DESIGN OPTIONS FOR DATA SPACES

Research Paper

Anna Gieß, Fraunhofer ISST, Germany, anna.giess@isst.fraunhofer.de

Frederik Möller, TU Braunschweig University, Fraunhofer ISST, Germany, f.moeller@tu-braunschweig.de

Thorsten Schoormann, University of Hildesheim, Fraunhofer ISST, Germany, thorsten.schoormann@uni-hildesheim.de

Boris Otto, TU Dortmund University, Fraunhofer ISST, Germany, boris.otto@tu-dortmund.de

Abstract

Data spaces receive considerable attention nowadays and are at the heart of numerous large-scale European research initiatives shaping the data economy. Their goal is to establish secure environments that enable cross-organizational data management and thereby collect, integrate, and make available heterogeneous data from various sources. Although we can observe a great interest in establishing new data spaces, questions of what exactly makes a data space and what it takes to design one remain open. To clarify that, we extracted and organized data space characteristics based on the analysis of 53 papers, as well as an empirical analysis of 47 real-world data spaces. We formalize the findings in a taxonomy to provide an intuitive tool that captures important data space design options. Our paper contributes to the understanding of an emerging artifact with significant implications for business and academia, namely data spaces.

Keywords: Data Sharing, Data Spaces, Conceptualization, Taxonomy.

1 Introduction

Novel digital technologies are a key driver for the digital economy because they produce and disseminate a variety of different data, both from people and organizations (Azkan et al., 2019). A study from 2017 uncovered that for 94% of all companies, the availability of industrial, high-quality IoT data (Internet of Things) is an essential prerequisite to participating in the digital economy and maintaining competitive advantages (Icks et al., 2017). Going a few years further until today, 73% of companies still are not ready to share data to generate value (Azkan et al., 2022). For this, organizations face a variety of barriers, such as missing trust, a lack of clear business value, and the fear of economic damage (Fassnacht et al., 2023). Supporting data sharing is, however, a pivotal component of the European data strategy (European Commission, 2020), making it one of the six priorities of the European Commission until 2024 (European Commission, 2022). As a result, we can observe an increasing interest in projects that develop data spaces in various domains aiming to establish technical infrastructure for secure data sharing. For illustration, the *Mobility Data Space* is dedicated to making mobility data (e.g., timetables, geoinformation, and weather data) available and to sharing this data securely (Mobility Data Space, 2021). Also, the data space seeks to develop mechanisms for monetizing data to provide economic incentives and secure data exchange. Another illustration is the Dutch data space *Smart Connected Supplier Network (SCSN)*, providing an ecosystem for manufacturing companies and suppliers to share data, such as communication and process data. It ensures that each participant (e.g., data providers and data receivers) adhere to a set of minimum standard of a commonly defined set of rules (SCSN, 2022). Examples of data usage rules include policies about the use of data, such as the restriction of time intervals in which a data receiver can use the data set by the data provider (Zrenner et al., 2019). While this concept of sharing data is not new, per se, it has drastically changed due to the availability of digital technologies and the pressure to share data for business purposes (Wallis et al., 2013). Furthermore, data sharing is promoted by the proliferation of communication gadgets that allow easy data transmission via the internet. As a result, conventional databases are becoming more networked and hold various data (e.g., structured and unstructured data, Singh, 2013). Database technologies have impacted enterprise data management development, especially in the last few years (Legner et al., 2020). New data characteristics challenge data management technologies, and traditional database management systems (DBMS) cannot meet these challenges. These novel data management challenges have led to a search for new data management technologies. A promising one is the data space (Curry and Ojo, 2020). To the best of our knowledge, there is no scientific publication showing the key dimensions and characteristics of data spaces yet, to understand them in their totality and in a more general way.

Although there are multiple ways to overcome that issue (e.g., creating a typology or ontology, Bailey, 1994), we chose to design a taxonomy as combining inductive and deductive reasoning allowed us to integrate scientific and practical knowledge into one artifact (Nickerson et al., 2013). By organizing knowledge for a specific research topic, taxonomies typically assist researchers and practitioners in understanding and analyzing complicated subjects (Glass and Vessey, 1995; Nickerson et al., 2013). A taxonomy can also be the foundation for the creation of a comprehensive theory (Williams et al., 2008). We think that the current moment is opportune for adopting an empirically-informed taxonomy-building approach, which leverages the practical insights extracted from a plethora of established data spaces (e.g., IDSA, 2022) and their significant relevance in European research and politics. Our endeavor aims to support both researchers and practitioners alike to understand, analyze, and develop (novel) data spaces by providing a holistic structure for the existing field of research and practice (Glass and Vessey, 1995). The contribution of this paper, in the form of a taxonomy, outlines data space design options that have a definitional character by structuring the concept (Bailey, 1994; Glass and Vessey, 1995). In doing this, we seek to respond to the major challenge concerning questions about how a data space can be conceptualized and designed. Therefore, we raise the following research question (RQ):

RQ: *What are the design options for designing data spaces?*

This paper is structured as follows. In the next section, we briefly state the research background of data spaces. Then we provide an overview of our methodical approach by following Kundisch et al. (2022), including the search, review, and analyze of the literature on data spaces as well as the iterative development and evaluation of a taxonomy. In accordance with the method, we present our main artifact, namely a taxonomy to structure data space design options and explain the captured elements. Afterward, we classify three illustrative use cases with data space experts by means of the taxonomy to demonstrate its applicability. Finally, we conclude our paper by elaborating on our contributions and outlining further research opportunities.

2 Research Context: Data Spaces

Until 1997, the term ‘data space’ was used for interactive 3D visualizations (e.g., Anupam et al., 1995; Petajan et al., 1997; Smotroff et al., 1994). Later, data space research mainly focused on the shared data space, representing the incorporation of database concepts in a programming language (e.g., Busi and Zavattaro, 2001; Cunningham and Gruia-Catalin, 1989; Busi and Zavattaro, 2003; Nagasaka and Motoyama, 2007). Franklin et al. (2005) introduced the term as a management system that collects and contains large-scale heterogeneous data distributed over various data sources in different formats (e.g., structured, semi-structured, and unstructured data). Data spaces are an emerging approach to data management that recognizes the difficulties and expenses of obtaining an upfront unifying schema across all sources in large-scale integration scenarios. Today, data management scenarios rarely enable all the data to fit into a conventional relational database (Franklin et al., 2005). Furthermore, data spaces are defined as a set of relationships underlying a set of participants (Singh and Jain, 2011). Table 1 provides an overview of the most prominent definitions of data spaces.

Definition	Source
"A dataspace […] should contain all of the information relevant to a particular organization regardless of its format and location, and model a rich collection of relationships between data repositories."	Franklin et al. (2005, p. 29)
"Dataspaces are not a data integration approach; rather, they are more of a data co-existence approach. The goal of dataspace support is to provide base functionality over all data sources, regardless of how integrated they are."	Halevy et al. (2006, p. 1)
"A dataspace system manages the large-scale heterogeneous collection of data distributed over various data sources in different formats. It addresses the structured, semistructured, and unstructured data in coordinated manner without presuming the semantic integration among them."	Singh (2013, p. 17)
„Data spaces support data sharing and data sovereignty in ecosystems as they are based on a distributed software infrastructure which provides the required software functionality."	Otto (2022, p. 5)

Table 1. Exemplary definitions of data spaces from literature.

Within a data space, data sources are not tightly controlled, and full semantic integration is not guaranteed (Curry and Ojo, 2020). Unlike data integration over DBMS, a data space does not have complete control over its data and gradually integrates data as necessary (Wang et al., 2016). A data space must cope with data and applications in a wide variety of formats available via many systems with distinct interfaces. The goal of a data space is to provide base functionality over all data sources, regardless if they conform to a specific schema or data constraint (Halevy et al., 2006).

In the past, the term 'data space' referred to an internal data management system, but today we mostly speak of data spaces as an enabler for data sharing between companies (Otto, 2022). This definition also fits our understanding of data spaces as reflected in practice; for example, in emerging reference architectures for data spaces to enable data sharing. Data sharing describes the process of granting access to third parties (e.g., other companies, individuals, or public institutions) data sets in a domain-independent manner. The shared data is frequently utilized to create new services and apps. The data provider expects to receive compensation in the form of money or other benefits (such as data). Depending on the use case, the (legal) agreements between the data producers, data consumers, and other roles govern what the data may be used for and how it is made available (Jussen et al., 2023).

