceptual underpinnings and definitions proposed in the literature. In Table 1, we show some prominent definitions of DIs as examples, which reflect the versatility of the term and demonstrate its importance in various domains. Most definitions characterize the DI as a *mediator*, organizing inter-organizational data sharing as either “in the middle” between providers and consumers or as all entities supporting data-sharing activities (e.g., Janssen & Singh, 2022). For example, a DI is defined in the DGA “as neutral third parties that connect individuals and companies on one side with data users on the other” (European Commission, 2023). Micheli et al. (2023) propose a broader definition, proposing DIs as entities providing stakeholders access to data. Ichihashi (2021) discusses DIs as mitigating agents exchanging

**Table 1** Definitions of DIs (findings from literature, highlighted by the authors)

Definition	Context	Author
“is a role that <b>facilitates the use of data</b> for other actors.”	Data ecosystems	Oliveira et al., (2019, p. 609)
“a <b>mediator</b> between those who wish to make their data available, and those who seek to leverage that data. The intermediary works to <b>govern the data</b> in specific ways and <b>provides</b> some degree of confidence regarding <b>how the data will be used</b> ”	Policy	Janssen and Singh (2022, p. 2)
“Data intermediaries <b>offering services that connect the different actors</b> have the potential to contribute to the efficient pooling of data as well as to the <b>facilitation of bilateral data sharing</b> . Specialised data intermediaries that are <b>independent</b> from both data holders and data users can have a facilitating role in the emergence of new data-driven ecosystems independent from any player with a significant degree of market power.”	EU legislation	European Commission (2020, p. 16)
“Data intermediaries can take many forms; but what they share is a primary purpose of <b>facilitating and managing data relations</b> between data rights holders (such as people or businesses), depending on the parties’ relationships, intentions and resources.”	Practical report	Flanagan and Warren (2022, p. 9)
“Data intermediaries for more inclusive data governance allow a broader range of stakeholders to <b>access, control and share data</b> , and <b>support data subjects and data holders</b> in deciding the purposes for which data is managed, as well as facilitating the exercise by data subjects of their rights over personal data, with the likely effect of producing further benefits from the same data and thus redistributing data value (social, public or private) across more actors and/ or society”	Policy	Micheli et al., (2023, p. 11)

personal data between consumers and upstream organizations, like Google and Facebook, to receive rewards. Open data intermediaries are another type of data intermediary. These are defined as “third-party actors who provide specialized resources and capabilities to (i) enhance the supply, flow, and/or use of open data and/or (ii) strengthen the relationships among various open data stakeholders” (Shaharudin et al., 2023, pp. 14–15). However, due to their relation to open data, they do not fall under the DGA.

Figure 1 shows the interaction of a “traditional” electronic market intermediary (Sarker et al., 1995). The intermediary takes on an organizing role and receives remuneration for the transaction, which takes place between provider and consumer. The right-hand side shows a model we adapted for a data-sharing scenario using a DI. In this model, in addition to the rewards for the services and transactions, the intermediary provides data and services in return for a fee. It is the intermediary’s task to develop a relationship of trust to motivate the actors to share data. The data producers and data consumers do not need to exchange data via the intermediary. The intermediary ensures that the appropriate actors are matched with each other and then negotiates the conditions of data sharing among themselves. They make use of possible further services of the intermediary, such as the provision of an infrastructure necessary for the data exchange.

The task of a data intermediary is to provide services for data sharing (Carovano & Finck, 2023; European Commission, 2022), so-called DISs. DIs occupy the *middle seat* between different actors when it comes to data sharing (Schweihoff et al., 2023). For example, the DI can provide a platform (Eisenmann et al., 2011), connect different actors (Kioses et al., 2007), or take on various trustworthy tasks in the data-sharing process (Otto et al., 2019). The DGA governs the responsibilities, rights, and duties of a DI (European Commission, 2022) and defines DISs as follows: “‘data intermediation service’ means a service which aims to establish commercial relationships for the purposes of data sharing between an undetermined number of data subjects and data holders on the one hand and data users on the other,

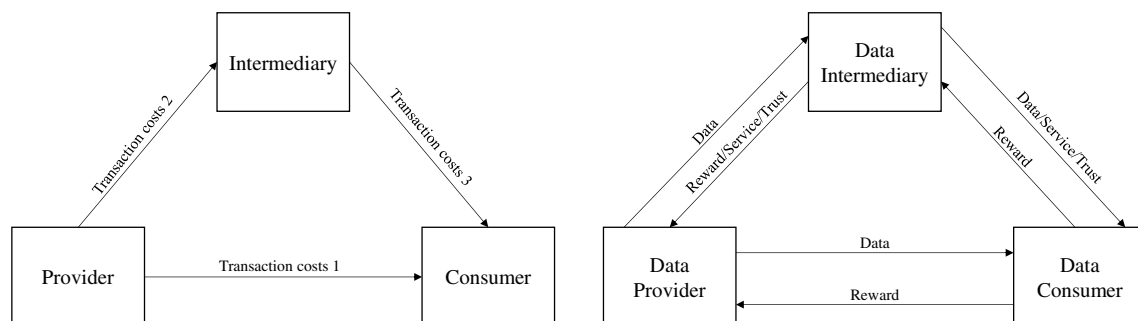
through technical, legal or other means, including for the purpose of exercising the rights of data subjects in relation to personal data” (European Commission, 2022, L 152/19).

In summary, the DI acts as an authority mediating between various stakeholders who wish to share data (Flanagan & Warren, 2022; Janssen & Singh, 2022; Oliveira et al., 2019). It is subject to legal requirements such as the DGA, which stipulates the form in which the DI may offer its services (European Commission, 2022; Richter, 2023). DIs should support stakeholders in implementing the data-sharing process (Micheli et al., 2023). Recent practice reports (e.g., Flanagan and Warren (2022), Micheli et al. (2023) or Centre for Data Ethics and Innovation (2021)) on the topic of DI show that the phenomenon of “data intermediaries” is currently of great interest to practitioners and researchers. Even if the DGA sets out a specific direction for intermediaries and many articles already discuss the DGA in combination with DIs (e.g., Carovano & Finck, 2023 or Richter, 2023), we think it is necessary first to understand the status quo of DISs and DIs in order to give suitable implications for companies.

## Research design

### Taxonomy design

We report the taxonomy design procedure using the method of Kundisch et al. (2021), who propose 18 steps organized into six phases (i.e., the DSR phases of Peffers et al., 2007). They are *identify problem and motivate* (1), *define objectives of a solution* (2), *design and development* (3), *demonstration* (4), *evaluation* (5), and *communication* (6). Each consists of a set of lower-threshold steps for taxonomy design, which we outline below. Taxonomies are a suitable method to “conceptualize phenomena based on their dimensions and characteristics” (Kundisch et al., 2021, p. 421), which fits our research goal. We chose the method of Kundisch et al. (2021) because it is an update to the de facto standard (see also Möller et al., 2022) and the proven method for taxonomy design in IS research of Nickerson et al. (2013).



**Fig. 1** The interrelationship between the actors involved in data sharing (left, Sarker et al. (1995); right, own adaption on data sharing)

**Table 2** Overview of taxonomy design iterations

Iteration	Approach	Method	Result
1	Conceptual-to-empirical	Systematic literature review	Meta-dimensions, initial characteristics
2	Empirical-to-conceptual	Empirical objects	Taxonomy
3	Empirical-to-conceptual	Empirical objects	Revised taxonomy
4	Empirical-to-conceptual	Empirical objects	Final taxonomy

**Phase 1 (identify problem and motivate)** Taxonomy builders are advised to *specify the observed phenomenon clearly (step 1)*. In our case, this refers to *DISs*, which we have characterized above (see the “**Data intermediaries**” section). **Step 2** outlines the *target user group(s)*, which are researchers exploring different facets of *DISs* as well as managers facing issues in how to use and utilize *DISs*. Next, taxonomy designers should formulate the intended purpose of the taxonomy (**step 3**). We pursue a dual mission in using the taxonomy. First, we aim to span a conceptual playing field outlining comprehensive characteristics of *DIs* and their services. Therefore, we are required to conceptualize an abstraction that reflects the sum of all possibilities for *DISs* existing in a *DI* (Weber, 1949). Second, we tackle it using the taxonomy to extract a set of archetypes, that is, configurations of data intermediation services that occur together by means of cluster analysis.

**Phase 2 (define objectives of a solution)** The second phase demarcates the solution space for the taxonomy and aims to capture it in a *meta-characteristic*, illustrating the “beginning” and “the most comprehensive characteristic that will serve as the basis for the choice of characteristics in the taxonomy” (Nickerson et al., 2013, p. 343). Our meta-characteristic is as follows: *comprehensive classification of data intermediation services*. In terms of explicating ending conditions, we follow standard practice and adopt subjective and objective ending conditions proposed by Nickerson et al. (2013). Additionally, we set evaluation goals to demonstrate the importance of data intermediaries in the taxonomy adequately.

**Phase 3 (design and development)** In phase 3, we oscillate between different strategies (empirical-to-conceptual or conceptual-to-empirical) to build the initial taxonomy and its subsequent iterations (see Fig. 4). For this, we initially reviewed the literature (iteration 1) and then analyzed real-life cases (iterations 2–4), extracting knowledge about data intermediaries (see Table 2). Based on our findings, we classified the services.

**Literature review** We followed standard practice in taxonomy design to conceptualize initial characteristics and

meta-dimensions by performing a **systematic literature review** (Kundisch et al., 2021) (see Fig. 2). Subsequently, we pursued a concept-centric literature review strategy (Webster & Watson, 2002). Since we did not constrain our *DI* definition to a specific type, such as data marketplaces, we defined keywords broad enough to identify a wide corpus of literature. We searched for “intermediary” and omitted the prefix “data” to be as comprehensive as possible and added complementary keywords from our experience that define types of *DI* or could potentially be used synonymously, namely “trustee,” “trusted third-party,” and “marketplace.” We combined these keywords with both “data sharing” and “data exchange” to accommodate for the interchangeable nature of these terms (Jussen et al., 2023a). This strategy was necessary because of the heterogeneity of terms used in data sharing. For example, Nwachock A Koul & Morin (2018) find that “The **data sharing** process is managed by a **data marketplace**, a **trusted third-party** handling the market participants’ request and managing the agreements between them.”<sup>1</sup> We searched for literature in established databases and focused on peer-reviewed conference proceedings and journals (Levy & Ellis, 2006). Specifically, we explored AISEL (indexes IS conferences and some journals), Scopus (extensive database indexing almost all relevant IS journals), IEEE Xplore (indexing additional literature from adjacent fields, such as engineering), and Web of Science (indexing broad array of literature). We first checked the 871 hits against our selection criteria. These included that the language used in the paper was German or English and that the abstract and title dealt with *DIs*. Next, we filtered the literature and included only those papers that reported or discussed a type of *DI* — such as data marketplaces (e.g., Agahari et al., 2021; Figueredo et al. (2020); or Demchenko et al. (2019)) or data trustees (e.g., Azkan et al., 2022). This allowed us to reduce the sample to 256 papers. After eliminating all duplicates, the 201 papers remained in the sample. In the second iteration of the literature search, we checked the content of the papers to investigate whether they addressed *DIs* to reduce the sample to 60 papers. In the final third iteration, we examined the papers to determine whether they included the service offering of *DIs* to a final selection of 48 papers.

<sup>1</sup> Highlighting of relevant terms by the authors (bold).

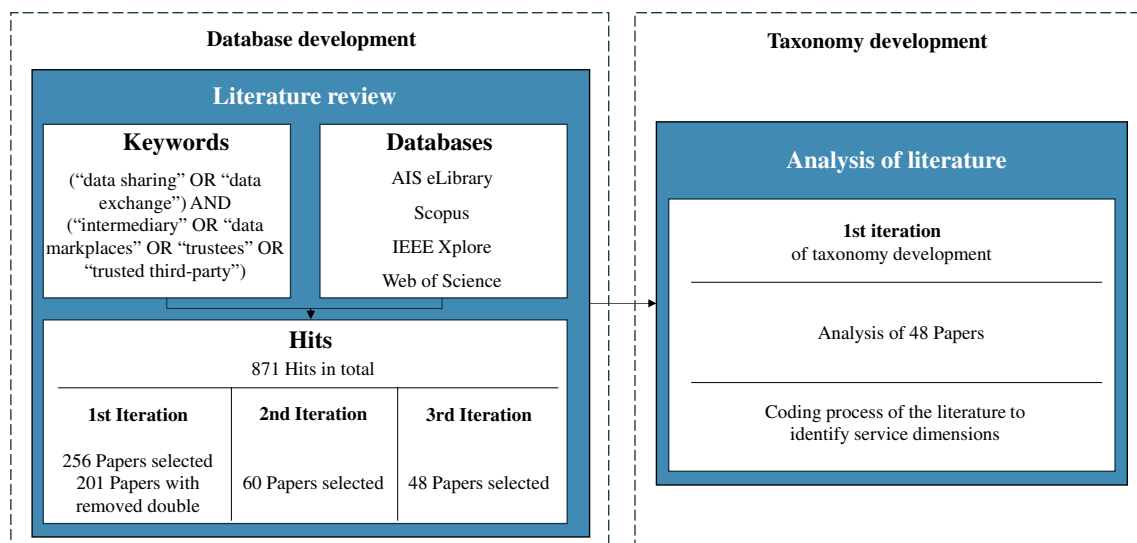


Fig. 2 Research design, literature review

**Analysis of the literature (first iteration of taxonomy development)** In the first iteration, we used a conceptual approach to evaluate the literature and analyze it using a Gioia diagram to explore which characteristics and services are attributed to DIs. We selected Gioia et al.'s (2013) approach as it is suitable for the iterative aggregation of textual data. Finally, we identified five service dimensions consisting of 15 characteristics: transaction services, governance services, sovereignty services, technology services, and data services. Table 3 shows an example of our literature coding.

**Public data analysis** To enrich our findings with real-world insights, we analyzed various companies that we identified as DIs (see Fig. 3). We built a database in three iterations to select the relevant companies for our research. The first sample of 116 companies we found through the MyData Operator Awardees (Langford et al., 2022; MyData, 2023), the Data Governance Act (European Commission, 2023), the Centre for Data Ethics and Innovation UK (Centre for Data Ethics & Innovation, 2021), and from previous interviews.

