What Does it Take to Connect?
Unveiling Characteristics of Data Space Connectors

Anna Gieß
Fraunhofer ISST / TU Braunschweig
anna.giess@isst.fraunhofer.de

Marius Hupperz
Fraunhofer ISST
marius.hupperz@isst.fraunhofer.de

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

Frederik Möller
TU Braunschweig / Fraunhofer ISST
frederik.moeller@tu-braunschweig.de

Abstract

Data spaces are a novel data management approach to collect large-scale heterogeneous data distributed over various data sources in different formats. To access these data spaces, users require so-called connectors to ensure technical compliance (e.g., usage control policies) and ensure that users play by the 'same rules'. While connectors are a critical component of data spaces and receive considerable attention in politics, practice, and research, there is still no shared understanding of what constitutes a connector. To address this gap, we analyzed 23 connector use cases, diverse types of practitioner literature (n = 14), 25 scientific papers, and a workshop with five experts to extract the characteristics of connectors. We synthesized our findings into a taxonomy that integrates insights from the conceptual and empirical analysis and illustrated it by classifying two connectors. Our paper contributes to our understanding of this novel artifact, which has implications for future businesses.

Keywords: Data Sharing, Data Space, Connector, Taxonomy, Conceptualization.

1. Introduction

"*Realize that everything connects to everything else*" – Leonardo DaVinci

In today's digital world, an immense amount of data is stored in different formats and places. To leverage these data effectively and generate value, they often must be exchanged between heterogeneous systems and applications beyond organizational boundaries (Jiang et al., 2021). As a result, cross-company data exchange is becoming increasingly important. This is evident, for example, by the *supply chain law* in Germany (ger. *LkSG*) that holds companies responsible for maintaining human rights and environmental rules across the entire supply chain. For instance, the data space initiative Catena-X Automotive Ecosystems wants to foster transparency through sovereign data sharing in automotive supply chains to enable companies to comply with the supply chain law (Martin-Jung, 2022). In addition to being mandated by law in some cases, cross-company data exchange can create new products and services. For instance, the agricultural data space Agri-gaia (2022) makes claims regarding plant growth in sub-areas or even individual plants using drone data or satellite pictures. With the help of this targeted analysis, farmers can manage their arable land more sustainably to maximize crop growth while minimizing resource use. Another example is the Mobility Data Space (2021), where vehicle manufacturers can use traffic data to identify dangerous situations and pass on information to other participants, or public transport operators can improve passenger information services.

Those data spaces lately receive growing attention from scholars and practitioners, which can be seen by an increase in the scientific literature (e.g., Beverungen et al., 2022; Donald et al., 2023; Gieß et al., 2023), an increase in establishing new data spaces (e.g., Data Space Radar, IDSA, 2023), as well as interest from leading companies such as Amazon Web Services (AWS) (Kolodziej & Vazquez, 2023). As a fundamental component allowing various actors with different data formats and sources to participate in such data spaces, gateways or so-called *data space connectors* are used (Gieß et al., 2023). These software components enable seamless and secure data exchange. Developing appropriate connectors for existing and emerging data spaces (i.e., engineer perspective) as well as making informed decisions about what connectors are suitable for a particular data space (i.e., user perspective) are crucial tasks.

However, due to rapid developments in this field, we can observe heterogeneity in types and attributes of connectors, which hinders purposeful building and choosing the most suitable ones. Against this, we aim to answer the following research question:

What are the options for designing a connector?

In pursuing to answer this, we followed a systematic approach to building and evaluating a taxonomy (Kundisch et al., 2022). A taxonomy aids in organizing a corpus of information that makes up a field and provides an adequate method for formulating conceptual knowledge (Glass & Vessey, 1995). It enables the classification of objects and their possible relations. Referring to this paper's aim, especially the combination of scientific and practical information, is helpful because the stream of data space connectors is still evolving and benefits from both conceptual and empirical reasoning.