Although the environment of the data spaces changed from internal data management to cross-company data sharing, most concepts were adopted. Data spaces, as opposed to central digital platforms, have a federated architecture and hence enable new options for value creation based on data ecosystems (Beverungen et al., 2022). In data ecosystems, actors interact and work together to find, archive, publish, consume, and reuse data, as well as to stimulate innovation (Oliveira et al., 2019). Prior research already proposes taxonomies for adjacent constructs (i.e., data ecosystem) next to data spaces (see Table 2). These taxonomies, for instance, focus on the governance and interactions within the data ecosystem or economic aspects (e.g., the purpose of participation). While related taxonomies already provide valuable insights, we still need knowledge on how to technically implement data sharing as well as a more holistic overview of design options that is informed by conceptual and empirical data to draw on insights that are available in both academia and practice.

Taxonomy for…	Source	Focus	Grounding
Data ecosystem design space	Curry and Ojo (2020, p. 37)	Governance, economic, and technical aspects at a high level with a limited number of characteristics	Own experiences
Data ecosystems	Gelhaar et al. (2020, p. 6)	Economic aspects of ecosystems	Literature
Ecosystem data governance	Lis and Otto (2021, p. 6070)	Governance and interactions within ecosystems	Literature

Table 2. Overview of related data ecosystem taxonomies.

3 Research Design

Taxonomies are artifacts to describe and classify existing or future objects of a particular domain (e.g., Glass and Vessey, 1995). They allow us to understand, analyze, and examine that domain based on its key characteristics and dimensions, wherefore they are an auspicious tool to advance the field of data spaces too. Taxonomies are widely accepted in research and practice and have been developed in industries and technologies that relate to our study (e.g., Baecker et al., 2021; Poser et al., 2022; Rosian et al., 2022). For designing our taxonomy, we adapted the method proposed by Kundisch et al. (2022) as it allows – in line with the well-established design science paradigm – for rigorous building and evaluating a taxonomy across several design cycles (see Figure 1).

	Problem identification and solution objectives	Design and development	Demonstration and evaluation	Communication
Design cycle 1	• Analysis of prior literature • Meta-characteristic • Ending conditions	• Iteration 1: Conceptual-to-empirical (53 papers) • Iteration 2: Empirical-to-conceptual (18 objects)	• Feedback (on applicability and understandability) from one data space expert	• Representation of taxonomy as morphological box • Report results
Design cycle 2	• Refinements based on feedback from cycle 1	• Iteration 3: Empirical-to-conceptual (29 objects)	• Feedback from one data space expert • Classification of three use cases with three practitioners	

Figure 1. The overall procedure for taxonomy design according to Kundisch et al. (2022).

Following Kundisch et al. (2022), we first specify the problem at hand and outline our research objectives. Then, in the design and development stage, we continue with iterations through two distinct approaches, namely 'conceptual-to-empirical' (deduction), in which the characteristics and dimensions are derived from relevant literature/theory and 'empirical-to-conceptual' (induction) in which real-world objects (i.e., existing data spaces) are analyzed for common characteristics and dimensions (see Table 3). Afterward, we checked if the ending conditions for valid and useful taxonomies were fulfilled, demonstrated the applicability of our taxonomy, and prepared the results for communication. In the detailed description of our procedure below, we refer to *steps 1-18* from Kundisch et al. (2022).

Design Cycle	Iteration	Source	Sample	Approach
Cycle 1	Iteration 1	ACM Digital, AIS eLibrary, IEEE Xplore, JSTOR, Science Direct, and Scopus	n = 53 papers	C2E
	Iteration 2	Data spaces from IDSA (2022)	n = 18 data spaces	E2C
Cycle 2	Iteration 3	Data spaces from EuPro Gigant (2022)	n = 29 data spaces	E2C

Table 3. Overview of input sources.

3.1 Problem identification and solution objectives

Given the above-mentioned limitations in the status quo of scientific literature as well as the increasing interest in establishing novel data spaces *(step 1)*, we aim to conceptualize existing knowledge to advance the understanding of and guide the design of data spaces *(step 3)*. This is important for several potential target groups, including researchers who are interested in theorizing about data spaces as well as practitioners who seek to create, contextualize, and implement concrete data spaces *(step 2)*.

In accordance with the taxonomy-building approach, a meta-characteristic needs to be determined *(step 4)*, which outlines the ultimate purpose of the taxonomy. The meta-characteristic serves as a foundation from which dimensions and characteristics need to be identified (Nickerson et al., 2013). Our meta-characteristic, "design options for data spaces", reflects the fact that we strive to create design options. Next, to determine when to stop the iterative building procedure (i.e., ending conditions), we adopted the ending conditions from Nickerson et al. (2013), with one exception: Following Szopinski et al. (2020a) guidelines for reducing the complexity of the artifact, we refrained from producing a taxonomy that contains only mutually exclusive dimensions *(step 5)*.

3.2 Design cycle 1: initial taxonomy

3.2.1 Design and development

To develop an initial taxonomy *(steps 6-10)*, the taxonomy builder can follow either a conceptual-to-empirical (C2E) approach or an empirical-to-conceptual (E2C) approach *(step 6)*. The C2E approach focuses on conceptualizing characteristics and dimensions before examining the objects, and a taxonomy is created afterward. The E2C approach focuses on identifying subsets of objects and extracting characteristics from the objects before grouping them into a taxonomy. Both approaches can be iteratively used (Nickerson et al., 2013).

At the beginning of the design, to give additional structure to the taxonomy, we employ the concept of meta-dimensions (Möller et al., 2021b). We chose a general framework to ensure that we could include very heterogeneous data spaces. For that purpose, we chose the meta-dimensions of economic, technical, and governance (Curry and Ojo, 2020; Glass and Vessey, 1995). Even if the data space is more of a technical construct, we also conducted the economic and governance part to ensure the different roles and services, which came together with a business model, within a data space were included in our investigations. For example, AI4EU (2022) is offering a matchmaking service for connecting businesses, and SCSN (2022) is offering several services from service providers, which make it easier for companies in the manufacturing industry to connect to the SCSN network.

In the first iteration (C2E), we reviewed the literature on data spaces *(steps 7c-8c)*. Therefore, the first step is an exhaustive structured literature review (Cooper, 1988) following the established guidelines of vom Brocke et al. (2009) and Webster and Watson (2002). Vom Brocke et al. (2009) proposes a procedure along with five activities for defining the review scope, conceptualizing the topic, searching the literature, analyzing the literature, and defining the research agenda. First, we defined the scope of the literature review as papers about data spaces published in established scientific databases of information systems research (ACM Digital, AIS eLibrary, IEEE Xplore, JSTOR, Science Direct, and Scopus). Second, we provided working definitions of data spaces for further research. After a quick scan of the collected literature, we identified the definitions from Table 1 as the most cited ones. Third, the literature search provided 2836 papers with the search string ‘data space’ (see Figure 2).

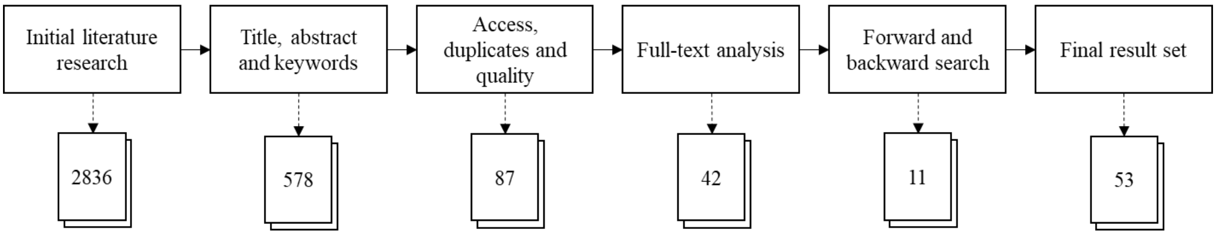

Figure 2. Structured literature review process.