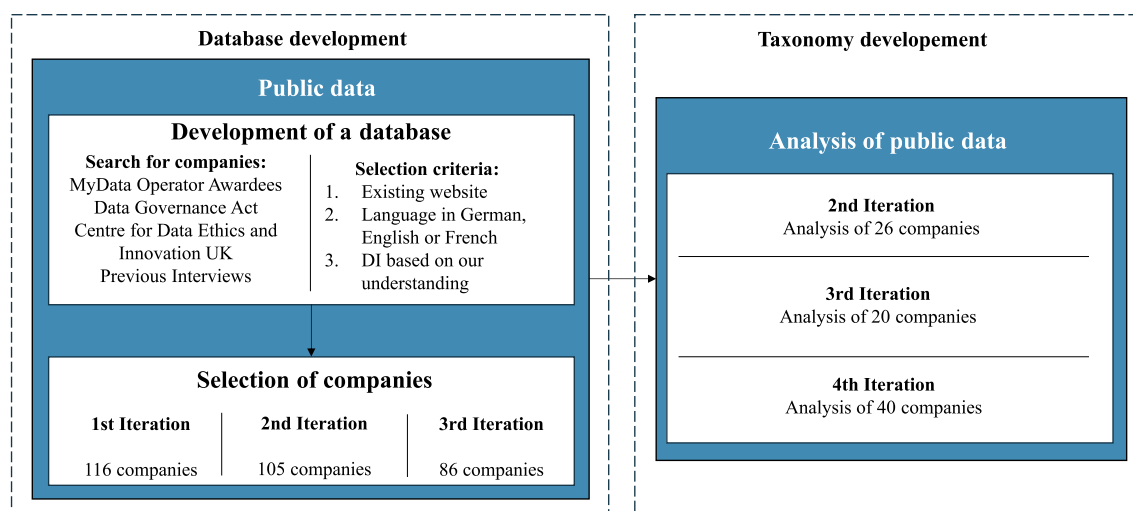
We reviewed all companies based on the following selection criteria (SE): (SE1) must have an existing website; (SE2) the website must be in German, English, or French; (SE3) the company is a DI based on the definition outlined in the DGA. Our final selection comprises 86 companies.

**Analysis of public data (second iteration of taxonomy development)** In the second iteration of the taxonomy development, we analyzed 26 of the selected companies. We used the dimensions from the first iteration (transaction, governance, sovereignty, technology, and data services) as a basis and categorized the findings into the dimensions. Through the analysis, we identify one further service dimension and 12 additional characteristics.

**Analysis of public data (third iteration of taxonomy development)** In the third iteration, we evaluated the companies to classify them according to the presence of the services based on the results of the second iteration (1 for agree, 0 for disagree). We checked if the services we found covered

Table 3 Examples of our literature coding

Example literature statement	Coding
"The role of the data trustee ensures that the platform <b>can enforce governance mechanisms</b> , such as enabling data exchange in <b>compliance with the law</b> and <b>managing access rights</b> ." Azkan et al., (2022, p. 7)	Governance mechanisms Compliance with law Access management
"However, data marketplace operators could also <b>offer computational infrastructure</b> as a service to ease the burden for data providers and consumers." Agahari et al., (2021, p. 9)	Provision of an infrastructure
"In case of <b>conflict resolution</b> , the clearing house acts as trusted third party by <b>verifying the validity and consistency of the logged data</b> hashes with the data provider's loggings." Dalmolen et al., (2019, p. 11)	Conflict resolution Verification



**Fig. 3** Research design, public data analysis

all the possibilities or if we should add further services. For this purpose, we investigated 20 companies. Finally, we add three services to the list: usage policies, pseudonymization, and certification.

**Analysis of public data (fourth iteration of taxonomy development)** In the fourth iteration, we checked whether we had missed any services in the previous iterations. Our approach was like the third iteration. We analyzed 40 other companies and were unable to identify additional service categories that had not been covered before. As we have not made any further changes to the taxonomy and all of Nickerson et al.'s (2013) end conditions have been met, we decided to end the taxonomy development. Figure 4 shows the taxonomy iterations.

**Phase 4 (demonstration)** In the demonstration phase, we apply three companies of our search process to the resulting taxonomy and demonstrate their fit. Table 4 shows the companies classified in our taxonomy (Catena-X (2024), Bundesdruckerei GmbH (2024), and DataGuard (2024)).

**Phase 5 (evaluation)** In the evaluation phase, we contacted the CEO of one of the companies (soviety) that we identified in our search process. In a workshop setting, we prepared a Miro board (Miro, 2024) and discussed both the classification of soviety in our taxonomy, as well as evaluated the resulting patterns (see below).

### Clustering data intermediation services

As the services are not exclusive, we identify reoccurring configurations of DISs. These typical combinations are called “archetypes” or “patterns.” Merriam-Webster

(2024a) defines archetypes as “the original pattern or model of which all things of the same type are representations or copies” and as a “perfect example.” To generate the patterns of typical DISs, we carried out a cluster analysis to check which services frequently occur in which combinations. Following K. D. Bailey (1994, p. 6), cluster analysis (similar, *pattern analysis*) is thematically based in psychology as “a quantitative method of classification.” In preparation for the cluster analysis, we coded all the companies found in the taxonomy using an Excel spreadsheet in which all services were listed. If services applied, we marked this with a 1 in the table; if the services did not apply, we marked this with a 0. We used RStudio (Version R 4.3.1) for the cluster analysis (RStudio Team, 2024) with the package “cluster” (Maechler et al., 2021) to derive a Euclidean distance matrix (Hellbrück, 2011). For the clustering method, we utilized agglomerative hierarchical clustering (AHC), according to Ward (1963). We use Ward’s method as it “minimize[s] within-group dispersion at each binary fusion” and “looks for clusters in multivariate Euclidean space” (Murtagh & Legendre, 2014, p. 275). When using AHC, the individual objects under consideration are placed in a “binary, rooted tree” (Murtagh & Contreras, 2017, p. 2). The final clusters were visualized in a dendrogram (Murtagh, 1984). A dendrogram is a way to show how many interrelated groups exist within the data visually and to derive relationships as it summarizes “the proximity and classification relationships” (Murtagh & Contreras, 2017, p. 3). The dendrogram (see Appendix 4) shows the sections in eight clusters.

We analyzed the resulting clusters with the entire team of authors. To identify the optimal number of clusters, we tested and analyzed different numbers of clusters. It is a typical procedure when using AHC to start with an initial

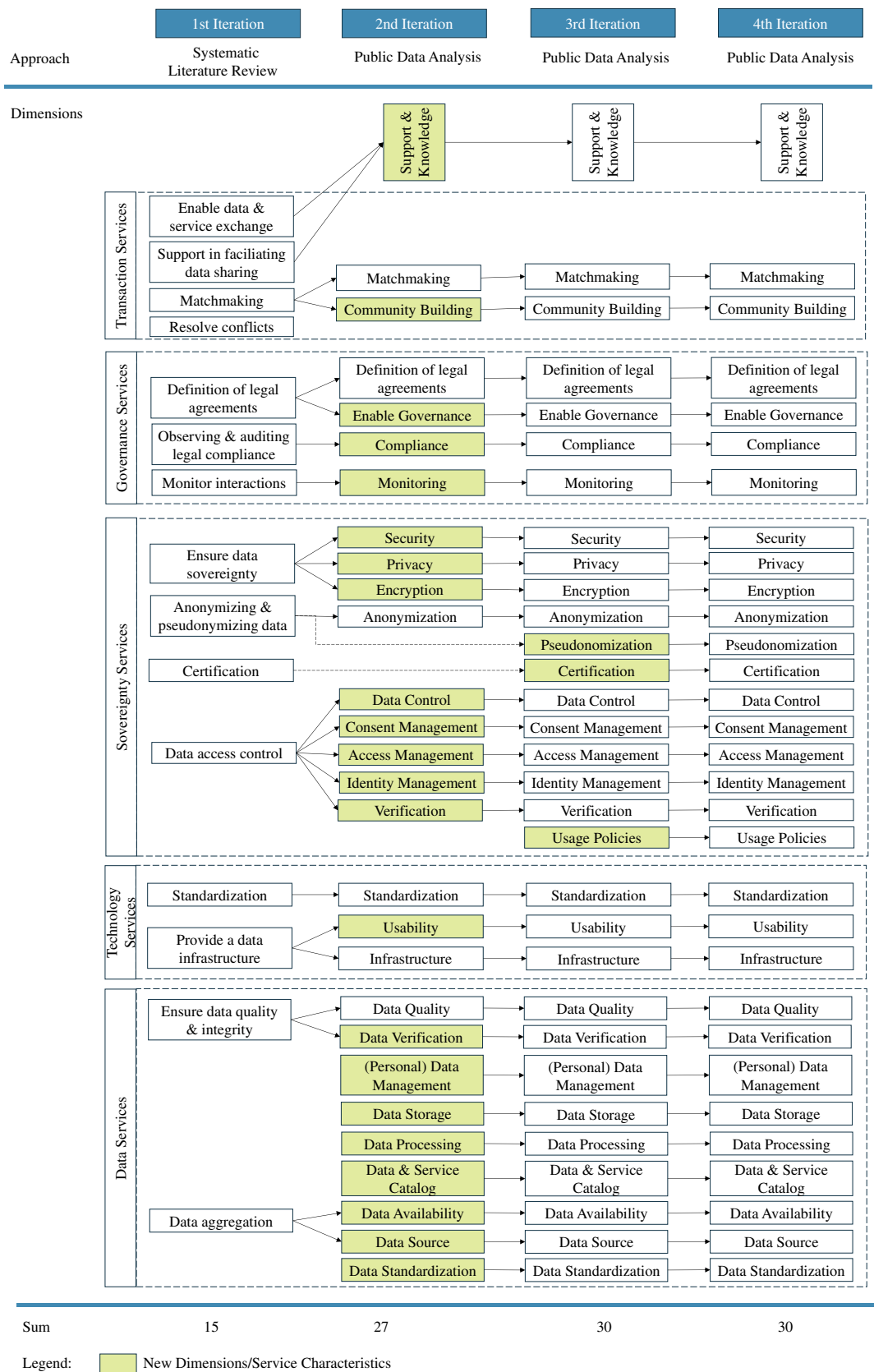


Fig. 4 Taxonomy design iterations

**Table 4** Demonstration of Catena-X (blue circle), Bundesdruckerei (orange square), and DataGuard (green triangle)

Dimension	Characteristics									
Transaction Services	Matchmaking					Community Building				
Governance Services	Monitoring		Compliance		Enable Governance		Define Legal Agreements			
Sovereignty Services	Data Control	Consent Management	Access Management		Identity Management		Verification		Security	
	Privacy	Anonymization	Pseudonymization		Encryption		Usage Policies		Certification	
Technology Services	Infrastructure			Usability			Standardization			
Data Services	Data Quality	Data Verification		(Personal) Data Management		Data Storage		Data Processing (Analysis/ Visualization/ Aggregation/ Reporting)		
	Data & Service Catalog	Data Availability		Data Source (Provision/ Generation)		Data Standardization				
Support/ Knowledge										
Legend: Catena-X   Bundesdruckerei   DataGuard										

cut into  $k$  clusters without having previously calculated which number of clusters would be optimal (Murtagh & Contreras, 2017). In an initial cut of the clusters, we intuitively decided on ten clusters. With ten clusters, however, most of the clusters consisted of around five companies and no longer showed any combinations of services but rather a specialization in one service. The selection of the number of clusters is based on a qualitative analysis of comparable characteristics. To do this, we first evaluated the dendrogram with six, eight, and ten clusters (see Appendix 4). Ten clusters could thus be excluded, as, in our view, the granularity of the services was lost. To decide between the remaining clusters, we decided to compare the cluster numbers 6, 7, and 8 in more detail. We calculated the percentage shares of the services and compared them with each other. The focus of our analysis was to discover a significant number of the clusters. We found that eight clusters are the optimal number, as certain cluster characteristics “disappeared” with a lower number as they were merged into larger cluster groups (e.g., the transaction focus from cluster 7 disappeared completely with a lower number of clusters). In our opinion, eight clusters reflect the optimal mix of a generalized overview of service patterns but also provide the necessary finer nuances that show the variance of service combinations. We then compared the eight clusters for their core differences and thus derived the typical patterns of the DIS. Therefore, we compared the percentages of occurrence of the individual services in the respective clusters (see Appendix 5) and worked out the characteristics that distinguish the respective clusters from one another.

## Findings: conceptualizing data intermediation services

### Understanding data intermediation services

The basis for conceptualizing DISs is the final taxonomy, which reflects the set of DISs we found in our sample (see Fig. 5): transaction services (1), governance services (2), sovereignty services (3), technology services (4), data services (5), and support/knowledge services (6). In practice, these services do not occur at the same time, but in our sample, each object occupied at least one of those non-exclusive service configurations. Figure 5 shows the underlying taxonomy that we used to procure the visualization of DI as a cumulation of each distinct DIS pattern, which we explain below. It indicates *mandatory*, *essential*, and *optional* elements that we identified in the analysis phase. *Mandatory* elements are present in every instance (100%), *essential* elements appear in **over 80%** of the instances, and *optional* elements are found in all of the instances **below 80%**.