The paper is structured as follows. In the next section, we briefly state the research background of data spaces and the associated connectors. Then we describe the research design, including the process of building and evaluating the taxonomy. Following that, we present our resulting taxonomy and illustrate its applicability via two real-world cases. Finally, we conclude with our paper and highlight contributions and further research opportunities.

2. Research Background

2.1. Data Spaces

"Data sharing is the domain-independent process of giving third parties access to the data sets of others" (Jussen et al., 2023, p. 3688). Data spaces are technical enablers of inter-organizational data sharing (Otto, 2022a). They act as a conceptual framework for managing data in scenarios involving multiple stakeholders who exchange data (Solmaz et al., 2022). Data spaces operate without the need for physical data integration, preserving data at its source and providing on-demand accessibility. Rather than imposing a centralized schema, data spaces facilitate semantic linking and integration (Otto, 2022b). This federated architecture, in contrast to centralized digital platforms, offers novel possibilities for value creation within data ecosystems (Beverungen et al., 2022). In this 'shared playing field' of data ecosystems, users can collaborate and interact by discovering, storing, publishing, or consuming data for various purposes (Oliveira et al., 2019). To enhance collaboration, participants can engage in multiple data spaces simultaneously, assuming the roles of both data holders and users. This flexibility arises from the overlapping and nested nature of data spaces (Otto, 2022b). To access a data space, companies can use a connector (Gieß et al., 2023). Figure 1 visualizes how data sharing works within a data space and shows the focus of this paper, namely connectors.

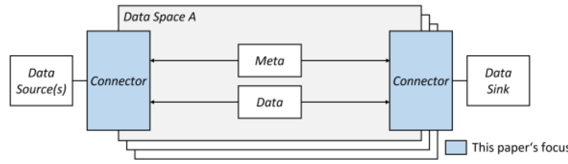

Figure 1. Data exchange within a data space extended from Otto and Burmann (2021).

2.2. Data Space Connectors

A connector establishes the connection to the data sources of the data providers, manages metadata of the data sources and the terms of use of the data, and sends or receives the data, including the terms of use (Otto & Burmann, 2021). Table 1 shows additional exemplary definitions of data space connectors focusing on their main functions.

Table 1. Exemplary definitions of data space connectors.

Definition	Source
"Connectors […] form a distributed network. Each connector runs different services, possibly from third parties, used for the exchange and processing of data.	Brost et al. (2018, p. 39)
"Connectors […] implement data endpoints at the sites of data providers and data consumers. "	Landolfi et al. (2019, p. 552)
"Connectors enforce data sovereignty based on data usage policies."	Firdausy et al. (2022b, p. 120)
"Data connectors are essential for enabling trust and interoperability in data sharing and exchange within data spaces, which are designed to provide data sovereignty."	Giussani and Steinbuss (2023, p. 5)

To summarize, a connector links different entities through secure exchange and trusted sharing of data, thus supporting data sovereignty (Landolfi et al., 2019). Data sovereignty refers to the idea of self-determination and the capability of a data provider to maintain control over their own data assets (Hellmeier & von Scherenberg, 2023). The decentralized data storage allows for the physical retention of each piece of data until it is transmitted to a reliable third party (Landolfi et al., 2019). To address security risks contracts that help to link policies to data assets are used. The contract definition contains an access control policy (e.g., only partners can see the assets), usage policy (e.g., data must stay in Europe), and asset selector (e.g., rules apply to specific assets) (Marino, 2022; Zrenner et al., 2019). These terms of use can be technically enforced by the connector (Otto & Burmann, 2021). Only if the data consumer agrees to the usage policy fully the data consumer may use the data (Landolfi et al., 2019).