Upon gathering potentially relevant publications, we analyzed and filtered the literature following different exclusion criteria. The term ‘data space’ should either be a keyword, part of the title or mentioned in the abstract, which resulted in 578 papers. All duplicates and papers which were not accessible and not yet published were excluded. Further, we applied several quality aspects as proposed by Cooper (1988) and vom Brocke et al. (2009), such as a publication must be methodologically consistent and argue comprehensibly. We also excluded papers published before 2005 that used the term 'data space' but meant something different from our research addresses. The final literature corpus consisted of 87 papers, which we considered for detailed analysis. Finally, the thematic focus of the paper should be the data space instead of merely mentioning it incidentally. After applying all exclusion criteria, we identified 42 papers as relevant. As proposed by Webster and Watson (2002), we conducted a backward search, which resulted in 11 additional papers. In total, 53 papers were examined.

After examining the papers, we chose the E2C approach for the second iteration *(steps 7e-9e)*. Through the collection and analysis of real-world use cases of data spaces, we aimed to extend our findings from the literature and provide further empirical evidence. For this, we have created a list[1] of 47 data spaces consisting of data space projects known to ourselves, the public repository of the IDSA[2] and EuPro Gigant[3]. We have chosen these two directories because they are the only comprehensive ones known to us at the time of writing. The public repository of the International Data Space Association (the data space radar) is dedicated to fostering data space design (IDSA, 2022). Separately, the EuPro Gigant project to build a cross-site, digitally networked production ecosystem provides an overview of data spaces in the EU. In this iteration, we started with the analysis of the data spaces from the IDSA radar, which resulted in 18 data spaces. We chose this iterative approach because the understanding of the data often develops as the analysis progresses (Srivastava and Hopwood, 2009).

3.2.2 Demonstration and evaluation

To collect insights on the applicability and utility of our initial taxonomy *(steps 15-17)*, we present our intermediary results in three meetings with an average duration of 50 minutes to an expert from a data space umbrella organization. This organization brings together several data spaces and projects to create a digital economy based on the International Data Space (IDS) standard in which all participants can realize the value of their data. Doing this ensured that what we found in the literature and publicly available sources mirrored the 'real' world of data space design. We derived new avenues to analyze the data and sharpened concepts and their wording from this feedback. For example, we added the characteristic 'pseudonymous' to the dimension 'data privacy' or demoted the element 'identity management' into the dimension 'trusted exchange' (see Table 4) *(step 10)*.

Taxonomy operations on taxonomy elements		**Taxonomy element before taxonomy operation**	**Taxonomy element after taxonomy operation**
Add (insert a new element)		Data privacy: Anonymous	Data privacy: Anonymous | Pseudonymous
Update	*Rename* (change the name of an element)	Data processing: Real-time	Data processing: Stream
	Swap (change the order of two elements)	Domain: Cross-domain | Domain-specific	Domain: Domain-specific | Cross-domain
	Merge (join at least two elements into one element)	Data sharing logic: Data marketplace | Data portal | Data platform	Data sharing logic: Data platform
	Demote (move an element to a lower level of abstraction)	Identity management: Level 1 | Level 2 | Level 3	Trusted exchange: Trust by identity management

[1] Full list available upon request

[2] https://internationaldataspaces.org/adopt/data-space-radar/

[3] https://euprogigant.com/wissens-hub/internationale-datenraeume/

Taxonomy operations on taxonomy elements	Taxonomy element before taxonomy operation	Taxonomy element after taxonomy operation
Delete (remove an existing element)	Data sharing policies Mandatory | Voluntary	Data sharing policies Set by data space | Set by data provider

Note: Element of higher order = grey background, element of lower order = white background.

Table 4. Exemplary taxonomy operations, according to Kundisch et al. (2022).

3.3 Design cycle 2: refined taxonomy

3.3.1 Design and development

Based on the insights gathered from the evaluation with an expert in data spaces, and because not all ending conditions were fulfilled in the first cycle *(steps 11-12)*, we conducted another E2C iteration *(steps 7e-9e)*. In this iteration, we engaged with the data spaces from the repository of EuPro Gigant (2022) and further data spaces that we are aware of. Hence, a total of 29 data spaces were consulted.

3.3.2 Demonstration and evaluation

To ensure that our results were valid, we checked the ending conditions again *(steps 11-14)*. Since we were able to meet all 13 ending conditions, we decided to finalize the taxonomy (see Table 5).

Ending conditions		Development				
		Cycle 1			Cycle 2	
		It.1	It.2	F.1	It.3	F.2
Objective	All objects or a representative sample of objects have been examined	-	-	-	●	●
	No object was merged with another or split into multiple ones.	-	-	●	●	●
	At least one object is classified for every characteristic of every dimension.	●	●	●	●	●
	No new dimensions or characteristics were added.	-	-	-	●	●
	Dimensions or characteristics were neither merged nor split.	-	-	-	-	●
	Each dimension is unique and not duplicated.	●	●	●	●	●
	Every characteristic is unique within its dimension.	●	●	●	●	●
	Each cell is unique and not repeated.	-	●	●	●	●
Subjective	Conciseness – no unnecessary dimensions and characteristics	-	-	-	●	●
	Robustness – dimensions and characteristics differentiate objects.	-	-	●	●	●
	Comprehensiveness – all objects can be classified.	-	-	-	●	●
	Extendibility – dimensions and characteristics can be added easily.	-	-	●	●	●
	Explanatory – dimensions and characteristics can describe all objects.	-	-	-	●	●

Note: It. = Design iteration, F. = Feedback.

Table 5. Ending conditions for each iteration, according to Nickerson et al. (2013).

In order to investigate the applicability of the taxonomy *(steps 15-17)*, we performed another feedback round and used the taxonomy with experts to classify real-world cases. We had the chance to present our results to experts from several data space organizations to classify their characteristics in our dimensions to ensure that our taxonomy reflects the reality of data space design options. To realize a heterogenous view, we chose data spaces from different domains. In detail, we chose the Health-X Data Space, the Mobility Data Space, and the Energy Data Space (see Section 5). We spoke with an expert on each data space and went through the taxonomy line by line to see whether the data space could be classified in the taxonomy. On average, a conversation took 30 minutes.

3.4 Communication

To increase the comprehensibility of our taxonomy, we decided to represent it as a morphological box *(step 18)*. This is a typical style of visualization, which is employed in both practical and academic settings wherefore we believe it is easy to understand and use by our target user groups (Szopinski et al., 2020b; Möller et al., 2021b). Following our procedure, we propose a taxonomy based on a literature review of 53 published papers and an empirical analysis of 47 data spaces.

4 A Taxonomy of Design Options for Data Spaces

The final taxonomy consists of 17 dimensions with 50 characteristics (see Table 6). To increase the comprehensibility of our taxonomy, we indicate for each dimension whether its characteristics are mutually exclusive (E) or non-exclusive (N).

MD	Dimension (D_n)	Characteristics ($C_{n,m}$)						E/N
Economic	Domain	Domain-specific			Cross-domain			E
	Funding	Public	Private	Private-public partnership				E
	Data space access	Free			Fee			E
	Reward	Money	Data	Service	Reputation	None		N
	Value added services	Yes			No			E
Technical	Data structure	Structured	Semi-structured	Unstructured				N
	Data type	Raw data	Processed data	Metadata				N
	Data processing	Stream			Batch			N
	Architecture	Centralized	Decentralized	Hybrid				E
	Data sharing logic	P2P data sharing	Data platform	Data sharing via intermediaries				N
	Data harmonization	Data models			Data catalog			N
	Access technology	Standardized connector			Portal			N
	Trusted exchange	Trust by identity management			Trust by certification			N
Governance	Data privacy	Anonymous	Pseudonymous	Non-anonymous	Various			E
	Data classification scheme	Domain	Origin	Topicality	Size	Data format	…	N
	Data sharing policies	Set by data space			Set by data provider			N
	Traceability and control	Space dimension	Time dimension	Use dimension	None			N

Note: E = Exclusive, N = Non-exclusive.