**Mandatory elements** include the provision of data-sharing infrastructure (e.g., data marketplaces), supporting data providers and consumers during the process, and signaling trust. For instance, providing technical infrastructure as well as support and knowledge is an essential DIS that each DI offers. Irrespective of the affiliation to the service patterns and the associated specialization or generalization of the services, all the companies provide infrastructure or support as a foundation for additional services. One example is sovity (2024), which offers support in the implementation of data sharing, as well as access to the necessary infrastructure. Gebäudedaten.ch (2024) provides the infrastructure through which data is made available. **Essential elements**

Dimension	Characteristics						
Transaction Services	Matchmaking (26%)			Community Building (79%)			
Governance Services	Monitoring (42%)		Compliance (83%)		Enable Governance (81%)	Define Legal Agreements (66%)	
Sovereignty Services	Data Control (79%)	Consent Management (86%)	Access Management (78%)		Identity Management (83%)	Verification (36%)	Security (71%)
	Privacy (58%)	Anonymization (28%)	Pseudonymization (13%)		Encryption (43%)	Usage Policies (70%)	Certification (26%)
Technology Services	Infrastructure (100%)			Usability (52%)		Standardization (52%)	
Data Services	Data Quality (35%)	Data Verification (27%)	(Personal) Data Management (57%)		Data Storage (53%)	Data Processing (Analysis/ Visualization/ Aggregation/ Reporting) (65%)	
	Data & Service Catalog (67%)	Data Availability (58%)	Data Source (Provision/ Generation) (55%)		Data Standardization (21%)		
Support/ Knowledge (100%)							
Mandatory elements (100%): Infrastructure, Support/Knowledge							
Essential elements (>80%): Compliance, Enable Governance, Consent Management, Identity Management							
Optional elements (< 80%): Matchmaking, Community Building, Monitoring, Define Legal Agreements, Data Control, Access Management, Verification, Security, Privacy, Anonymization, Pseudonymization, Encryption, Usage Policies, Certification, Usability, Standardization, Data Quality, Data Verification, (Personal) Data Management, Data Storage, Data Processing, Data & Service Catalog, Data Availability, Data Source, Data Standardization							

Fig. 5 Overview of the DIS

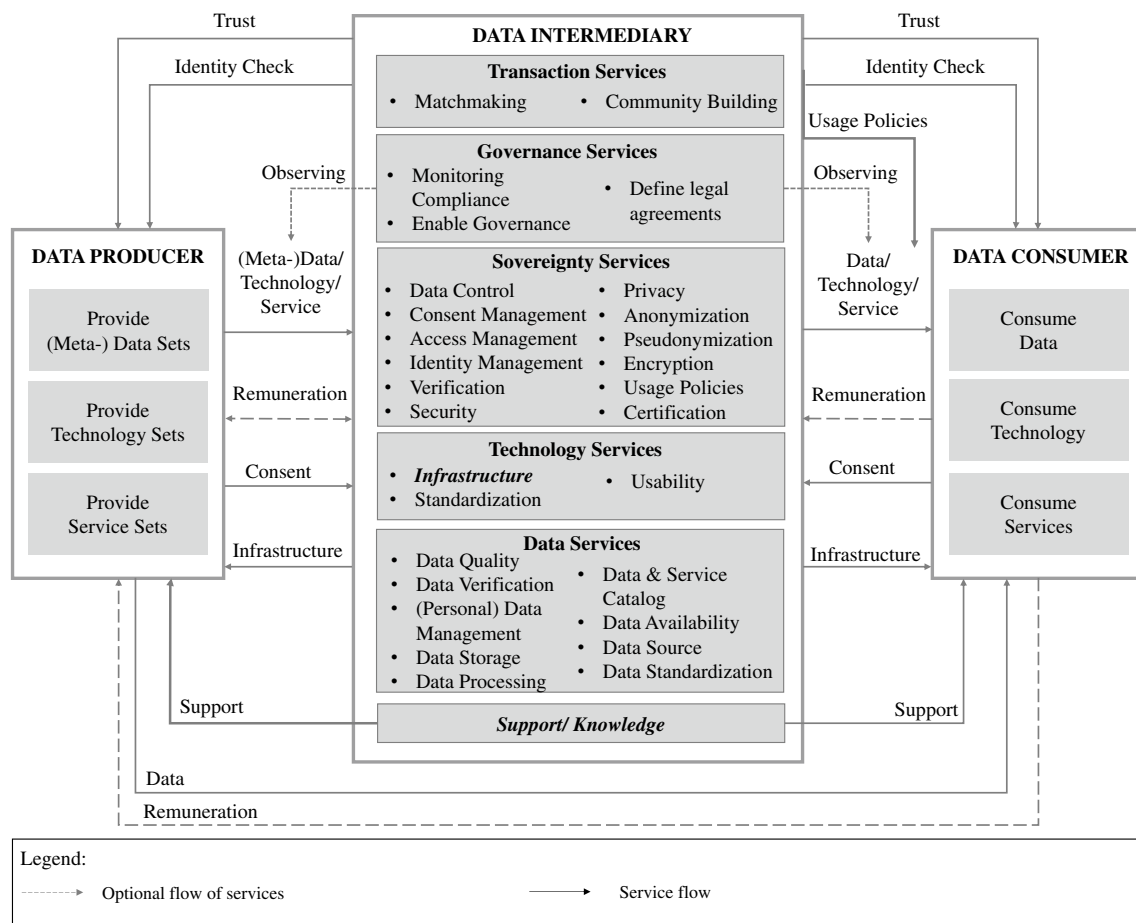
are enabling compliance and governance, managing consent for data sharing, and verifying the identities of the different actors. **Optional elements** are those pertaining to specific services that aim to solve a specific problem. This includes anonymization of data, which might be relevant in some cases but is not mandatory across all potential data-sharing cases. The fact that not all services are mandatory is shown again by the example of sovti (2024), as they do not offer, for example, any data quality or data verification services.

### The connection between DI and DISs

In our understanding, a DI provides DISs. For this reason, it is important to understand how DISs relate to DIs. Initially, we procured a generalized understanding that reflects the sum total of potential the DISs can inhibit. This understanding is necessary to comprehend the DI as an umbrella for a range of instances, which we complement through configurations of DISs. Figure 6 visualizes our understanding of the connection between a DI and DISs. The illustration is *utopian* (Weber, 1949) in the sense that it represents a DI consisting of all the potential DISs that could be offered, which, in reality, is highly unlikely or potentially impossible. As previously mentioned, our analysis shows that the services are offered at different frequencies. This leads us to conclude that it is almost a utopian idea for a DI to offer all services. The illustration is based on Stachon et al. (2023), who show how a data trustee works. In our case, we use the

findings from the taxonomy (Fig. 5) and combine them into one figure to show the connection between the services and the data intermediary in the data-sharing context.

Given its nature as a two-sided digital platform (Ditfurth & Lienemann, 2022), which “facilitates the use of data for other actors” (Oliveira et al., 2019 p. 609), it orchestrates data sharing and related services for *data consumers* and *data producers*. Each DI organizes the *flow of data* from a data producer to a data consumer, but not every DI monetizes the sharing process. The DI is at the center of data-sharing activities. In detail, data sharing through the data intermediary works as follows: The data producer wants to share their (meta-)data (1). Potential motivation for this can be monetization (e.g., in data marketplaces) or to comply with legal obligations (e.g., the Supply Chain Act) (2). The data consumer requires data, technology, or services for a specific use case and considers a data intermediary as a data source (3). Depending on the use case, the DI is selected based on the DISs it offers. For instance, the data consumer could require data that has clear usage policies and ensures the legality of using the data — something typically found in data trusts (e.g., Stachon et al., 2023). If the DI now receives a request that someone wants to share data, it first checks the identity of the actor. This is to prevent data misuse, and only the correct recipients should receive the data intended for them. Once the DI has obtained consent for data sharing and compliance with the specified policies, it usually provides the necessary infrastructure. In some cases, the data or service is only made available in



**Fig. 6** The connection of DIs and DISs

return for an equivalent value as monetary compensation or other forms of remuneration (Jussen et al., 2024b). During data exchange, the DI monitors the process and compliance with usage and access policies. In addition to the scenario in which the data is exchanged via the DI, there is another scenario in which the DI merely matches the data consumer and data producer. These actors can exchange their data and the optional counter values independently of the DI.

### Patterns of data intermediation services

The analysis of the services shows that DIs offer combinations of DISs. To identify these combinations, we carried out a cluster analysis with a final sample of 86 companies (see Fig. 3). We identified eight patterns with service combinations that occur together: *privacy and anonymization*, *data control*, *infrastructure*, *data catalog*, *governance and sovereignty*, *identity management*, *transaction*, and *enabling data sharing*. In this section, we present all eight clusters and their characteristics. Table 5 provides a complete overview of all patterns.

#### Pattern 1: Enable privacy and anonymization

Pattern 1 establishes **privacy through anonymization**. This is evident by the use of encryption and/or anonymization of data to prevent unauthorized users from gaining access to sensitive data. DIs that belong to this pattern offer consent and access management. Users retain full control over their data and ensure that their data cannot be used without authorization. The data is analyzed, visualized, or offered in a data catalog. The characterizing services of the pattern are:

- Definition of legal agreements
- Privacy
- Anonymization
- Encryption

Various technologies or combinations with other services support **privacy** and the protection of personal data. Datavillage (2024) operates a platform that enables companies to handle sensitive data securely and ensures privacy through technologies (e.g., gateway-based architecture and

**Table 5** Patterns of data intermediation services

	Pattern 1	Pattern 2	Pattern 3	Pattern 4	Pattern 5	Pattern 6	Pattern 7	Pattern 8
Name	Privacy and anonymization	Data control	Infrastructure	Data catalog	Governance and sovereignty	Identity management	Transaction	Enabling data sharing
Count	6	9	11	9	25	15	5	6
Unique selling point (> 90%)	<ul style="list-style-type: none"> <li>• Definition of legal agreements</li> <li>• Data control</li> <li>• Consent management</li> <li>• Privacy</li> <li>• Anonymization</li> <li>• Usability</li> </ul>	<ul style="list-style-type: none"> <li>• Community building</li> <li>• Compliance</li> <li>• Enable governance</li> <li>• Data control</li> <li>• Access management</li> <li>• Identity management</li> <li>• Usage policies</li> </ul>	<ul style="list-style-type: none"> <li>• Consent management</li> <li>• Identity management</li> </ul>	<ul style="list-style-type: none"> <li>• Data control</li> <li>• Consent management</li> <li>• Data catalog</li> </ul>	<ul style="list-style-type: none"> <li>• Compliance</li> <li>• Enable governance</li> <li>• Consent management</li> <li>• Access management</li> <li>• Security</li> <li>• Data processing</li> <li>• Data availability</li> <li>• Data source</li> </ul>	<ul style="list-style-type: none"> <li>• Identity management</li> </ul>	<ul style="list-style-type: none"> <li>• Community building</li> <li>• Usage policies</li> <li>• Data processing</li> </ul>	<ul style="list-style-type: none"> <li>• Compliance</li> <li>• Identity management</li> <li>• Usage policies</li> <li>• Usability</li> <li>• Data processing</li> </ul>
Additional typical services (70–90%)	<ul style="list-style-type: none"> <li>• Support/knowledge</li> <li>• Compliance</li> <li>• Enable governance</li> <li>• Access management</li> <li>• Encryption</li> <li>• Standardization</li> <li>• Data processing</li> <li>• Data and service catalog</li> </ul>	<ul style="list-style-type: none"> <li>• Consent management</li> <li>• Verification</li> <li>• Security</li> <li>• Usability</li> <li>• Standardization</li> <li>• Data and service catalog</li> </ul>	<ul style="list-style-type: none"> <li>• Compliance</li> <li>• Enable governance</li> <li>• Access management</li> <li>• Usage policies</li> </ul>	<ul style="list-style-type: none"> <li>• Community building</li> <li>• Access management</li> <li>• Identity management</li> <li>• Security</li> <li>• Privacy</li> <li>• Data storage</li> <li>• Data availability</li> <li>• Data source</li> </ul>	<ul style="list-style-type: none"> <li>• Community building</li> <li>• Definition of legal agreements</li> <li>• Data control</li> <li>• Identity management</li> <li>• Privacy</li> <li>• Usage policies</li> <li>• (personal) data management</li> <li>• Data and service catalog</li> </ul>	<ul style="list-style-type: none"> <li>• Community building</li> <li>• Enable governance</li> <li>• Consent management</li> <li>• Access management</li> <li>• Standardization</li> </ul>	<ul style="list-style-type: none"> <li>• Matchmaking</li> </ul>	<ul style="list-style-type: none"> <li>• Community building</li> <li>• Data control</li> <li>• Consent management</li> <li>• Security</li> <li>• Data quality</li> </ul>
Examples	<a href="https://www.datavillage.me/">https://www.datavillage.me/</a> <a href="https://worlddataexchange.com/">https://worlddataexchange.com/</a>	<a href="https://www.meeco.me/">https://www.meeco.me/</a> <a href="https://www.onecub.com/">https://www.onecub.com/</a>	<a href="https://www.mydataashare.com/">https://www.mydataashare.com/</a> <a href="https://www.gebaeudedaten.ch/">https://www.gebaeudedaten.ch/</a>	<a href="https://schluss.org/">https://schluss.org/</a> <a href="https://dataforgoodfoundation.org/">https://dataforgoodfoundation.org/</a>	<a href="https://www.snowflake.com/en/">https://www.snowflake.com/en/</a> <a href="https://www.dawex.com/en/">https://www.dawex.com/en/</a>	<a href="https://www.comunity.de/">https://www.comunity.de/</a> <a href="https://www.gravity.to.net/">https://www.gravity.to.net/</a>	<a href="https://bdrtrust.org/">https://bdrtrust.org/</a> <a href="https://www.startupcommons.org/">https://www.startupcommons.org/</a>	<a href="https://sovity.de/en/">https://sovity.de/en/</a> <a href="https://sovity-en.de/">https://sovity-en.de/</a>

encryption). Further privacy-preserving features are offered, such as “trusted execution and granular consent” and all data is encrypted and anonymized (Datavillage, 2024). One way to ensure data sovereignty is to change the data in such a way that it is no longer possible to uncover the data provider (through **anonymization or pseudonymization**) or that the data is not accessible to read (**encryption**). World Data Exchange (2024) offers privacy-enhancing services. These services are created to be “private by design” by encrypting the data and not permanently storing the data (i.e., deleting it after transmission). In addition to the existing legal requirements, agreements are necessary to secure the process of data sharing. The DI defines the required **legal agreements** with his expertise in data sharing and legal knowledge. Legal agreements are data management and governance frameworks or contracts, including data usage/sharing agreements (Bastiaansen et al., 2020; Dalmolen et al., 2019; Demchenko et al., 2019; Noorian et al., 2014).