Table 6. Design options for data spaces visualized as a morphological box.

Furthermore, to provide additional structure, we clustered the dimensions into three meta-dimensions:

- The first meta-dimension subsumes dimensions relevant to the **economic** design of data spaces. While principally a technical artifact, designing data spaces requires considering the business model elements and dynamics of competitive environments (Curry and Ojo, 2020).
- The second meta-dimension is **technical**. It relates to the characteristics of the technical architecture of data space and the characteristics of the shared data within, which is crucial to enable data sharing (Otto, 2022).
- The third meta-dimension is **governance**. It focuses on building trust between individual participants and across the entire data space. This is a necessary and significant component of functional and sustainable data space (Fernandez et al., 2020) and is accompanied by the governance characteristics of data ecosystems.

4.1 Meta-dimension: economic

In the first meta-dimension, we find that data spaces differ regarding the **domain (D_1)** in which they are deployed, which naturally impacts what kind of data they specialize in. Examples of this have emerged in different domains **($C_{1,1}$)** (Schahovska, 2011), such as *mobility* (Mobility Data Space, 2021), *logistics* (DE4L, 2022), *agriculture* (agdatahub, 2022), and independently from a certain domain in cross-domain **($C_{1,2}$)** data spaces (Du et al., 2012; EuPro Gigant, 2022). Another example is the E015 digital ecosystem which focuses on one region instead of one domain (Regione Lombardia, 2022).

Given that some data spaces are the results of research projects, and some are based in industry, they differ in how they receive **funding (D_2)**. Frequently, data spaces are the product of public funding **($C_{2,1}$)** in research projects (e.g., by the federal ministry for research, Alonso et al., 2018). Naturally, they can also emerge in the industry through private funding **($C_{2,2}$)** or change funding schemes based on their life cycle (e.g., once public funding has ended, it has to shift to private funding). A prime instance of a private-public partnership **($C_{2,3}$)** is the Germany-based data space *Catena-X Automotive Network*, the funding of which is nearly equally split between public and private funding (Catena-X, 2022).

Users can **access data spaces (D_3)** in two options. On the one hand, access to the data space can be free of charge **($C_{3,1}$)** (Catena-X, 2022; DatenMarktplatz.NRW, 2020) On the other hand, access might require the data space users to pay a fee **($C_{3,2}$)** (agdatahub, 2022; SCSN, 2022). This fee can be paid in the form of a fixed value, a subscription model (Al-Zahrani, 2020), or a usage-based payment model (Muschalle et al., 2013). Next to how to access data spaces, users are concerned with the **rewards (D_4)** they get in exchange for their data or must give to receive third-party data. In terms of payment, the most basic payment technique is a defined reward for a specific amount of data. This might be a set monetary value **($C_{4,1}$)** since basic monetary compensation is versatile (Badewitz et al., 2020). Besides direct monetary compensation, some rewards use bartering (Fernandez et al., 2020). This can be the trading of data for data **($C_{4,2}$)**, and the data providers can be offered some service **($C_{4,3}$)**, or technology in exchange for their data (Woerner and Wixom, 2015). A service can contain everything feasible. For example, agdatahub (2022) offers management, technical, legal, and marketing support. Another type of reward mentioned in the literature is reputation **($C_{4,4}$)**, as possible community recognition (Xie et al., 2020) or appraisal from other users (Thomas and Leiponen, 2016). Lastly, a data provider can make its data freely **($C_{4,5}$)** available, which may be the case when governmental authorities or non-profit organizations are required to share their data or when data providers are attempting to attract new clients (Muschalle et al., 2013; EnDaSpace, 2021). However, the user may have to pay a fee for access to the data space, even if the data itself is free. **Value-added services (D_5)** can be offered **($C_{5,1}$)** within the data space to increase revenue and offer additional business potential. As an example, CDQ (2022) offers data quality as a service, which comes with challenges concerning the development, configuration, and deployment of such offerings (Badewitz et al., 2020). Given the plethora of potential services, our taxonomy only distinguishes between if data spaces provide value-added services or not **($C_{5,2}$)**.

4.2 Meta-dimension: technical

In the second meta-dimension, we examine the characteristics of the data shared in a data space before getting to its technical architecture. According to Franklin et al. (2005), the **data structure (D_6)** in data spaces is heterogeneous and can be structured **($C_{6,1}$)**, semi-structured **($C_{6,2}$)**, or unstructured **($C_{6,3}$)** (Singh, 2013). Structured data are usually managed in relational DBMS. In contrast, unstructured data, such as text messages or videos, do not follow a specific format. Semi-structured data does not follow a conventional database system but, for instance, provides self-describing elements (Hashem et al., 2015). The **data types (D_7)** differ as bulk data (raw **($C_{7,1}$)** and processed data **($C_{7,2}$)**) and metadata **($C_{7,3}$)** to enable good scalability. Metadata provides information about other data so that users can discover relevant data sources (Franklin et al., 2005). **Data processing (D_8)** happens as a stream **($C_{8,1}$)** or as a batch upload **($C_{8,2}$)**. For example, things in a smart environment, like connected devices or sensors, can produce real-time data streams (Curry et al., 2019).

The **architecture (D_9)** represents the central technical infrastructure in a data space. It has impacts on several factors, including data security, data management, and trust between the data space participants (Al-Zahrani, 2020). Data space architectures can either be centralized **($C_{9,1}$)** (Li et al., 2022) or decentralized **($C_{9,2}$)** (Arellanes and Lau, 2019; George et al., 2016). Hybrid architectures **($C_{9,3}$)** combine centralized and distributed technologies (Abraham et al., 2019; Große et al., 2020; Shen et al., 2020).

Given the plethora of architectural designs, the underlying **data-sharing logic (D_{10})** differs too. First, peer-to-peer (P2P) data sharing **($C_{10,1}$)** can be performed (Halevy et al., 2005). In this case, the data provider and data receiver directly exchange their data. Another option is sharing data via a platform **($C_{10,2}$)**. This characteristic also includes data marketplaces, data portals, or data sharing via a hub (Curry et al., 2019; Labadie et al., 2020). A data intermediary **($C_{10,3}$)** acts as a go-between for organizations that want to make their data available and those that want to use it (Janssen and Singh 2022). This could be, for example, a not-for-profit organization. Finally, a combination of several data sharing logics is possible. For instance, DatenMarktplatz.NRW (2020) and Catena-X (2022) store metadata centralized while storing the bulk data decentralized to achieve privacy by performing P2P data sharing.

To **harmonize (D_{11})** the uploaded data, data models **($C_{11,1}$)** are employed. Data models can be semantic models (EuPro Gigant, 2022), hash functions (DE4L, 2022), Worker-Scripts (DatenMarktplatz.NRW, 2020), Human-in-the-loop models (Demeter, 2020), and domain-specific data models (Agri-gaia, 2022). Also, catalogs **($C_{11,2}$)** can be employed to harmonize uploaded (meta) data (Advaneo, 2022). While these catalogs are a dictionary of knowledge about the data and processes used to manage and consume the data, data models are more abstract models that organize the data and standardize their relationships.

There are a variety of **access technologies (D_{12})**. First, data spaces use dedicated connectors **($C_{12,1}$)** (EnDaSpace, 2021; Mobility Data Space, 2021; SCSN, 2022). A connector can be either an internal one from a participating organization or an external one that executes data exchange between organizations (Braud et al., 2021). The underpinning technology can be standardized (e.g., IDS connector). Another way to access a data space is via a portal **($C_{12,2}$)** (Demeter, 2020), accessible through webpages or apps (DaWID, 2022). Also, a combination of both access technologies is possible.

Trusted data exchange (D_{13}) is ensured by identity providers that have identity management **($C_{13,1}$)** in several forms, including a membership form (AI4EU, 2022), authentication (agdatahub, 2022), and authorization (Demeter, 2020). Another possibility is providing trust via a certification authority **($C_{13,2}$)**, managing and issuing digital certificates to data space participants (Nesheim et al., 2021; EnDaSpace, 2021). Because this dimension focuses on the technical implementation for ensuring trust and not trust in general, we still position it as part of the technical meta-dimension and not as governance.