### Pattern 2: Enable data control

The second pattern, **data control**, characterizes users retaining control over their data. The characterizing services of the pattern are:

- Data control
- Access control
- Enable governance

Data control is a core service of sovereignty services. The service focuses on giving users control over their data through the combination of other services. The service users have the opportunity to decide which tools/services they want to use (JLINC, 2022). One way to ensure data control is to restrict data access. **Access management** means that the data provider or DI grants access to the data. The decision with whom to share the data rests with the data provider (see OPEN BANKING, 2024). Access is only authorized for a certain period or under certain conditions and can be revoked at any time. The role of the intermediary has often been associated with regulating access to data (Kurtz et al., 2019; Marotta et al., 2021). Another way to gain control over the data is to **enable governance**. Lack of know-how, time, or human resources leads to challenges in legal compliance. The DIs take on the task of ensuring conformity with the applicable laws, and users do not have to deal with legal issues themselves. Users rely on the fact that the services offered by their intermediary are already compliant or provide them with ready-to-use tools that support them in governance issues. In the most straightforward implementation, it means compliance with the GDPR (e.g., Geens, 2024).

### Pattern 3: Providing infrastructure

By **providing the infrastructure** to its customers, the DI reduces the complexity of the data-sharing process for data providers and data consumers and enables trust and transparency (Agahari et al., 2021; Noorian et al., 2014). The characteristic service of the pattern is providing infrastructure. There are several options for providing infrastructure: software, hardware, data-related services, and platforms (Fuerstenau & Auschra, 2016; Kurtz et al., 2019; Öksüz, 2014; Schmidt, 2022; Woroch & Strobel, 2022). In practice, we find the provision of infrastructure in the form of a *marketplace* (e.g., Mobility Data Space, 2024), an *API* (e.g., Visions, 2021), or a *connector* (e.g., soivity 2024). The pattern covers the minimum necessary elements for data exchange as well as additional services that are part of the provision of infrastructure, such as consent and identity management, ensuring infrastructure compliance, access management, and usage policies for the infrastructure. Examples of the pattern are the building database from Switzerland (orig. “Gebäudedaten.ch”) and MyDataShare (Gebäudedaten.ch, 2024; MyDataShare, 2023).

### Pattern 4: Providing a data catalog

The fourth pattern focuses on providing a **data catalog**. A DI provides the aggregated or generated data in the form of a data or service catalog. Other DIs only offer various services, which we refer to as service catalogs. The characterizing services of the pattern are:

- Data catalog
- Consent management
- Data availability
- Data storage
- Data source

To be able to share data, it is necessary to obtain the consent of the data provider for the data exchange. With **consent management**, the data provider gives consent for the data to be exchanged or allows other parties to access their data. This consent applies for an undefined or defined period or relates to the frequency of access. In some instances, each individual’s access is authorized. The data consumer can submit consent requests to gain data access. DIs provide an overview of the consent given in the visual form to facilitate an overview of the data-sharing processes (Data for Good Foundation, 2024). For example, users of Schluss’ (2024) services decide for themselves who is allowed to access the data and decide for themselves which data they want to pass. They explicitly consent to the sharing of data. The Data for Good Foundation (2024) offers “consent-as-a-service” to give users the opportunity to retain control over their data

while also exuding confidence by ensuring compliance with current and future EU regulations.

### Data availability

Represents the mobilization of data. Following Figueredo et al. (2020), it is the DI's task to make the data available to consumers. DIs from healthcare make data on rare diseases available to their consumers that they would otherwise not have been able to obtain. Other DIs generate the data they provide themselves, such as the Data for Good Foundation (2024), which pursues the goal of making data accessible to users. **Data storage** involves the secure preservation of user data. This can use end-to-end encrypted cloud storage or in a personal vault, for example, Schluss (2024) or Cozy (2024). Other forms of data storage mean that the data is not stored on the DI's infrastructure but that the intermediaries offer the data storage solution that the data is stored on the users' devices (CitizienMe, 2024).

### Pattern 5: Enable governance and sovereignty

The fifth pattern represents the ***governance and sovereignty*** DIs. The core services focus on governance aspects, such as compliance and governance enabling, as well as sovereignty services, such as access and consent management. Moreover, data services such as data availability and the provision of data sources are part of the narrower offering. The characterizing services of the pattern are, in comparison to the other patterns:

- Security
- Data processing
- Management of (personal) data

**Security** is ensured through various service offerings or the combination of different services. This can be done by defining certain safety requirements. Mydex (2024) has developed a security model that defines how security is ensured when using Mydex, and iGrant.io (2024) provides digital wallets that comply with the applicable security requirements. Security services are guaranteed by compliance with various certified principles, as shown by HIE of ONE (2021). In addition to ensuring a secure environment for data sharing, DIs from the fifth pattern offer data services that set them apart from other DIs. These include the processing of data and the management of (personal) data.

**Data processing** deals with the treatment of data and includes ***data analysis*** (OpenSAFELY, 2024; RACOON, 2024; Snowflake, 2023), ***visualization*** (RACOON, 2024; Snowflake, 2023; Visions, 2021), ***aggregation*** from various data sources, and ***data reporting*** (Visions, 2021). **Data analysis** refers to the service that interprets or evaluates the

data, as shown by the evaluation of resource capacities from Snowflake (2023). These analysis services investigate the data automatically (RACOON, 2024) and are closely linked to the ***visualization of the data***. Visualizations of data are helpful to get an overview of the available data. RACOON (2024) and Visions (2021) use visualizations to display clearly with the help of a dashboard. **Data aggregation** is the merging of data from different sources. DIs such as the West German Biobank Essen or the UK Biobank collect samples and data from volunteers and make the data available to researchers (WBE, 2019). The German Centre for Cancer Registry Data (2013) aggregates the data from various registries and provides clinical data on different types of cancer. **Reports** present information about which data has been shared. These reports are passed to other players, as in the Visions (2021) example. Reporting means that research results developed on the provided data must be reported back to the DI, which publishes them for the data provider (cf. WBE (2019)).

The ***management (personal) data*** means that the DI supports its customers in organizing and using their data. AWS Data Exchange (2023) offers different options for data management as they provide an overview of all available data sets (such as data files, data tables, and APIs). Other data management options include the organization of health data (Pensions Dashboards Programme, 2024). The aim is to make the data available to users in one place on a platform or in an app (Dawex, 2023; Pensions Dashboards Programme, 2024; Yivi, 2024). Data management refers to the management of data-sharing partners. The DI service consists of bringing together all of the user's customers and keeping them up to date automatically, as offered by CDQ (2024). Similarly, this service is part of the Dawex platform, which offers support for user collaboration (Dawex, 2023).

### Pattern 6: Enable identity management

The sixth pattern focuses on identity management since access to sensitive data represents a major security risk for the data provider and a certain level of trust for the data consumer. The characterizing services of the pattern are:

- Identity management
- Standardization

By offering ***identity management***, the DI supports the required trust for data sharing among the participants in the data exchange. One possibility is the provision of a digital identity card or digital wallet (e.g., identities of individuals or companies), as offered by Findy (2024) or Comuny (2024). By verifying the identities of users, a trustworthy exchange of data is guaranteed, and processes are made more

efficient (Findy, 2024). Another form of identity management is the registration of all users who want to participate, as is the case with the Mobility Data Space (2024) or Smart Connected Supplier Network (2024). DIs such as Comuny (2024) require verification of personal details such as e-mail address, age, or biometric facts.

In addition to enabling identity management, the sixth pattern includes other services: providing the infrastructure, establishing the necessary standards for data exchange, regulating access to the data, ensuring governance and consent management, and bringing together various stakeholders. For instance, shared standards facilitate the data exchange between different actors, and the task of the intermediary is to provide the necessary *standards* (Dalmolen et al., 2019). The Smart Connected Supplier Network (2024) (short, SCSN) offers a message standard that all service providers in the SCSN network must use.

### Pattern 7: Enable transaction

The seventh pattern deals with the implementation of transaction services in data sharing to connect the right players so that data-sharing activities can be carried out. The characterizing services of the pattern are:

- Community building
- Matchmaking

DIs support data providers and data consumers in building their networks for data sharing as the building process of data ecosystems or data spaces. Perscheid et al., (2020, p.1) declare the coordination function of an intermediary to build useful networks for data-sharing activities, and Agahari et al., (2021, p.9) identify one service of a DI as the matchmaking service that brings data providers and data consumers together. This includes not only the matchmaking of demand and supply but also of data providers and data consumers or different markets (Agahari et al., 2021; Demchenko et al., 2019; Huang et al., 2021; Schrieck et al., 2016). Startup Commons (2024) supports the formation of ecosystems by providing knowledge and enabling a global exchange between different players who share their knowledge. Personium (2020) offers an infrastructure that enables secure data exchange between data storage devices and connects them. This allows data sharing between different groups of individuals and the connection of stakeholders with similar thematic concerns, such as doctors and nursing staff or teachers, tutors, and students. DIs are not only matching actors who already know each other, they also match previously unknown parties. In addition to personal contact with experts, BDT (2024) offers technical tools such as chatbots to give stakeholders access to knowledge at any time.

### Pattern 8: Enabling data sharing

Pattern 8 focuses on the *enabling of data sharing* and establishing of data-sharing processes. The characterizing services of the pattern are:

- Compliance
- Usability
- Usage policies
- Data quality

Enabling data sharing can be technical-driven, like the provision of infrastructure or connectors, but also the general support in the establishment process. It includes ensuring that the services are compliant with applicable regulations, that access and consent management are regulated, that usage policies for data are established, and that control and security over the data are guaranteed. DIs of this pattern support the development of ecosystems or the establishment of networks for data sharing and data services such as data processing or data quality services. DIs like Agdatahub (2024), Catena-X (2024), or soivity (2024) are part of the eighth pattern.

To ensure that the service is adapted to the conditions of the respective user, DIs offer *usability* services. This means that the service is not rigid and fixed but that adjustments can be made, such as scalability (e.g., soivity (2024) or Streamr (2024)), interoperability (e.g., soivity (2024) or Catena-X (2024)), or tailor-made solutions (e.g., Numbers (2024) or soivity (2024)). Numbers (2024) gives software developers the opportunity to build new applications for their specific use. *Usage policies* define in which form and under which conditions the shared data can be used. These are time conditions, the number of downloads, or the number of users. Depending on the DI, the usage policies are defined by the data provider or by the DI itself. With DIs like soivity (2024) or Agdatahub (2024), users of the service can define their data usage conditions.

*Compliance* with the applicable laws, such as the General Data Protection Regulation (short, GDPR) (European Commission, 2016) or the Data Governance Act (European Commission, 2020), is one of the central aspects of sharing data. MyDataShare (2023) or OCKTO (2022) ensure that their customers are compliant with the applicable laws. Users of Myfairdata (2020) send GDPR requests, and onecub (2024) guarantees compliance by warranting that the entire infrastructure is aligned with the GDPR. It is not only the GDPR or Data Governance Act that is observed by DIs.

When receiving external data, there are often concerns regarding the *data quality*. This is highlighted by Demchenko et al. (2019) that the data must be of appropriate quality, and Huang et al. (2021) emphasize that data quality increases the motivation to participate in data sharing.

In practice, we find various forms of data quality services that are related to data verification. The data of the German Centre for Cancer Registry Data (2013) is checked for completeness and plausibility. Other DIs indicate where their data comes from to be transparent and allow users to decide whether or not they consider the data quality to be appropriate based on its origin (CDQ, 2024; Gebäudedaten.ch, 2024). This transparency on data quality can be visualized in the form of a dashboard (DATA SENTINEL, 2024).

## Demonstration and evaluation

Following the method of Kundisch et al. (2021), evaluating and demonstrating that the taxonomy works are the penultimate steps to reporting the taxonomy. In particular, we aim to evaluate the usefulness of the taxonomy, both for presenting service configurations accurately as well as to support innovation (see Kundisch et al., 2021; Szopinski et al., 2019). To configure the evaluation, we drew from Szopinski et al. (2019) and reached out to the CEO of a company in our sample — soivity — to whom we had access. Soivity is a Fraunhofer spin-off with expertise in the field of sovereign data sharing through building data spaces (soivity, 2024). They are involved in the development of data space standards, and their vision is “to empower effortless and safe connection, collaboration, and innovation based on data for everyone” (soivity, 2024). The core offering of soivity is connecting customers to existing data spaces and building their data space instances. This encompasses the provision of expert knowledge, easy-to-implement, standardized, certified solutions, and software products for data spaces. The value-generating partnership and customer network comprises not only industrial companies such as BMW and Volkswagen but also initiatives such as the Mobility Data Space and research institutes, including the German Aerospace Center (German, Deutsches Zentrum für Luft- und Raumfahrt; short, DLR).

This enabled us to acquire an outside view of the taxonomy and patterns from practice. We prepared a digital workshop in Miro and asked the expert to apply the taxonomy in two ways. First (1), to map soivity using the taxonomy (the ability of the taxonomy to **represent** DISs), and second (2), to use the taxonomy to identify potential service extensions (the ability of the taxonomy to support **innovation**). Workshops are a suitable method for gaining new insights or evaluating existing ones (Thoring et al., 2020). Due to its expert knowledge and know-how for the technical implementation of data sharing, soivity is a typical representative of the “Enable Data Sharing” pattern. The workshop consisted of four specific tasks that were discussed together. Some parts needed to be actively completed, while other parts only had to be evaluated and commented on. **First (1)**, we categorized soivity in our taxonomy. The services marked (in Table 6) in

dark gray are very applicable, and the services marked in light gray are slightly applicable. Through this classification, we recognized that not all services necessarily have to be highly present. In the **second step (2)**, we checked whether soivity could be found in the intended pattern. Here, we describe an element of the pattern as an example and then show how soivity implements it. The information was taken from the soivity website.<sup>2</sup> As part of the evaluation, we focused on the aspect of providing infrastructure with additional tools, such as identity and consent management. Soivity (2024) implements this through a connector that uses authentication mechanisms. In discussions with our workshop partner, we identify a match with the eighth pattern. To confirm this agreement more precisely, in a **third step (3)**, we derived 16 goals from the publicly available data on the soivity website and compared them with the typical characteristics of the pattern. During the workshop, we reviewed our classification together with soivity. Ten matches with the characteristics of the eighth pattern were identified.

In the **fourth** and final part of the workshop (4), we derived ideas for expanding the service configuration based on the unaddressed goals that we identified in the third step. We were able to receive two potential extensions in the workshop. Firstly, an extension to data quality or data verification could be an essential option. As the data format should be the same when exchanging data, a service should be implemented on the provider side that takes care of the conformity of the data. Data processing is playing an increasingly important role, especially in relation to the implementation of various use cases, as is the case with soivity. Table 6 shows the results of our evaluation with soivity inspired by Schöbel et al. (2020).