4.3 Meta-dimension: governance

In the third meta-dimension, we highlight that **privacy (D_{14})** can be ensured if data can be shared anonymously **($C_{14,1}$)**, (Curry et al., 2019); so, communication parties hide their identity (Sun et al., 2020). In the context of protecting personal data, pseudonymization **($C_{14,2}$)** plays a crucial role (DaWID, 2022).

In some cases, however, it might be important to share data non-anonymously **($C_{14,3}$)** (SCSN, 2022) or configure privacy **($C_{14,4}$)** depending on the underlying use case of the data space (HEALTH-X, 2022).

Several categories can be used to classify the data. Given the multiplicity of **data classification schemes (D_{15})**, the taxonomy only presents the most common ones: domain **($C_{15,1}$)**, origin **($C_{15,2}$)** (e.g., provider) (Demeter, 2020), topicality **($C_{15,3}$)**, size **($C_{15,4}$)** (DE4L, 2022), and data format **($C_{15,5}$)** (Advaneo, 2022). As there are many other possibilities, we also provide space for further classification schemes **($C_{15,6}$)**.

Data sharing policies (D_{16}) can either be mandatory if set by the data space **($C_{16,1}$)** (AI4EU, 2022), or voluntary if the data provider **($C_{16,2}$)** can choose them (Mobility Data Space, 2021). Data sharing policies set by a data provider also include policies set by a data owner in case of a different owner than the provider (DaWID, 2022), as inheritance is an important topic when it comes to policies (Huang et al., 1991). Also, it is plausible to have combinations of policies (DaWID, 2022; SCSN, 2022).

A data provider's willingness to share its data is determined by the extent to which it may retain ownership and sovereignty over its data and hence control its subsequent usage (Fecher et al., 2015). That’s why **data tracing (D_{17})** is essential for data sharing and goes along with data control. Tracing can happen in different ways: the trace in the space dimension **($C_{17,1}$)** by analyzing the relational data evolution and the provenance. This also includes access control which restricts access to the data repository by discriminating between those having access and those who do not, also called role-based access rights (Agri-gaia, 2022; Catena-X, 2022; DE4L, 2022); the change trace in the time dimension **($C_{17,2}$)** by analyzing the operational data evolution (Cheng et al., 2013; DatenMarktplatz.NRW, 2020; EuPro Gigant, 2022). Extending this, traceability in use dimension **($C_{17,3}$)** is possible, which also comes along with Usage Control to determine the abilities of actors (e.g., for what purpose they can use the data) (Mobility Data Space, 2021; EnDaSpace, 2021; Demeter, 2020). To ensure full data sovereignty, a mixture of different kinds of traceability and control can be implemented (EuPro Gigant, 2022). Lastly, there can be no data tracing **($C_{17,4}$)**. This trait is connected to the open data paradigm, in which data is made publicly available and may be utilized by anybody (Fadler and Legner, 2020; Link et al., 2017). However, it can still be the case that the open data is traced, or terms of use are defined.

5 Demonstration: Illustrative Application

In the following section, we indicate the applicability of the taxonomy with three illustrative examples as proposed by Kundisch et al. (2022) (see Figure 3). For each case, we had access to the public data we used to construct the taxonomy as well as to informants involved in the projects (three informants for three example cases). The informants were asked to fill out the taxonomy based on their knowledge.

The first case is the *Health-X Data Space*[4]. The data space is citizen-facing (patients) and uses Gaia-X standards to build a secure platform for handling personal data. Subsequently, the focus is on the European health domain and on the generation of individual business models based on personal data for service providers or equipment manufacturers.

The second case is the *Mobility Data Space*[5], which is a central hub for the secure exchange of mobility data. In this, it fosters the exploration of mobility data in catalogs, monetization of mobility data, and the design of new services and business models based on mobility data. Given its character as a data space, a key feature is the self-determination of data providers regarding the use of data (i.e., who is allowed to use them under which obligations).

The third case, *Energy Data Space*[6], focuses on data in the energy sector, as the generation from wind energy and photovoltaic systems is weather-dependent and thus creates a significant need for information and communication. In this project, the concept of International Data Spaces was used to demonstrate the communication flow between a wind turbine and an electrolyzer for hydrogen

[4] https://www.health-x.org/

[5] https://mobility-dataspace.eu/

[6] https://www.iee.fraunhofer.de/de/projekte/suche/2021/EnDaSpace.html

production. As a digital service, a schedule was calculated from wind turbine operating data and electricity market information to ensure the economical production of green hydrogen.

The illustrative application of the taxonomy (see Figure 3) shows that we can distinguish different design options of data spaces within the taxonomy, meaning it fulfills its purpose (Peffers et al., 2012). Furthermore, since data spaces have different characteristics and are designed in different ways, the taxonomy shows that we can characterize essential design options of data spaces. Since we strive to generate design options for data spaces in our meta-characteristic (see Section 3.1), and the figure below shows the applicability of the taxonomy, we have successfully implemented our meta-characteristic, "design options for data spaces".

MD	Dimension	Characteristics					
Economic	Domain	**Domain-Specific**	Cross-Domain				
	Funding	Public	Private	**Private-public partnership**			
	Data space access	**Free**	Fee				
	Reward	**Money**	**Data**	**Service**	Reputation	None	
	Value added services	Yes	**No**				
Technical	Data structure	**Structured**	Semi-structured	Unstructured			
	Data type	Raw data	**Processed data**	**Metadata**			
	Data processing	Stream	**Batch**				
	Architecture	Centralized	Decentralized	**Hybrid**			
	Data sharing logic	**P2P data sharing**	Data platform	Data sharing via intermediaries			
	Data harmonization	**Data models**	Data catalog				
	Access technology	**Standardized connector**	Portal				
	Trusted exchange	**Trust by identity management**	**Trust by certification**				
Governance	Data privacy	**Anonymous**	Pseudonymous	Non-anonymous	Various		
	Data classification scheme	Domain	**Origin**	Topicality	Size	Data format	…
	Data sharing policies	Set by data space	**Set by data provider**				
	Traceability & control	**Space dimension**	**Use dimension**	Time dimension	None		

Mobility Data Space

MD	Dimension	Characteristics					
Economic	Domain	**Domain-specific**	Cross-domain				
	Funding	**Public**	Private	Private-public partnership			
	Data space access	**Free**	Fee				
	Reward	Money	Data	**Service**	Reputation	None	
	Value added services	Yes	**No**				
Technical	Data structure	**Structured**	**Semi-structured**	**Unstructured**			
	Data type	**Raw data**	Processed data	**Metadata**			
	Data processing	**Stream**	Batch				
	Architecture	Centralized	**Decentralized**	Hybrid			
	Data sharing logic	**P2P data sharing**	Data platform	Data sharing via intermediaries			
	Data harmonization	**Data models**	**Data catalog**				
	Access technology	**Standardized connector**	**Portal**				
	Trusted exchange	**Trust by identity management**	Trust by certification				
Governance	Data privacy	Anonymous	Pseudonymous	Non-anonymous	**Various**		
	Data classification scheme	**Domain**	**Origin**	**Topicality**	Size	Data format	…
	Data sharing policies	Set by data space	**Set by data provider**				
	Traceability & control	**Space dimension**	**Use dimension**	**Time dimension**	None		

HEALTH-X

MD	Dimension	Characteristics					
Economic	Domain	**Domain-specific**	Cross-domain				
	Funding	**Public**	Private	Private-public partnership			
	Data space access	**Free**	Fee				
	Reward	Money	Data	**Service**	Reputation	None	
	Value added services	**Yes**	No				
Technical	Data structure	**Structured**	**Semi-structured**	Unstructured			
	Data type	**Raw data**	**Processed data**	**Metadata**			
	Data processing	**Stream**	**Batch**				
	Architecture	Centralized	Decentralized	**Hybrid**			
	Data sharing logic	**P2P data sharing**	Data platform	**Data sharing via intermediaries**			
	Data harmonization	**Data models**	**Data catalog**				
	Access technology	**Standardized connector**	Portal				
	Trusted exchange	**Trust by identity management**	**Trust by certification**				
Governance	Data privacy	Anonymous	Pseudonymous	**Non-anonymous**	Various		
	Data classification scheme	**Domain**	**Origin**	Topicality	Size	Data format	…
	Data sharing policies	**Set by data space**	**Set by data provider**				
	Traceability & control	**Space dimension**	**Use dimension**	Time dimension	None		

Energy Data Space

Figure 3. An illustrative application of the taxonomy of data spaces, including the Mobility Data Space (top left), HEALTH-X (top right), and Energy Data Space (bottom).