We derived **three key learnings** from the workshop in terms of content. First (1), definitions and explanations of the services must be provided in addition to the pure terminology to avoid misunderstandings. For this reason, we created a glossary following the workshop that explains all services briefly (see Appendix 3). Second (2), services can be offered in various forms within the company. Third (3), insights into service configurations can be gained based on our patterns.

## Discussion

During our analysis and extraction of the patterns, we accumulated knowledge about DISs, which we will discuss below. **First (1)**, the DGA regulates DISs in the EU, and organizations need to identify and categorize their importance and relevance in relation to their services. While the

<sup>2</sup> Website of soivity, <https://soivity.de/en/connect-to-data-space-en/>

**Table 6** The case of soivity: analyzing conformity with DIS patterns

Understanding design characteristics : soivity's classification in the taxonomy (1)	Dimension	Characteristics					
	Transaction Service	Matchmaking			Community Building		
	Governance Service	Monitoring	Compliance	Enable Governance		Define legal agreements	
	Sovereignty Services	Data Control	Consent Management	Access Management	Identity Management	Verification	Security
		Privacy	Anonymization	Pseudonymization	Encryption	Usage Policies	Certification
	Technology Services	Infrastructure		Usability		Standardization	
	Data Services	Data Quality	Data Verification	(Personal) Data Management		Data Storage	Data Processing (Analysis/ Visualization/ Aggregation/ Reporting)
Data & Service Catalog		Data Availability	Data Source (Provision/ Generation)		Data Standardization		
Support/ Knowledge							

Verification of the soivity and pattern 8 conformity (2)	<div>Enabling Data Sharing</div> <div>Pattern 8</div>	<div>Providing a Data Sharing Infrastructure with all the necessary Tools like Identity Management &amp; Consent Management</div> <div>Description of an aspect of the pattern</div>	<div>"The Connector acts as secure API gateway for data sources. You connect various IT systems or databases to the Connector with authentication mechanisms and define access and usage conditions to share data. Different formats are no longer an obstacle since automatically converted, this enables standardized communication. Your data offerings, which are published in a catalog, automatically become legally binding technical contracts with consumers in the Data Space accepting your terms of use."</div> <div>Description of the implementation of soivity</div>
--	---	--	---

Goals of intermediation services (3)	<b>Goals of soivity (taken from the soivity website)</b>	<b>Matching goals</b>																																
	<table><tr><th>#</th><th>Goals of soivity</th></tr><tr><td>1</td><td>Technology validated in the industry</td></tr><tr><td>2</td><td>Expert knowledge</td></tr><tr><td>3</td><td>Conserving resources and reducing costs</td></tr><tr><td>4</td><td>Added value through secure data exchange</td></tr><tr><td>5</td><td>Plug-and-play without programming</td></tr><tr><td>6</td><td>Scalable solution</td></tr><tr><td>7</td><td>Data remains with the data owner</td></tr><tr><td>8</td><td>Integration into the connector</td></tr><tr><td>9</td><td>Sovereign and secure data transfer through standardized communication protocols</td></tr><tr><td>10</td><td>Establishment of terms of use</td></tr><tr><td>11</td><td>Development of a data contract</td></tr><tr><td>12</td><td>Offering a data catalog</td></tr><tr><td>13</td><td>Offering a dashboard for an overview of the data used</td></tr><tr><td>14</td><td>Monitoring data transfer and capacities</td></tr><tr><td>15</td><td>Wizard for data offers</td></tr><tr><td>16</td><td>Information on the data connection</td></tr></table>		#	Goals of soivity	1	Technology validated in the industry	2	Expert knowledge	3	Conserving resources and reducing costs	4	Added value through secure data exchange	5	Plug-and-play without programming	6	Scalable solution	7	Data remains with the data owner	8	Integration into the connector	9	Sovereign and secure data transfer through standardized communication protocols	10	Establishment of terms of use	11	Development of a data contract	12	Offering a data catalog	13	Offering a dashboard for an overview of the data used	14	Monitoring data transfer and capacities	15	Wizard for data offers
#	Goals of soivity																																	
1	Technology validated in the industry																																	
2	Expert knowledge																																	
3	Conserving resources and reducing costs																																	
4	Added value through secure data exchange																																	
5	Plug-and-play without programming																																	
6	Scalable solution																																	
7	Data remains with the data owner																																	
8	Integration into the connector																																	
9	Sovereign and secure data transfer through standardized communication protocols																																	
10	Establishment of terms of use																																	
11	Development of a data contract																																	
12	Offering a data catalog																																	
13	Offering a dashboard for an overview of the data used																																	
14	Monitoring data transfer and capacities																																	
15	Wizard for data offers																																	
16	Information on the data connection																																	
		<table><tr><th>Service</th><th>Yes/No</th></tr><tr><td>Compliance</td><td>✓</td></tr><tr><td>Identity Management</td><td>✓</td></tr><tr><td>Usage Policies</td><td>✓</td></tr><tr><td>Usability</td><td>✓</td></tr><tr><td>Data Processing</td><td>✗</td></tr><tr><td>Community Building</td><td>✓</td></tr><tr><td>Data Control</td><td>✓</td></tr><tr><td>Consent Management</td><td>✓</td></tr><tr><td>Security</td><td>✓</td></tr><tr><td>Data Quality</td><td>✗</td></tr><tr><td>Infrastructure</td><td>✓</td></tr><tr><td>Support/ Knowledge</td><td>✓</td></tr></table>	Service	Yes/No	Compliance	✓	Identity Management	✓	Usage Policies	✓	Usability	✓	Data Processing	✗	Community Building	✓	Data Control	✓	Consent Management	✓	Security	✓	Data Quality	✗	Infrastructure	✓	Support/ Knowledge	✓						
Service	Yes/No																																	
Compliance	✓																																	
Identity Management	✓																																	
Usage Policies	✓																																	
Usability	✓																																	
Data Processing	✗																																	
Community Building	✓																																	
Data Control	✓																																	
Consent Management	✓																																	
Security	✓																																	
Data Quality	✗																																	
Infrastructure	✓																																	
Support/ Knowledge	✓																																	

Business model improvement options (4)	<b>Option 1: Data Quality (Data Verification)</b>	<b>Option 2: Data Processing</b>
	Data quality is an important part of data sharing. One option for expanding the business model based on the pattern could be to focus on data quality through data verification. The same data format is essential and should be implemented on the provider side with the support of a DI.	Data processing is essential, especially regarding the implementation of data sharing in various use cases. The processing of data, especially the visualization, can be a beneficial extension to the existing business model.

patterns are not specifically tailored to the DGA, the de facto reality of its existence places our patterns as starting points for discussions in organizations. **Second (2)**, some combinations of services occur more frequently than others — both in their distribution in the sample and in the spectrum of services each pattern covers. **Third (3)**, there are different concentrations of companies in the clusters. Some clusters are more strongly represented, while other clusters have fewer companies. **Fourth (4)**, our research has resulted in a set of findings that have an impact on current research on data ecosystems and the data economy.

### Data intermediation services and the DGA

Our research provides an overview of potential DISs grounded in the literature corpus and the examination of real-world cases. While we cannot answer which DIS falls under the DGA, organizations should reflect on the services they offer and assess whether they fall under the DGA to prevent potential sanctions if they do not comply, even unknowingly. Given that this paper does not provide legal consultancy or expertise, we leave it to the practitioners to draw from our framework and investigate their compliance with the DGA. Instead, we would like to show that there is a range of different services that could fall under the DGA due to the broad definition of DISs. This includes not only the services offered by large corporations but also smaller companies. If covered by the DGA, the DI must follow specific criteria in its activities (European Commission, 2023):

- No monetarization of the data
- Ensure neutrality
- Avoid conflicts of interest
- Structural separation between the data intermediation service and other services provided
- No dependence between commercial terms (incl. pricing) for the provision for usage of other services
- Usage of data and metadata acquired only to improve the data intermediation services

Our analysis shows it depends on the other underlying conditions under which the services are provided. Services that directly affect the data, e.g., provision or further processing of the data, such as de-identification or analysis of the data, are more affected by the DGA than general sovereignty services or services that affect the network. Here, companies must precisely define the extent to which they will continue to use the provided data and ensure that this data is not subsequently monetized. In general, a clear separation must be created between DIS and the company's other activities. Aspects of the DGA, such as

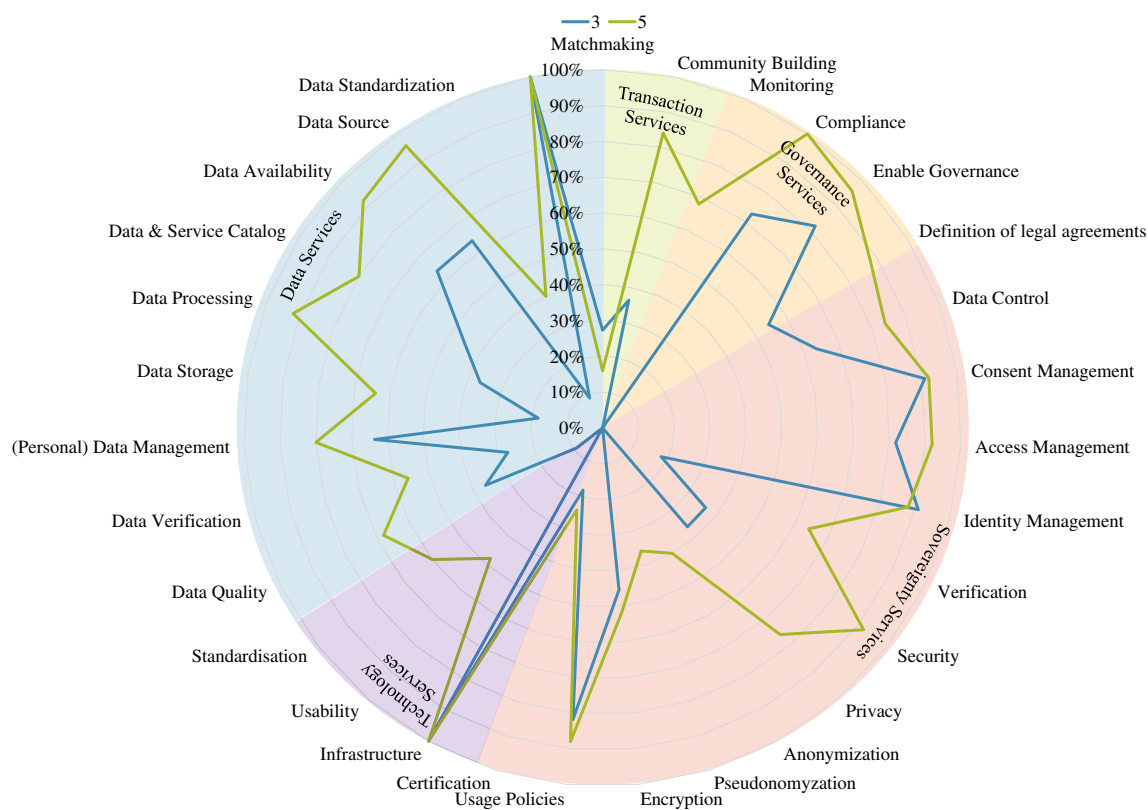
avoiding conflicts of interest and maintaining neutrality, are not related to the services offered but are the responsibility of the company offering them. For companies, this means that they must be aware of the restrictions of the DGA and adapt their services to the new circumstances. Another major question that arises is how to implement the DGA in practice. The requirement for companies to be certified means that lean processes must be created that make it easy to understand how to become a certified DI. Since the DGA entry into force in September 2023, only two DIs have been certified to date (May 2024) (European Commission, 2024b).

This means that existing service offerings must be checked for conformity, but also that the potential for new service opportunities must be recognized. The number of DIs registered with the EU rose from two to eight between May and June (European Commission, 2024b). Relatively speaking, this number seems small at first and raises the question as to why so few companies are currently registered. Is it due to concerns about overly strict monitoring by the EU, or do companies not yet recognize themselves as DIs that fall under the DGA? Is there a lack of guidelines to support companies in the process? The EU is now also calling on companies to comply with the DGA (European Commission, 2024a). This is consistent with the fact that relatively few companies have been registered as DIs to date.

### Distribution of the patterns

The patterns differ in the *composition* of the services and in the *intensity* of how the services are represented within the cluster. There are service clusters that focus on offering a *small selection* of services, such as the provision of infrastructure (see Fig. 7 in blue), and service patterns that provide a *wide range* of different services (see Fig. 7Err or ! **Reference source not found.** in green). We show that there is a broad spectrum of service distributions and that variants of specialization or a general service offering are possible. We visualize these two patterns to create awareness of the distinct characteristics of the patterns and the spectrum in between. Pattern 3 has the lowest number of typical services, and pattern 5 has the most services (based on our analysis, see Fig. 5). **Pattern 3** represents a DI that provides the infrastructure required for data sharing. This essentially includes the technical provision of the infrastructure components, as well as the regulation of who can access the infrastructure and whether it is compliant with the applicable regulations. **Pattern 5** offers a wide variety of services. We decided to compare these two patterns since they exposit the most significant differences in their design (see Fig. 7). All visualized patterns are in the appendix (see Appendix 6).

Comparing the service patterns results in a categorization based on their design. For this, we sorted the patterns

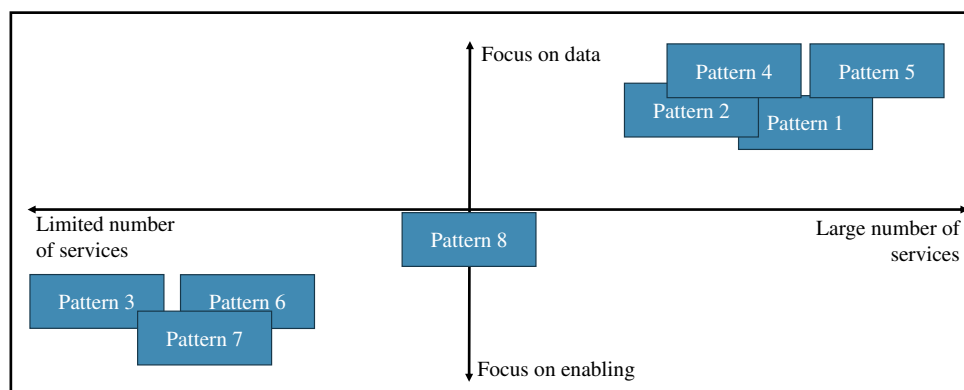


**Fig. 7** Visualization and comparison of patterns 3 and 5

according to the density of services and thematic focus. For the latter, we differentiated between *data-as-a-service* and *enabling-data-sharing-as-a-service*. Investigating the distribution resulted in a two-by-two matrix (see Fig. 8). The distinction of the patterns in the visualization is a qualitative separation for visualization purposes that is not based on a determinate scale. The aim is to visualize how close the patterns are in relation to their nearest neighbors, i.e., similar patterns, and how they contrast in difference with those “further away.” Patterns 3, 6, and 7 primarily concentrate on providing technology components (e.g., infrastructure or

identity management) and the development of overarching data-sharing networks. Typically, other data-related services are not involved. Patterns 1, 2, 4, and 5 emphasize the provision of data catalogs and other data-related services, such as (personal) data management services and data processing. Since infrastructure and support services are mandatory for DIs (see above), they are also included in their service configuration. Pattern 8, in this categorization, is the “outlier” as it is positioned between the clusters explained above. Subsequently, organizations in this cluster primarily offer data sharing enabling processes and individual

**Fig. 8** Comparison of the pattern



data-related services such as data quality and data processing as complements.

Given this distribution, we propose a few reasons to explain this distribution. Based on our sample and observations, we find that, potentially:

- As *specialists*, the companies focus on a specific service or specific service combinations (e.g., Gebäudedaten.ch (2024) and DIABETES.SERVICES (2024) from pattern 3).
- *Startups* or *pioneers* initially start with a small range of services and only gradually plan to expand their range of services (e.g., soivity (2024) and Catena-X (2024) from pattern 8).
- *Established companies* already offer a wide range of services (e.g., CDQ (2024) and Bundesdruckerei GmbH (2024) from pattern 5).

Given the rapid development of DIs and the evolving landscape, the classification should not be static. Our evaluation with soivity has shown that the company, which was founded as a start-up, is still actively working on constantly expanding its range of services. This is not an isolated case. Pioneers like Catena-X's lighthouse projects are constantly being developed and expanded.

### Specimen in each cluster

The number of companies varies significantly between the patterns. Some patterns are smaller with less than ten companies, and other patterns are larger with more than ten companies. Figure 9 shows the distribution of the companies in the patterns. The median average number of companies in the clusters is 10.75. Patterns 3, 5, and 6 are above the median, and patterns 1, 2, 4, 7, and 8 have fewer companies than the median. While the number of services was previously clearly indicated by two thematic focuses (data-driven and enabling-driven), no thematic reason is initially apparent here. A closer analysis of the service patterns shows that patterns 3 and 6 tend to offer specific but frequently required services. These include identity services and the provision of infrastructure. These services are often required for data-sharing processes. Patterns 1, 2, 4, 7, and 8 tend to have more specific service combinations. This explains the lower number of companies within the patterns, as specialization quickly excludes other patterns.

### DISs, data ecosystems, and the data economy in the EU

Existing research on data ecosystems spurred a rich literature on the foundational triangle of data providers, data

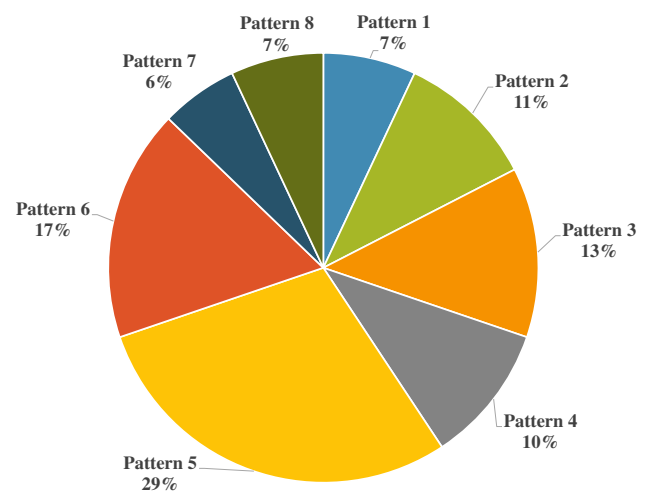


Fig. 9 Comparison of the count of the companies in the patterns

consumers, and data intermediaries (Oliveira & Lóscio, 2018). Some research has provided in-depth contextualization with a specific type of data infrastructure, such as data spaces (e.g., Otto & Jarke, 2019) or data marketplaces (e.g., Bergman et al., 2022). Our work continues and complements this research and focuses on the services offered by the DI. We show that there is currently a wide range of possible services that DIs can offer which provides an overview of how the market is currently positioned with DIs. This also allows conclusions to be drawn about the identification of potential gaps and how existing business models can be expanded to cover the widest possible range of needs. In further research, these patterns can be mirrored against the legal situation in order to check whether the existing DIs and DISs also play the role that they should ideally play in the data economy in the sense of the EU. Due to a large number of potential services and its unique position in the ecosystem as a mediator, “data intermediation services are expected to play a key role in the data economy” (European Commission 2022, L 152/10) as (regulated) digital platforms specialized in data sharing. This is especially relevant against the background of the continuously developing data economy. DIs (as intended by the DGA) are poised to provide neutral and trusted data intermediation services that enable more (potential) data consumers access to data (Carovano & Finck, 2023). Subsequently, DIs hold a critical position at the heart of the data economy, as they could accelerate and facilitate the access and (re-)use of its foundational resource — data (e.g., Micheli et al., 2020). Resulting, our work supports organizations in managing the “hype” around DIs, which inevitably will mean that getting accustomed to DISs and DIs is not optional but mandatory. Participating for value and considering legal compliance with the DGA makes it essential for companies to consider if they could fall under the DGA. For instance, some of the most prominent barriers to engaging in

data sharing are *trust* and *technology* (e.g., Ebel et al., 2021). Both could be mediated or at least lessened when choosing the right DI providing fitting DISs.

## Contributions, limitations, and outlook

The paper has multiple contributions. For the **literature on data ecosystems** (e.g., Oliveira & Lóscio, 2018), we contribute an overview of a specific role in data ecosystems — the *data intermediary* — and provide an abstracted, idealized type that subsumes all relevant services based on our analysis. The underlying contribution is the shared understanding of what is possible for DIs and the more detailed consideration of DISs. The dual nature summarizes and details our understanding of data intermediaries simultaneously. While prior research outlined different types of data intermediaries (e.g., Ditzfurth & Lienemann, 2022; Jussen et al., 2024a), our work extends this by proposing a comprehensive set of DISs that are decoupled from specific DI instances. The term DI has been used in many ways in the literature to date. The presented definitions show that DIs are relevant for a wide variety of stakeholder groups. We offer a specialized view of the possible services of DIs that can benefit all interested stakeholders, regardless of the industry. DIS is still a very new topic in the IS domain. There are already some initial papers that look into DIs, but often with a strong focus on legal issues (e.g., Carovano & Finck, 2023; Vogel, 2022; or Richter, 2023).

Our evaluation with soviety indicates contributions to **managers**. We provide an assessment tool for DIs and DISs, giving them the opportunity to add to their business model or identify extensions of existing services. We offer initial approaches that show the extent to which data providers and data consumers benefit from DIs. For example, initial boundaries can be overcome by compensating for the lack of know-how among the players in DIs and thus enabling secure data sharing. Significantly, the DGA sets the requirements for DIs and forces organizations to deal with these regulations, especially considering the broad definition it proposes. While our paper naturally does not give legal advice, it can help assess managers whether they might fall under the DGA and subsequently have to fulfill a host of requirements. In particular, we provide a more detailed view of specific services (e.g., technology) and subsequently make this assessment possible.

For **policymakers**, we offer insights into practice and research. This can help gain an understanding of how DIS practice is currently structured and identify possible gaps where companies need support with implementation. We offer a starting point that policymakers can use to support companies (e.g., new research projects or workshops). Our

archetypes help policymakers to understand what is possible in practice and with which functions. Based on these findings, regulations can be adapted or aligned, for example, particularly regarding the possibility of companies being certified as DIs by the EU.

Our research has several **limitations**. The taxonomy is based on a systematic review of the academic literature and a sample of DISs (i.e., companies). While we aimed to follow systematic procedures and leveraged the team of authors to mitigate issues of interpretation and data selection, some limitations naturally remain. For one, we aimed to collect a comprehensive sample of the literature exploring DISs, but we cannot guarantee that we have overlooked individual articles. Additionally, we drew the literature from standard data bases in IS research (see the research design), which could be extended by using additional data bases and complementary search terms. Our sample of companies can only cover a subsection of the total (potential) DIs on the market. Subsequently, we had to “make a cut” at some point and identify a suitable sample for our research, which we had to assess based on our (i.e., the team of authors) understanding of what a DIS is. To ensure the most neutral analysis possible, we conducted the literature review with two authors and discussed the results with a team of three authors. Due to the growing interest in DIs, we expect a sharp increase in the amount of relevant literature and DIs on the market. To sum it up, our research only shows a snapshot of time at a point in which the DI concept is emerging (e.g., due to the DGA). Building on our findings with a similar study in the near future could provide a valuable update. Two authors carried out the analysis of the companies, and the results were discussed and determined by a team of all authors. The same applies to the cluster analysis. Furthermore, we are not experts in legal matters and cannot make any legally binding statements about the extent to which the services we mentioned are influenced by the DGA, which is still possible. In the future, we need interdisciplinary research contributions from legal experts to be able to make legally binding statements about the offered services.

Our research offers some starting points for **further research**. The database can be expanded to include more in-depth insights through workshops or interviews with experts from research and practice, enabling a more concrete and detailed description of the individual services. Future research contributions should focus on the design of the necessary business models for DIs. Now, our research only represents a part of a DI business model. Further research could, e.g., focus on the monetization of services. Ultimately, it is the task of an intermediary to overcome barriers to data sharing (see Jussen et al., 2023b). To achieve this goal, it is the task of the research community to identify the necessary mechanisms and make them usable for the industry.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s12525-024-00729-9>.

**Acknowledgment** This research and development project is funded by the German Federal Ministry of Education and Research (BMBF) within the Incentives and Economics of Data Sharing Funding Action (IEDS004). The author is responsible for the content of this publication.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Agahari, W., Dolci, R., & Reuver, M. de (2021). Business model implications of privacy-preserving technologies in data marketplaces: the case of multi-party computation. In *Proceedings of the 29th European Conference on Information Systems*, A Virtual AIS Conference.
- Agdatahub (Ed.). (2024). *European operator of consents and data exchange platform*. <https://agdatahub.eu/en/>.
- AWS Data Exchange. (2023). *AWS Data Exchange is on a mission to increase speed to value for third-party data sets in the cloud*. [https://aws.amazon.com/data-exchange/?nc1=h\\_ls&adx-cards2.sort-by=item.additionalFields.eventDate&adx-cards2.sort-order=desc](https://aws.amazon.com/data-exchange/?nc1=h_ls&adx-cards2.sort-by=item.additionalFields.eventDate&adx-cards2.sort-order=desc) last accessed: 13.05.2024.
- Azkan, C., Möller, F., Ebel, M., Iqbal, T., Otto, B., & Pöppelbuss, J. (2022). Hunting the treasure: Modeling data ecosystem participant value. In *Proceedings of the 43rd International Conference on Information Systems*, Copenhagen: Denmark.
- Bailey, J. (1996). The Emergence of electronic market intermediaries. In *Proceedings of the 16th International Conference on Information Systems*, Cleveland: USA.
- Bailey, K. D. (1994). *Typologies and taxonomies: An introduction to classification techniques (07–102)*. Sage University Paper series on Quantitative Applications in the social Sciences.
- Bastiaansen, H., Dalmolen, S., Kollenstart, M., & van Engers, T. M. (2020). User-centric network-model for data control with interoperable legal data sharing artefacts. In *Proceedings of the 24th Pacific Asia Conference on Information Systems*, Dubai: UAE.
- BDT. (2024). *BDT (Benefits Data Trust): Helping people today, modernizing benefits access for tomorrow*. <https://bdtrust.org/> last accessed: 05.03.2024.
- Bergman, R., Abbas, A. E., Jung, S., Werker, C., & de Reuver, M. (2022). Business model archetypes for data marketplaces in the automotive industry. *Electronic Markets*, 32(2), 747–765. <https://doi.org/10.1007/s12525-022-00547-x>
- Bhargava, H. K., Choudhary, V., Krishnan, R., & Lee, K. B. (1999). Analysis of an intermediary's strategies in electronic markets for decision technologies. In *Proceedings of the 19th International Conference on Information Systems*, Charlotte: USA.
- Bitkom e.V. (Ed.). (2023). *Unternehmen wollen Daten nutzen, aber nicht teilen*. [https://www.bitkom.org/Presse/Presseinformation/Datenoekonomie-Unternehmen-nutzen-Daten#\\_](https://www.bitkom.org/Presse/Presseinformation/Datenoekonomie-Unternehmen-nutzen-Daten#_). Accessed 8 Aug 2024
- Bouwman, H., Faber, E., & van der Spek, J. (2005). Connecting future scenarios to business models of insurance intermediaries. In *18th Bled eConference eIntegration in Action*, Slovenia.
- Bundesdruckerei GmbH. (2024). *Bundesdruckerei GmbH: Data trustee: Data trustee platform with a trust center service on demand*. <https://www.bundesdruckerei-gmbh.de/en/solutions/data-trustee> last accessed: 05.02.2024.
- Candelon, F., Bergues, G. S. de, Martinez, D. Z., Shekar, H. C., & Aguiar, M. (2024). *The benefits of sharing now outweigh the risks*. <https://www.bcg.com/publications/2024/the-benefits-of-data-sharing-now-outweigh-the-risks> last accessed: 07.05.2024.
- Capgemini Research Institute (Ed.). (2021). *Data sharing masters: How smart organizations use data ecosystems to gain an unbeatable competitive edge*. <https://www.capgemini.com/de-de/insights/research/data-sharing-masters/>. Accessed 22 Feb 2023
- Carovano, G., & Finck, M. (2023). Regulating data intermediaries: The impact of the Data Governance Act on the EU's data economy. *Computer Law & Security Review*, 50, 105830. <https://doi.org/10.1016/j.clsr.2023.105830>
- Catena-X. (2024). *Catena-X: Your automotive network*. <https://catena-x.net/en/> last accessed: 26.01.2024.
- CDQ. (2024). *CDQ*. <https://www.cdq.com/> last accessed: 05.03.2024.
- Centre for Data Ethics and Innovation. (2021). *Unlocking the value of data: Exploring the role of data intermediaries: An exploration of the role intermediaries could play in supporting responsible data sharing*. [https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment\\_data/file/1004925/Data\\_intermediaries\\_-\\_accessible\\_version.pdf](https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1004925/Data_intermediaries_-_accessible_version.pdf). Accessed 13 Mar 2024
- Chrusciel, D., & Zahedi, F. (1999). Seller-based vs. buyer-based internet intermediaries: A research design. In *Proceedings of the 5th Americas Conference on Information Systems*, Milwaukee: USA.
- Cichy, P., Salge, T.-O., & Kohli, R. (2021). Privacy concerns and data sharing in the Internet of Things: Mixed methods evidence from connected cars. *MIS Quarterly*, 45(4), 1863–1892. <https://doi.org/10.24451/ARBOR.15064>
- CitizenMe. (2024). *CitizenMe: The data platform for humanity*. <https://www.citizenme.com/> last accessed: 14.02.2024.
- Comuny. (2024). *Comuny: Empowering personal data*. <https://www.comuny.de/warum-comuny/> last accessed: 05.02.2024.
- Corrold, T. (1878). Discovery of the intermediary host of Filaria sanguinis hominis (F. Bancrofti). *The Lancet*, 111(2837), 69. [https://doi.org/10.1016/S0140-6736\(02\)42528-7](https://doi.org/10.1016/S0140-6736(02)42528-7)
- Cozy. (2024). *Cozy: A smart personal cloud to gather all your data*. <https://cozy.io/en/> last accessed: 28.02.2024.
- Culot, G., Orzes, G., Sartor, M., & Nassimbeni, G. (2024). The data sharing conundrum: Revisiting established theory in the age of digital transformation. *Supply Chain Management: An International Journal*, 29(7), 1–27. <https://doi.org/10.1108/SCM-07-2023-0362>
- Dalmolen, S., Bastiaansen, H., Kollenstart, M., & Punter, M. (2019). Infrastructural sovereignty over agreement and transaction data ('metadata') in an open network-model for multilateral sharing of sensitive data. In *Proceedings of the 40th International Conference on Information Systems*, Munich: Germany.
- Data for Good Foundation. (2024). *Data for Good Foundation*. <https://dataforgoodfoundation.org/> last accessed: 07.02.2024.
- DATA SENTINEL. (2024). *DATA SENTINEL: Advanced automation to enable data trust*. <https://www.data-sentinel.com/> last accessed: 05.02.2024.