6 Discussion and Conclusion

This paper developed a taxonomy to capture and organize design options for an emerging artifact, namely data spaces. To the best of our knowledge, no other taxonomies about data spaces exist that synthesize current knowledge from the literature and supplement it with empirical items.

We can draw various conclusions about theory and practice from our findings. In terms of **scientific contributions**, our work advances knowledge of the rapidly developing and mostly unexplored research area of data spaces. In particular, our taxonomy aims to expand the existing body of knowledge on design options for data spaces by making use of conceptual and empirical insights as well as to contribute to the specification of a shared understanding of this complex topic. The taxonomy is a mechanism to store the knowledge we have collected from the literature corpus and the data spaces we have analyzed. For researchers, we expect the taxonomy to contribute a 'big picture' to an emerging research field that is of high importance for society and industry. Also, given the significant European landscape of research investigating data spaces, the taxonomy proposes a domain-agnostic overview of design options that others can contextualize and tailor in their projects to their application scenarios in their domains. The blurriness of definitions also sharpens what we understand as a data space and builds a 'playing field' for others to go into more detail in specific elements.

The results of this research also offer versatile **contributions for practitioners**. For example, industry-driven projects such as Catena-X are dedicated to developing data spaces for industries at large scale. Subsequently, practitioners need to collect knowledge about how to design data spaces tailored to their individual needs (e.g., design decisions on whether anonymous or non-anonymous data is necessary). From a design perspective, the taxonomy assists these practitioners in tackling this real-world 'wicked' design problem (Hevner et al., 2004). Besides this, we have outlined the value of inter-organizational data sharing. The establishment of data spaces assists and helps companies to reap the benefits of cross-organizational data sharing (Otto et al., 2020).

Our study is, naturally, subject to several **limitations** that must be considered when interpreting the results. Because of the ongoing fast technical and organizational advancement in digitalization, as well as the fact that it is still a relatively unexplored study subject (Azkan et al., 2019), the concepts around data spaces are continually developing. As a result, our taxonomy is a time-bound picture that must be updated periodically to remain relevant and consider new dimensions and characteristics created by the advancement of digitalization. Although the taxonomy is based on a review of the scientific literature as well as actual use cases of data spaces, the data collection itself is subject to interpretation. Thus other researchers may derive other dimensions and features based on their influences, preferences, and biases. The depth and scope of the taxonomy are also limited due to the selection of literature and the search activities conducted (e.g., only backward search) as well as use cases. To account for the subjective procedure, we aimed to select a representative set across sectors and performed interim evaluations with data space experts. However, the three data spaces chosen for evaluation were also part of the selected use cases, so it would have been surprising if they could not be classified. Lastly, due to the varying degrees of maturity in practice, the empirical samples examined are unlikely to cover all domains in which data spaces can develop, which means that the transferability of the results cannot be fully guaranteed and, instead, leave room for further study practice-oriented research.

In general, the constraints point to potential **future research** directions. The development of archetypical patterns for data spaces is a common next step in information systems taxonomy research (Azkan et al., 2020). Based on the archetypes found, it may be explored if some archetypes are more effective than others, and design guidelines for data spaces might be derived (Möller et al., 2021a). For that, we have already started an interview series to formalize further knowledge of data space organizations. As a result, we may expect new dimensions or characteristics to emerge. For example, this study does not uncover any characteristics of the clearing house that documents the data-sharing activities. In addition, the compatibility of the different data spaces, so-called interoperability, has not been investigated. However, these challenges are becoming increasingly important.

Acknowledgment

This work was supported by the German Federal Ministry of Education and Research (BMBF) and is part of the research project "IEDS – Incentives and Economics of Data Sharing" (IEDS0001).

References

Abraham, R., J. Schneider and J. vom Brocke (2019). "Data governance: A conceptual framework, structured review, and research agenda" *International Journal of Information Management* 49, 424–438.

Advaneo (2022). *Access to the world of data*. URL: https://www.advaneo.de/# (visited on November 09, 2022).

agdatahub (2022). *European operator of consents and data exchange platform*. URL: https://agdatahub.eu/en/ (visited on November 05, 2022).

Agri-gaia (2022). *Ein agrarwirtschaftliches KI-Ökosystem für die Agrar- und Ernährungswirtschaft*. URL: https://www.agri-gaia.de/ (visited on October 18, 2022).

AI4EU (2022). *Europe's AI-on-Demand Platform*. URL: https://www.ai4europe.eu/ (visited on October 17, 2022).

Alonso, Á., A. Pozo, J. M. Cantera, F. de La Vega and J. J. Hierro (2018). "Industrial Data Space Architecture Implementation Using FIWARE" *Sensors* 18 (7), 2226.

Al-Zahrani, F. A. (2020). "Subscription-Based Data-Sharing Model Using Blockchain and Data as a Service" *IEEE Access* 8, 115966–115981.

Anupam, V., S. Dar, T. Leibfried and E. Petajan (1995). "Research report. DataSpace: 3-D visualizations of large databases". In: *Proceedings on Information Visualization*: IEEE Comput. Soc. Press. Atlanta, USA.

Arellanes, D. and K. Lau (2019). "Decentralized Data Flows in Algebraic Service Compositions for the Scalability of IoT Systems". In: *IEEE 5th World Forum on Internet of Things (WF-IoT).* Limerick: Ireland.

Azkan, C., I. Gür, J. Gelhaar, A. Gieß, T. Groß, M. Hupperz, H. Kett, D. Kutzias, O. Strauß, J. Büchel, V. Demary, B. Engels, H. Goecke, A. Mertens, K.-H. Röhl, C. Rusche, M. Scheufen, B. Schröder, F. Möller, J. Schweihoff, I. Jussen, V. Dahms, R. Andres, D. Erdsiek, T. Niebel, D. Rehse, V. Rost, J. Walter and S. Valet (2022). *Incentives and Economics of Data Sharing*. *Fields of action of cross-company data exchange and status quo of the German economy:* Fraunhofer-Gesellschaft e.V.

Azkan, C., L. Iggena, I. Gür, F. O. Möller and B. Otto (2020). "A Taxonomy for Data-Driven Services in Manufacturing Industries". In: *Proceedings of the 24th Pacific Asia Conference on Information Systems.* Dubai, UAE.

Azkan, C., M. Spiekermann and H. Goecke (2019). "Uncovering Research Streams in the Data Economy Using Text Mining Algorithms" *Technology Innovation Management Review* 9 (11), 62–74.

Badewitz, W., S. Kloker and C. Weinhardt (2020). "The Data Provision Game: Researching Revenue Sharing in Collaborative Data Networks". In: *IEEE 22nd Conference on Business Informatics.* Antwerp, Belgium.

Baecker, J., T. P. Böttcher and J. Weking (2021). "How Companies Create Value From Data - A Taxonomy on Data, Approaches, and Resulting Business Value". In: *Proceedings of the 29th European Conference on Information Systems.* Marrakech, Morocco.

Bailey, K. D. (1994). *Typologies and Taxonomies*. *An Introduction to Classification Techniques.* Thousand Oaks, Calif.: SAGE Publications.

Beverungen, D., T. Hess, A. Köster and C. Lehrer (2022). "From private digital platforms to public data spaces: implications for the digital transformation" *Electronic Markets* 32 (2), 493–501.

Braud, A., G. Fromentoux, B. Radier and O. Le Grand (2021). "The Road to European Digital Sovereignty with GAIA-X and IDSA" *IEEE Network* 35 (2), 4–5.

Busi, N. and G. Zavattaro (2001). "Publish/subscribe vs. shared dataspace coordination infrastructures. Is it just a matter of taste?" *Proceedings Tenth IEEE International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises. WET ICE 2001*, 328–333.