- Data Sharing Coalition (Ed.). (2020). *Obtaining green loans by sharing data*. <https://datasharingcoalition.eu/use-cases/obtaining-green-loans-by-sharing-energy-data/>. Accessed 10 Nov 2023
- DataGuard. (2024). *DataGuard: Mitigate risks, boost compliance*. <https://www.dataguard.de/en-de/> last accessed: 05.02.2024.
- Datavillage. (2024). *Datavillage: Share insights not data*. <https://www.datavillage.me/> last accessed: 16.02.2024.
- Dawex. (2023). *Data exchange platform*. <https://www.dawex.com/de/> last accessed: 05.03.2024.
- Demchenko, Y., Cushing, R., Los, W., Grosso, P., Laat, C. de, & Gommans, L. (2019). Open data market architecture and functional components. In *Proceedings of the 17th International Conference on High Performance Computing & Simulation*, Dublin: Ireland
- DIABETES.SERVICES. (2024). *DIABETES.SERVICES: More data, better treatment*. <https://diabetes.services/> last accessed: 05.02.2024.
- Ebel, E., Kasah, T., Koerber, B., Mohr, N., Müllerklein, D., & Roggen-dorf, M. (2021). *Overcoming barriers to multiter supplier col-laboration*. McKinsey. <https://www.mckinsey.com/capabilities/operations/our-insights/overcoming-barriers-to-multiter-suppl-ier-collaboration>.
- Eisenmann, T., Parker, G., & van Alstyne, M. (2011). Platform envel-opment. *Strategic Management Journal*, 32(12), 1270–1285. <https://doi.org/10.1002/smj.935>
- European Commission (Ed.). (2020). *Regulation of the European Par-liament and of the Council: On European data governance (Data Governance Act)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020PC0767&from=DE>. Accessed 6 Mar 2024
- European Commission. (2022). *Regulation (EU) 2022/868 of the Euro-pean Parliament and of the Council of 30 May 2022 on Euro-pean data governance and amending Regulation (EU) 2018/1724 (Data Governance Act)*. <https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:32022R0868>
- European Commission. (2023). *Data Governance Act explained*. <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act-explained#ecl-inpage-14ihlqt9> last accessed: 13.11.2023.
- European Commission. (2024a). *Commission calls on 18 Member States to comply with the EU Data Governance Act*. <https://digit-al-strategy.ec.europa.eu/en/news/commission-calls-18-member-states-comply-eu-data-governance-act> last accessed: 13.08.2024.
- European Commission. (2024b). *EU register of data intermediation services*. <https://digital-strategy.ec.europa.eu/en/policies/data-intermediary-services> last accessed: 20.03.2024.
- European Commission. (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regula-tion)*. <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> last accessed: 06.03.2024.
- Fassnacht, M., Benz, C., Heinz, D., Leimstoll, J., & Satzger, G. (2023). Barriers to data sharing among private sector organizations. In *Proceedings of the 56th Hawaii International Conference on System Sciences*, Hawaii: USA.
- Figueredo, K., Seed, D., & Wang, C. (2020). A scalable, standards-based approach for IoT data sharing and ecosystem monetization. *IEEE Internet of Things Journal*, 9(8), 5645–5652. <https://doi.org/10.1109/JIOT.2020.3023035>
- Findy. (2024). *Findy: Verifiable data network*. <https://findy.fi/en/front-page/> last accessed: 01.03.2024.
- Fischer, M., Heim, D., Hofmann, A., Janiesch, C., Klima, C., & Winkelmann, A. (2020). A taxonomy and archetypes of smart services for smart living. *Electronic Markets*, 30(1), 131–149. <https://doi.org/10.1007/s12525-019-00384-5>
- Flanagan, A. J., & Warren, S. (2022). *Advancing digital agency: The power of data intermediaries*. [https://www3.weforum.org/docs/WEF\\_Advancing\\_towards\\_Digital\\_Agency\\_2022.pdf](https://www3.weforum.org/docs/WEF_Advancing_towards_Digital_Agency_2022.pdf). Accessed 16 Sep 2023
- Fuerstenau, D., & Auschra, C. (2016). Open digital platforms in health care: Implementation and scaling strategies. In *Proceedings of the 27th International Conference on Information Systems*, Dub-lin: Ireland.
- Gebäudedaten.ch. (2024). *Gebäudedaten: Gebäudeinformationen auf Knopfdruck*. <https://www.gebaeuedaten.ch/> last accessed: 13.02.2024.
- Geens. (2024). *Why Geens?* <https://geens.com/> last accessed: 08.02.2024.
- German Centre for Cancer Registry Data. (2013). *German Centre for Cancer Registry Data: Tasks*. [https://www.krebsdaten.de/Krebs/EN/Content/ZfKD/Tasks/tasks\\_node.html](https://www.krebsdaten.de/Krebs/EN/Content/ZfKD/Tasks/tasks_node.html) last accessed: 13.02.2024.
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualita-tive rigor in inductive research. *Organizational Research Meth-ods*, 16(1), 15–31. <https://doi.org/10.1177/1094428112452151>
- Hein, A., Schrieck, M., Riasanow, T., Setzke, D. S., Wiesche, M., Böhm, M., & Krcmar, H. (2020). Digital platform ecosystems. *Electronic Markets*, 30(1), 87–98. <https://doi.org/10.1007/s12525-019-00377-4>
- Hellbrück, R. P. (2011). *Angewandte Statistik mit R: Eine Einführung für Ökonomen und Sozialwissenschaftler* (2., überarb. Aufl.). Lehrbuch. Gabler.
- HIE of ONE. (2021). *HIE of ONE: Managing personal health informa-tion shouldn't be so hard*. <https://hieofone.com/> last accessed: 29.02.2024.
- Hoffman, W., Bick, R., Boral, A., Henke, N., Olukoya, D., Rifai, K., Roth, M., & Youldon, T. (2019). *Collaborating for the common good: Navigating public-private data partnerships*. <https://www.mckinsey.com/capabilities/quantumblack/our-insights/collaborat-ing-for-the-common-good> last accessed: 15.08.2024.
- Huang, L., Dou, Y., Liu, Y., Wang, J., Chen, G., Zhang, X., & Wang, R. (2021). Toward a research framework to conceptualize data as a factor of production: The data marketplace perspective. *Funda-mental Research*, 1(5), 586–594. <https://doi.org/10.1016/j.fmre.2021.08.006>
- Ichihashi, S. (2021). Competing data intermediaries. *The RAND Jour-nal of Economics*, 52(3), 515–537. <https://doi.org/10.1111/1756-2171.12382>
- iGrant.io. (2024). *iGrant.io: Your data, your choice: Infrastructure for data exchange and verification services with digital wallets*. <https://igrant.io/> last accessed: 08.02.2024.
- Janssen, H., & Singh, J. (2022). Data intermediary. *Internet. Policy Review*, 11(1), 1–9. <https://doi.org/10.14763/2022.1.1644>
- JLINC. (2022). *JLINC: Interoperable data exchange*. <https://www.jlinc.com/> last accessed: 27.02.2024.
- Joint Research Centre. (2023). *Data intermediaries for more inclu-sive data governance: How do they work?* [https://joint-resea-rch-centre.ec.europa.eu/jrc-news-and-updates/data-intermedia-ries-more-inclusive-data-governance-how-do-they-work-2023-10-04\\_en](https://joint-resea-rch-centre.ec.europa.eu/jrc-news-and-updates/data-intermedia-ries-more-inclusive-data-governance-how-do-they-work-2023-10-04_en) last accessed: 10.02.2024.
- Jussen, I., Fassnacht, M., Schweihoff, J., & Möller, F. (2024a). Reach-ing for the stars: exploring value constellations in inter-organiza-tional data sharing. In *Proceedings of the 32nd European Confer-ence on Information Systems*, Paphos, Cyprus.
- Jussen, I., Möller, F., Schweihoff, J., Gieß, A., Giussani, G., & Otto, B. (2024b). Issues in inter-organizational data sharing: Findings from practice and research challenges. *Data & Knowledge Engi-neering*, 150, 1–19. <https://doi.org/10.1016/j.datak.2024.102280>
- Jussen, I., Schweihoff, J., Dahms, V., Möller, F., & Otto, B. (2023a). Data sharing fundamentals: Definition and characteristics.