Busi, N. and G. Zavattaro (2003). "Expired data collection in shared dataspaces" *Theoretical Computer Science* 298 (3), 529–556.

Catena-X (2022). *Catena-X Automotive Network - Die Auffahrt zur Digital Economy*. URL: https://catena-x.net/de/ (visited on October 18, 2022).

CDQ (2022). *Corporate Data Quality - Sharing Data Excellence*. URL: https://www.cdq.com/ (visited on October 05, 2022).

Cheng, X., C. Hu, Y. Li and H. Zuo (2013). "Data Evolution Analysis of Virtual DataSpace for Managing the Big Data Lifecycle" *IEEE International Symposium on Parallel & Distributed Processing, Workshops and Phd Forum*.

Cooper, H. M. (1988). "Organizing Knowledge Syntheses: A Taxonomy of Literature Reviews" *Knowledge in Society* 1 (104).

Cunningham, C. and R. Gruia-Catalin (1989). *A UNITY-Style Programming Logic for a Shared Dataspace Language* Report Number: WUCS-89-05. All Computer Science and Engineering.

Curry, E., W. Derguech, S. Hasan, C. Kouroupetroglou and U. Hassan (2019). "A Real-time Linked Dataspace for the Internet of Things: Enabling ''Pay-As-You-Go'' Data Management in Smart Environments" *Future Generation Computer Science* 90, 405-422.

Curry, E. and A. Ojo (2020). "Enabling Knowledge Flows in an Intelligent Systems Data Ecosystem". In E. Curry (ed.) *Real-time Linked Dataspaces*, pp. 15–43. Cham: Springer Nature.

DatenMarktplatz.NRW (2020). *Werde Teil des ersten GAIA-X konformen IoT-Datenraums in NRW, aus NRW, für NRW*. URL: https://datenmarktplatz.nrw/ (visited on October 12, 2022).

DaWID (2022). *Datenzentrierte Wertschöpfungsplattform für Interaktive, assistierende Dienstleistungssysteme*. URL: https://www.dawid-projekt.de/projektbeschreibung/ (visited on October 16, 2022).

DE4L (2022). *Data Economy 4 Advanced Logistics*. URL: https://de4l.io/ (visited on October 18, 2022).

Demeter (2020). *Demeter - Empowering Farmers*. URL: https://h2020-demeter.eu/ (visited on October 18, 2022).

Du, Z., X. Fu, C. Zhao, T. Liu and Q. Liu (2012). "Multi-domain cloud social network service platform supporting online collaborations on campus". In: *IEEE 2nd International Conference on Cloud Computing and Intelligence Systems.* Hangzhou, China.

EnDaSpace (2021). *EnDaSpace - moderne Datenökonomie in der Energiewirtschaft*. URL: https://www.iee.fraunhofer.de/de/projekte/suche/2021/EnDaSpace.html (visited on October 16, 2022).

EuPro Gigant (2022). *Wissens Hub - Internationale Datenräume*. URL: https://euprogigant.com/wissens-hub/internationale-datenraeume/ (visited on October 12, 2022).

European Commission (2020). *Communication from the commission to the European parliament, the council, the European economic and social committee and the committee of the regions: A European strategy for data.* Brussels.

European Commission (2022). *The European Commission's priorities*. URL: https://ec.europa.eu/info/strategy/priorities-2019-2024_en (visited on November 03, 2022).

Fadler, M. and C. Legner (2020). "Who owns data in the enterprise? Rethinking data ownership in times of big data and analytics". In: *Proceedings of the 28th European Conference on Information Systems.* Marrakech, Morocco.

Fassnacht, M., C. Benz, D. Heinz, J. Leimstoll and G. Satzger (2023). "Analyzing Barriers to Data Sharing among Private Sector Organizations: Combined Insights from Research and Practice". In: *Proceedings of the 56th Hawaii International Conference on System Sciences.* Hawaii, USA.

Fecher, B., S. Friesike and M. Hebing (2015). "What Drives Academic Data Sharing?" *PloS one* 10 (2), e0118053.

Fernandez, R. C., P. Subramaniam and M. J. Franklin (2020). "Data Market Platforms: Trading Data Assets to Solve Data Problems" *Proceedings of the VLDB Endowment* 13 (12).

Franklin, M., A. Halevy and D. Maier (2005). "From Databases to Dataspaces: A New Abstraction for Information Management" *ACM SIGMOD Record* 34 (4), 27–33.

Gelhaar, J., T. Groß and B. Otto (2020). "A Taxonomy for Data Ecosystems". In: *Proceedings of the 53rd Hawaii International Conference on System Sciences.* Hawaii, USA.

George, H., L. Fletcher and F. Mandreoli (2016). "No users no dataspaces! Query-driven dataspace orchestration" *Proceedings of the 24th Italian Symposium on Advanced Database Systems SEBD.*

Glass, R. L. and I. Vessey (1995). "Contemporary Application-Domain Taxonomies" *IEEE Software* 12 (4), 63–76.

Große, N., D. Leisen, T. Gürpinar, R. S. Forsthövel, M. Henke and M. ten Hompel (2020). "Evaluation of (De-) Centralized IT technologies in the fields of Cyber-Physical Production Systems". In: *Proceedings of the 1st Conference on Production Systems and Logistics*. Hannover, Germany.

Halevy, A., M. Franklin and D. Maier (2006). *Principles of dataspace systems.* New York, NY, USA: Association for Computing Machinery.

Halevy, A., Z. G. Ives, D. Suciu and I. Tatarinov (2005). "Schema mediation for large-scale semantic data sharing" *The VLDB Journal* 14, 68–83.

Hashem, I. A. T., I. Yaqoob, N. B. Anuar, S. Mokhtar, A. Gani and U. Khan (2015). “The rise of “big data” on cloud computing: Review and open research issues” *Information Systems* 47, 98–115.

HEALTH-X (2022). *dataLOFT: Die Zukunft der Gesundheitsversorgung*. URL: https://www.health-x.org/ (visited on October 20, 2022).

Hevner, A., S. March, J. Park and S. Ram (2004). “Design Science in Information Systems Research” *MIS Quarterly* 28 (1), 75–105.

Huang, J., J. Stankovic, K. Ramamritham and D. Towsley (1991). “On Using Priority Inheritance In Real-Time Databases” *Proceedings Twelfth Real-Time Systems Symposium*.

Icks, A., C. Schröder, S. Brink, C. Dienes and S. Schneck (2017). “Digitalisierungsprozesse von KMU im Produzierenden Gewerbe. Research Report” *IfM-Materialien* (255).

IDSA (2022). *Data Space Radar: Faster IDS breakthroughs are within range*. URL: https://internationaldataspaces.org/adopt/data-space-radar/ (visited on October 12, 2022).

Jussen, I., J. Schweihoff, V. Dahms, F. O. Möller and B. Otto (2023). “Data Sharing Fundamentals: Characteristics and Definition”. In: *Proceedings of the 56th Hawaii International Conference on System Sciences.* Hawaii, USA.

Kundisch, D., J. Muntermann, A. M. Oberländer, D. Rau, M. Röglinger, T. Schoormann and D. Szopinski (2022). “An Update for Taxonomy Designers. Methodological Guidance from Information Systems Research” *Business & Information Systems Engineering* 64, 421–439.

Labadie, C., C. Legner, M. Eurich and M. Fadler (2020). “FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs”. In: *IEEE 22nd Conference on Business Informatics.* Antwerp, Belgium.

Legner, C., T. Pentek and B. Otto (2020). “Accumulating Design Knowledge with Reference Models: Insights from 12 Years’ Research into Data Management” *Journal of the Association for Information Systems* 21 (3), 735–770.

Li, P., K. Kai Cheng, P. Jiang and K. Katchasuwanmanee (2022). “Investigation on industrial dataspace for advanced machining workshops: enabling machining operations control with domain knowledge and application case studies” *Journal of Intelligent Manufacturing* 33, 103–119.