- In *Proceedings of the 56th Hawaii International Conference on System Sciences*, Hawaii: USA.
- Jussen, I., Schweihoff, J., & Möller, F. (2023b). Tensions in inter-organizational data sharing: Findings from literature and practice. In *Proceedings of the 25th IEEE International Conference on Business Informatics*, Prague: Czech Republic.
- Kim, T., & Talalayevsky, A. (2005). Internet shopping search: A decision theoretic perspective. In *Proceedings of the 11th Americas Conference on Information Systems*, Omaha: USA.
- Kioses, E., Pramatar, K., Doukidis, G., & Bardaki, C. (2007). Measuring the business value of electronic supply chain collaboration: The case of electronic invoicing. In *Proceedings of the 20th Bled eConference*, Bled: Slovenia.
- Kundisch, D., Muntermann, J., Oberländer, A. M., Rau, D., Röglinger, M., Schoormann, T., & Szopinski, D. (2021). An update for taxonomy designers: Methodological guidance from information systems research. *Business & Information Systems Engineering*, 64, 421–439. <https://doi.org/10.1007/s12599-021-00723-x>
- Kurtz, C., Wittner, F., Semmann, M., Schulz, W., & Böhm, T. (2019). The unlikely siblings in the GDPR family: A technological analysis of major platforms in the diffusion of personal data in service ecosystems. In *Proceedings of the 52nd Hawaii International Conference on System Sciences*, Hawaii: USA.
- Langford, J., Poikola, A., Janssen, W., Lähteenoja, V., & Rikken, M. (2022). *Understanding MyData operators*. <https://mydata.org/wp-content/uploads/2020/04/Understanding-Mydata-Operators.pdf>. Accessed 13 Mar 2024
- Lefebvre, H., Legner, C., & Teracino, E. A. (2023). *5 pillars for democratizing data at your organization*. <https://hbr.org/2023/11/5-pillars-for-democratizing-data-at-your-organization> last accessed: 21.04.2024.
- Legnere, H., & Hamer, A.-P. (2023). The emergence of data sharing along complex supply chains. *International Journal of Operations & Production Management*, 44(1), 292–297. <https://doi.org/10.1108/IJOPM-11-2022-0729>
- Levy, Y., & Ellis, T. (2006). A systems approach to conduct an effective literature review in support of information systems research. *International Journal of an Emerging Transdiscipline*, 9, 181–212. <https://doi.org/10.28945/479>
- Lipovetskaja, A., Ciftci, S. A., Schweihoff, J., Janiesch, C., & Möller, F. (2024). Business models for data trustees. In *Proceedings of the 19th International Conference on Wirtschaftsinformatik*, Würzburg: Germany.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2021). *cluster: Cluster analysis basics and extensions. R package version 2.0.7–1*. <https://cran.r-project.org/web/packages/cluster/cluster.pdf> last accessed: 03.01.2022.
- Marotta, V., Wu, Y., Zhang, K., & Acquisti, A. (2021). The welfare impact of targeted advertising technologies. *Information Systems Research*, 33(1), 131–151. <https://doi.org/10.1287/isre.2021.1024>
- Martens, B., Streel, A. de, Graef, I., Tombal, T., & Duch-Brown, N. (2020). *Business-to-business data sharing: An economic and legal analysis* (JRC Digital Economy Working Paper). European Commission. <https://ssrn.com/abstract=3658100>.
- Merriam-Webster. (2024a). *Definition of archetypes*. <https://www.merriam-webster.com/dictionary/archetypes> last accessed: 18.03.2024.
- Merriam-Webster. (2024b). *Definition of the word 'Intermediary'*. <https://www.merriam-webster.com/dictionary/intermediary> last accessed: 15.05.2024.
- Micheli, M., Ponti, M., Craglia, M., & Suman, A. B. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, 7(2), 1–15. <https://doi.org/10.1177/2053951720948087>
- Micheli, M., Farrell, E., Carballa-Smichowski, B., Posada-Sanchez, M., Signorelli, S., & Vespe, M. (2023). *Mapping the landscape of data intermediaries: Emerging models for more inclusive data governance*. <https://op.europa.eu/en/publication-detail/-/publication/379cf850-4553-11ee-9854-01aa75ed71a1>. Accessed 13 May 2024
- Miro (Ed.). (2024). *The online whiteboard for easy collaboration*. <https://miro.com/online-whiteboard/>. Accessed 6 Mar 2024
- Mobility Data Space. (2024). *Mobility Data Space*. <https://mobility-dataspace.eu/de> last accessed: 26.01.2024.
- Möller, F., Haße, H., Azkan, C., van der Valk, H., & Otto, B. (2021). Design of goal-oriented artifacts from morphological taxonomies: Progression from descriptive to prescriptive design knowledge. In *Proceedings of the 16th International Conference on Wirtschaftsinformatik* (pp. 523–538). Springer, Cham. [https://doi.org/10.1007/978-3-030-86790-4\\_36](https://doi.org/10.1007/978-3-030-86790-4_36)
- Möller, F., Jussen, I., Springer, V., Gieß, A., Schweihoff, J. C., Gelhaar, J., Guggenberger, T., & Otto, B. (2024). Industrial data ecosystems and data spaces. *Electronic Markets*, 34(41), 1–17. <https://doi.org/10.1007/s12525-024-00724-0>
- Möller, F., Stachon, M., Azkan, C., Schoormann, T., & Otto, B. (2022). Designing business model taxonomies – Synthesis and guidance from information systems research. *Electronic Markets*, 32(2), 01–726. <https://doi.org/10.1007/s12525-021-00507-x>
- Moloney, B. (2005). Investigation of the value chain for creative products: A phenomenological study of the perceptions of music intermediaries. In *Proceedings of the 16th Australasian Conference on Information Systems*, Sydney: Australia.
- Murtagh, F. (1984). Counting dendrograms: A survey. *Discrete Applied Mathematics*, 7(2), 191–199. [https://doi.org/10.1016/0166-218X\(84\)90066-0](https://doi.org/10.1016/0166-218X(84)90066-0)
- Murtagh, F., & Contreras, P. (2017). Algorithms for hierarchical clustering: An overview, II. *WIREs Data Mining and Knowledge Discovery*, 7(6), 1–16. <https://doi.org/10.1002/widm.1219>
- Murtagh, F., & Legendre, P. (2014). Ward's hierarchical agglomerative clustering method: Which algorithms implement ward's criterion? *Journal of Classification*, 31, 274–295. <https://doi.org/10.1007/s00357-014-9161-z>
- MyData. (2023). *The MyData Award is the gold standard for ethical use and sharing of personal data*. <https://mydata.org/participate/awards/> last accessed: 13.03.2024.
- MyDataShare. (2023). *MyDataShare: Empowering European personal data spaces*. <https://www.mydatashare.com/> last accessed: 05.02.2024.
- Mydex. (2024). *Mydex security model*. <https://dev.mydex.org/fyi/security-model.html> last accessed: 09.02.2024.
- Myfairdata. (2020). *Myfairdata: Secure and simplify your digital life*. <https://myfairdata.com/en/> last accessed: 05.02.2024.
- Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems*, 22(3), 336–359. <https://doi.org/10.1057/ejis.2012.26>
- Noorian, Z., Iyilade, J., Mohkami, M., & Vassileva, J. (2014). Trust mechanism for enforcing compliance to secondary data use contracts. In *Proceedings of the IEEE 13th International Conference on Trust, Security and Privacy in Computing and Communications*, Beijing: China.
- Numbers. (2024). *Numbers: Giving you the full picture*. <https://www.numbersprotocol.io/> last accessed: 09.02.2024.
- Nwachok A Koul, A. S., & Morin, J. (2018). Agreements framework for data market ecosystem. In *Proceedings of the 13th Pre-ICIS Workshop on Information Security and Privacy*, San Francisco: USA.
- OCKTO. (2022). *Ockto: Safe, compliant and ethical personal data management*. <https://ockto.eu/en/about-us> last accessed: 05.02.2024.
- Öksüz, A. (2014). Turning dark into white clouds – A framework on trust building in cloud providers via websites. In *Proceedings of the 20th Americas Conference on Information Systems*, Savannah: USA.

- Oliveira, M. I. S., Barros Lima, G. D. 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>
- Oliveira, M. I. S., & Lóscio, B. F. (2018). What is a data ecosystem? In *Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age*, New York: USA.
- onecub. (2024). *Trustworthy data: Data-sharing ecosystems - Governance & infrastructure*. <https://www.onecub.com/> last accessed: 05.04.2024.
- OPEN BANKING. (2024). *What is open banking?* <https://www.openbanking.org.uk/what-is-open-banking/> last accessed: 27.02.2024.
- OpenSAFELY. (2024). *OpenSAFELY: Secure analytics platform for NHS electronic health records*. <https://www.opensafely.org/> last accessed: 29.02.2024.
- Otto, B., Steinbuß, S., Teuscher, A., & Lohmann, S. (2019). *Reference architecture model: Version 3.0*. <https://www.fraunhofer.de/content/dam/zv/en/fields-of-research/industrial-data-space/IDS-Reference-Architecture-Model.pdf> last accessed: 11.09.2021.
- Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: Findings from the International Data Spaces case. *Electronic Markets*, 29(4), 561–580. <https://doi.org/10.1007/s12525-019-00362-x>
- Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77. <https://doi.org/10.2753/MIS0742-1222240302>
- Pensions Dashboards Programme. (2024). *Pensions Dashboards Programme: Access pensions information online, securely and all in one place*. <https://www.pensionsdashboardsprogramme.org.uk/> last accessed: 05.02.2024.
- Perscheid, G., Ostern, N. K., & Moormann, J. (2020). Determining platform governance: Framework for classifying governance types. In *Proceedings of the 41st International Conference on Information Systems*, Hyderabad: India.
- Personium. (2020). *Personium: An interconnectable open source PDS (Personal Data Space) server envisioning world wide web of protected data APIs*. <https://personium.io/en/index.html> last accessed: 05.03.2024.
- RACoon. (2024). *RACoon: Die Radiologie Kooperation im NUM*. <https://racoon.network/> last accessed: 13.02.2024.
- Radosevic, N., Duckham, M., Saiedur Rahaman, M., Ho, S., Williams, K., Hashem, T., & Tao, Y. (2023). Spatial data trusts: An emerging governance framework for sharing spatial data. *International Journal of Digital Earth*, 16(1), 1607–1639. <https://doi.org/10.1080/17538947.2023.2200042>
- Richter, H. (2023). Looking at the Data Governance Act and beyond: How to better integrate data intermediaries in the market order for data sharing. *GRUR International*, 72(5), 458–470. <https://doi.org/10.1093/grurint/ikad014>
- RStudio Team. (2024). *RStudio: Integrated development environment for R*. <http://www.rstudio.com/> last accessed: 18.03.2024.
- Sarker, M. B., Butler, B., & Steinfield, C. (1995). Intermediaries and Cybermediaries: Sarker, Butler and Steinfield. *Journal of Computer-Mediated Communication*, 1(3). <https://doi.org/10.1111/j.1083-6101.1995.tb00167.x>
- Schluss. (2024). *Schluss: Take back control of your data*. <https://schluss.org/> last accessed: 28.02.2024.
- Schmidt, M.-C. (2022). Determination of relational success? On the interrelation of digital industrial platforms and buyer-supplier relationships. In *Proceedings of the 26th Pacific Asia Conference on Information Systems*, Taipei - Sydney (Virtual Conference).
- Schöbel, S. M., Janson, A., & Söllner, M. (2020). Capturing the complexity of gamification elements: A holistic approach for analysing existing and deriving novel gamification designs. *European Journal of Information Systems*, 29(6), 641–668. <https://doi.org/10.1080/0960085X.2020.1796531>
- Schrieck, M., Wiesche, M., & Krcmar, H. (2016). Design and governance of platform ecosystems - Key concepts and issues for future research. In *Proceedings of the 24th European Conference on Information Systems*, Istanbul: Turkey.
- Schweihoff, J., Jussen, I., & Möller, F. (2023). Trust me, I'm an intermediary! Exploring data intermediation services. In *Proceedings of the 18th International Conference on Wirtschaftsinformatik*, Paderborn: Germany.
- Shaharudin, A., van Loenen, B., & Janssen, M. (2023). Towards a common definition of open data intermediaries. *Digital Government: Research and Practice*, 4(2), 1. <https://doi.org/10.1145/3585537>
- Sherer, S. A. (1995). Risk in interorganizational information systems. In *Proceedings of the 1st Americas Conference on Information Systems*, Pittsburgh: USA.
- Smart Connected Supplier Network. (2024). *Smart Connected Supplier Network (SCSN)*. <https://smart-connected.nl/en> last accessed: 26.02.2024.
- Snowflake (Ed.). (2023). *Snowflake: Customer stories: Snowflake Data Cloud*. <https://www.snowflake.com/en/why-snowflake/customers/>
- Sovity (Ed.). (2024). *sovity - Data sovereignty as a service*. <https://sovity.de/de/start/>
- Stachon, M., Möller, F., Guggenberger, T., Tomczyk, M., & Henning, J.-L. (2023). Understanding data trusts. In *Proceedings of the 31st European Conference on Information Systems*, Kristiansand: Norway.
- Startup Commons. (2024). *Startup Commons: Startup ecosystem orchestration*. <https://www.startupcommons.org/> last accessed: 04.03.2024.
- Streamr. (2024). *Streamr: Decentralized data broadcasting*. <https://streamr.network/> last accessed: 09.02.2024.
- Szopinski, D., Schoormann, T., & Kundisch, D. (2019). Because your taxonomy is worth it: Towards a framework for taxonomy evaluation. In *Proceedings of the 27th European Conference on Information Systems*, Stockholm: Sweden.
- Tang, X., Lan, T., Zhong, H., Li, D., & Miao, Q. (2024). An evolutionary game model for indirect data sharing in manufacturing big data consortium. *Expert Systems with Applications*, 255, 124807. <https://doi.org/10.1016/j.eswa.2024.124807>
- Thoring, K., Mueller, R. M., & Badke-Schaub, P. (2020). Assessing and changing an organization's innovation culture with the workspace catalyst canvas. *Proceedings of the 53rd Hawaii International Conference on System Sciences*, 417–426. <https://doi.org/10.24251/HICSS.2020.052>
- Trusted Third Party of the University Medicine Greifswald. (2024). *Trusted third party of the University Medicine Greifswald: The Data Trustees*. <https://www.ths-greifswald.de/en/homepage/> last accessed: 09.02.2024.
- Visions. (2021). *Visions: Connect and access thousands of data and services*. <https://visionspol.eu/en/home-exchange-personal-data/> last accessed: 07.02.2024.
- Vogel, Y. A. (2022). Stretching the limit, the functioning of the GDPR's notion of consent in the context of data intermediary services. *European Data Protection Law Review*, 8(2), 238–249. <https://doi.org/10.21552/edpl/2022/2/10>
- von Ditzfurth, L., & Lienemann, G. (2022). The Data Governance Act: – Promoting or restricting data intermediaries? *Competition and Regulation in Network Industries*, 23(4), 270–295. <https://doi.org/10.1177/17835917221141324>
- Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>

- Weber, M. (1949). *The methodology of the social sciences* (1st ed.). UK: The Free Press. <https://doi.org/10.4324/9781315124445>
- Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, 26(2), xiii–xxiii. <http://dl.acm.org/citation.cfm?id=2017160.2017162>.
- Weking, J., Mandalenakis, M., Hein, A., Hermes, S., Böhm, M., & Krcmar, H. (2020). The impact of blockchain technology on business models – A taxonomy and archetypal patterns. *Electronic Markets*, 30(2), 285–305. <https://doi.org/10.1007/s12525-019-00386-3>
- Wernick, A., Olk, C., & Grafenstein, M. von (2020). Defining data intermediaries - A clearer view through the lens of intellectual property governance. *Technology and Regulation*, 65–81. <https://doi.org/10.26116/techreg.2020.007>
- Westdeutsche Biobank Essen. (2019). *Die Westdeutsche Biobank Essen (WBE): Häufig gestellte Fragen*. <https://www.uni-due.de/med/biobank/faq> last accessed: 09.02.2024.
- Wixom, B. H., Sebastian, I. M., Gregory, R. W., & Piccoli, G. (2023). Level up to strategic data sharing. *MIT Sloan Management Review*, 64(2), 1–6.
- World Data Exchange. (2024). *World Data Exchange: The borderless data exchange platform for personal data ownership and consent-driven on-sharing*. <https://worlddataexchange.com/> last accessed: 03.07.2024.
- World Economic Forum (Ed.). (2021). *6 ways data sharing can shape a better future*. <https://www.weforum.org/agenda/2021/04/six-ways-data-sharing-can-shape-a-better-future/>.
- Woroch, R., & Strobel, G. (2022). Show me the money: How to monetize data in data-driven business models? In *Proceedings of the 17th International Conference on Wirtschaftsinformatik*, Nürnberg: Germany.
- Wysel, M., Baker, D., & Billingsley, W. (2021). Data sharing platforms: How value is created from agricultural data. *Agricultural Systems*, 193, 1–12. <https://doi.org/10.1016/j.agsy.2021.103241>
- Yivi. (2024). *Yivi: Back in charge of your digital data*. <https://www.yivi.app/en> last accessed: 08.02.2024.

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.