Link, G., K. Lumbard, K. Conboy, M. Feldmann, J. Feller, J. George, M. Germonprez, S. Goggins, D. Jeske, G. Kiely, K. Schuster and M. Willis (2017). “Contemporary Issues of Open Data in Information Systems Research: Considerations and Recommendations” *Communications of the Association for Information Systems* 41, 587–610.

Lis, D. and B. Otto (2021). “Towards a Taxonomy of Ecosystem Data Governance”. In: *Proceedings of the 54th Hawaii International Conference on System Sciences.* Hawaii, USA.

Mobility Data Space (2021). *Mobility Data Space - Data Sharing Community*. URL: https://mobility-dataspace.eu/de (visited on October 05, 2022).

Möller, F. O., H. Haße, C. Azkan, H. van der Valk and B. Otto (2021a). “Design of Goal-Oriented Artifacts from Morphological Taxonomies: Progression from Descriptive to Prescriptive Design Knowledge”. In S. Stieglitz, R. Schütte and F. Ahlemann (eds.) *Innovation Through Information Systems: Volume I: A Collection of Latest Research on Domain Issues*, pp. 523–538. Cham: Springer.

Möller, F. O., M. Stachon, C. Azkan, T. Schoormann and B. Otto (2021b). “Designing business model taxonomies – synthesis and guidance from information systems research” *Electronic Markets* 32, 701–726.

Muschalle, A., F. Stahl, A. Löser and G. Vossen (2013). “Pricing Approaches for Data Markets”. In M. Castellanos, U. Dayal and E. A. Rundensteiner (eds.) *Enabling real-time business intelligence*, pp. 129–144. Berlin, Heidelberg: Springer.

Nagasaka, Y. and H. Motoyama (2007). “A Shared Dataspace Communication Framework for Data Acquisition System”. In: *15th IEEE-NPSS Real-Time Conference.* Batavia, IL, USA.

Nesheim, D. A., K. Bernsmed, b. zernichow, Ø. Rødseth and P. H. Meland (2021). “Secure, Trustworthy and Efficient Information Exchange – Enabling Added Value through The Maritime Data Space and Public Key Infrastructure”. In: *20th International Conference on Computer and IT Applications in the Maritime Industries.* Mülheim, Germany.

Nickerson, R. C., U. Varshney and J. Muntermann (2013). “A Method for Taxonomy Development and its Application in Information Systems” *European Journal of Information Systems* 22 (3), 336–359.

Oliveira, M. I. S., Barros Lima, Glória de Fátima and B. Farias Lóscio (2019). "Investigations into Data Ecosystems: a systematic mapping study" *Knowledge and Information Systems (61:2)* 61, 589–630.

Otto, B. (2022). "The Evolution of Data Spaces". In B. Otto, M. ten Hompel and S. Wrobel (eds.) *Designing Data Spaces*. *The Ecosystem Approach to Competitive Advantage*, pp. 3–15. Cham: Springer International Publishing.

Otto, B., N. Mohr, M. Roggendorf and T. Guggenberger (2020). "Data sharing in industrial ecosystems: Driving value across entire production lines" *McKinsey & Company*.

Peffers, K., M. Rothenberger, T. Tuunanen and R. Vaezi (2012). "Design Science Research Evaluation". In: *Design Science Research in Information Systems. Advances in Theory and Practice*. Ed. by K. Peffers, M. Rothenberger, B. Kuechler. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 398–410.

Petajan, E. D., Y. D. Jean, D. Lieuwen and V. Anupam (1997). "Dataspace: An automated visualization system for large databases". In: *Visual Data Exploration and Analysis IV*. Ed. by G. G. Grinstein, R. F. Erbacher: SPIE, pp. 89–98. San Jose, CA.

Poser, M., C. Wiethof and E. A. C. Bittner (2022). "Integration of AI into Customer Service: A Taxonomy to Inform Design Decisions". In: *Proceedings of the 30th European Conference on Information Systems*. Timisoara, Romania.

Regione Lombardia (2022). *L'ECOSISTEMA DIGITALE: E015 digital ecosystem*. URL: https://www.e015.regione.lombardia.it/ (visited on November 10, 2022).

Rosian, M., M. Altendeitering and B. Otto (2022). "A Socio-Technical Analysis of Challenges in Managing Multi-Clouds" *Wirtschaftsinformatik 2022 Proceedings*.

Schahovska, N. (2011). "Datawarehouse and Dataspace – information base of Decision Support System". In: *11th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics*. Polyana-Svalyava, Ukraine.

SCSN (2022). *Smart connected supplier network - Get connected*. URL: https://smart-connected.nl/de (visited on October 05, 2022).

Shen, M., J. Duan, L. Zhu, J. Zhang, X. Du and M. Guizani (2020). "Blockchain-Based Incentives for Secure and Collaborative Data Sharing in Multiple Clouds" *IEEE Journal on Selected Areas in Communications* 38 (6), 1229–1241.

Singh, M. (2013). "A Framework for Data Modeling and Querying Dataspace Systems". In: *Proceedings of the 7th International Conference on Data Mining and Warehousing*. Bangalore, India.

Singh, M. and S. K. Jain (2011). "A Survey on Dataspace". In D. C. Wyld and M. Wozniak (eds.) *Advances in network security and applications*, pp. 608–621. Berlin: Springer.

Smotroff, I., L. Hirschman and S. Bayer (1994). "Integrating natural language with large dataspace visualization". In: *Digital Libraries Current Issues*. Ed. by N. R. Adam, B. K. Bhargava, Yesha, Y. (eds). Berlin, Heidelberg: Springer.

Srivastava, P. and N. Hopwood (2009). "A Practical Iterative Framework for Qualitative Data Analysis" *International Journal of Qualitative Methods* 8 (1), 76–84.

Sun, Y., L. Yin, Z. Sun, Z. Tian and X. Du (2020). "An IoT data sharing privacy preserving scheme". In: *IEEE INFOCOM 2020 - IEEE Conference on Computer Communications Workshops*. Toronto, ON, Canada.

Szopinski, D., T. Schoormann and D. Kundisch (2020a). "Criteria as a prelude for guiding taxonomy evaluation". In: *Proceedings of the 53rd Hawaii International Conference on System Sciences*. Hawaii, USA.

Szopinski, D., T. Schoormann and D. Kundisch (2020b). "Visualize Different: Towards Researching the Fit Between Taxonomy Visualizations and Taxonomy Tasks" *Proceedings of the Wirtschaftsinformatik (WI)*.

Thomas, L. D. W. and A. Leiponen (2016). "Big Data Commercialization" *IEEE Engineering Management Review* 44 (2), 74–90.

vom Brocke, J., A. Simons, B. Niehaves, K. Reimer, R. Plattfaut and A. Cleven (2009). "Reconstructing the Giant: On the Importance of Rigour in Documenting the Literature Search Process". In: *Proceedings of the 17th European Conference on Information Systems*. Verona, Italy.

Wallis, J. C., E. Rolando and C. L. Borgman (2013). "If We Share Data, Will Anyone Use Them? Data Sharing and Reuse in the Long Tail of Science and Technology" *PloS one* 8 (7), e67332.

Wang, Y., S. Song and L. Chen (2016). "A survey on accessing dataspaces" *ACM SIGMOD Record* 45 (2), 33–44.

Webster, J. and R. T. Watson (2002). "Analyzing the Past to Prepare for the Future: Writing a Literature Review" *MIS Quarterly* 26 (2), xiii–xxiii.

Williams, K., S. Chatterjee and M. Rossi (2008). "Design of emerging digital services: a taxonomy" *European Journal of Information Systems* 17 (5), 505–517.

Woerner, S. L. and B. H. Wixom (2015). "Big Data: Extending the Business Strategy Toolbox" *Journal of Information Technology* 30 (1), 60–62.

Xie, K., Y. Hu and J. Xiao (2020). "Antecedents of User's Intention to Share Big Data The influence of User Type". In: *Proceedings of the 24th Pacific Asia Conference on Information Systems.* Dubai, UAE.

Zrenner, J., F. O. Möller, C. Jung, A. Eitel and B. Otto (2019). "Usage control architecture options for data sovereignty in business ecosystems" *Journal of Enterprise Information Management* 32 (3), 477–495.