3. Research Design

3.1. General Approach

We took a design-oriented approach for building our taxonomy to generate a comprehensive overview of design options (Möller, Haße, et al., 2021). To that effect, we draw from taxonomies as tools for describing and classifying key characteristics of objects (in our case: data space connectors) (Glass & Vessey, 1995). Building taxonomies for this purpose is an established practice, for example, in the field of data sharing business models (Schweihoff et al., 2023) or data spaces (Gieß et al., 2023).

To design our taxonomy, we adapted the method proposed by Kundisch et al. (2022), which extends Nickerson et al. (2013)'s method and integrates it into a broader framework of Peffers et al. (2007)'s DSR method. This adaptation aligns with the principles of design science research (Hevner et al., 2004). Following the approach outlined by Kundisch et al. (2022), we initially formulated the problem situation and outlined our research objectives, including the meta-characteristics that define the purpose of the taxonomy. Our taxonomy design involved iterative iterations using two distinct approaches: "conceptual-to-empirical" and "empirical-to-conceptual." These approaches facilitated the identification of common characteristics and dimensions, and we continued iterating until achieving theoretical saturation, which occurs when all selected ending conditions are satisfied (Nickerson et al., 2013). Subsequently, we evaluated the criteria for determining valid and useful taxonomies, showcased the practical applicability of our taxonomy through illustrative examples, and formalized our findings for effective communication purposes.

3.2. Problem Situation and Objectives

Considering the multifaceted requirements associated with informed decision-making in the creation of novel data space connectors and the selection of connectors that align with individual user needs, as well as the existing limitations within the current literature and the inherent heterogeneity of real-world objects, our objective is to advance the understanding of data space connectors by synthesizing and conceptualizing existing knowledge. This endeavor holds significance for diverse audiences, encompassing both researchers interested in the theoretical exploration of data space connectors and practitioners seeking practical implementation or selection of such connectors. As a consequence, we define the meta-characteristic as follows: 'key characteristics of data space connectors'. This meta-characteristic serves as the overarching principle for discerning and identifying essential attributes (Nickerson et al., 2013).

3.3. Taxonomy Design and Development

For the taxonomy design, one can select between approaches for 'conceptual-to-empirical' (deduction), in which the characteristics and dimensions are derived from relevant literature as well as 'empirical-to-conceptual' (induction) in which real-world objects (i.e., existing data space connectors) and practical insights (i.e., from experts) are analyzed for common characteristics and dimensions. In this paper, we performed three major iterations (see Table 2), which are described in the following sub-sections.

First Iteration (empirical-to-conceptual). To ensure that the research benefits from both research and industry knowledge to answer the research questions, the state-of-art, and the state-of-practice are conducted (Garousi et al., 2019). In doing so, we chose the E2C approach for the first iteration. Through the analysis of data space connectors, we aimed to extend our findings from the literature and provide further empirical evidence. For this, we collected empirical examples of data space connectors from the repository of the Industrial Data Space Association (the data connector report), which is dedicated to give an overview of existing connector use cases. However, these use cases are mainly based on two solutions, the Eclipse Data Space Components (EDC)[1] and the Dataspace Connector[2]. In the first iteration, we analyzed 23 data space connector use cases.

[1] https://projects.eclipse.org/projects/technology.edc [Accessed 02.06.2023].

[2] https://international-data-spaces-association.github.io/DataspaceConnector/ [Accessed 02.06.2023].

Furthermore, we included grey literature in this iteration. This is useful for both researchers and practitioners especially when the context has a large effect on the implementation and outcome which is typically the case in software engineering (Benzies et al., 2006; Gargani & Donaldson, 2011; Garousi et al., 2019). In total, we analyzed seven videos with an average length of 58 minutes and 30 seconds, three whitepapers, and four GitHub repositories.

Table 2. Overview of iterations and sources.[3]

<table>
<tr><th colspan="3">Source</th><th colspan="3">Iteration</th></tr>
<tr><th>Type</th><th>Origin</th><th>Amount</th><th>1</th><th>2</th><th>3</th></tr>
<tr><td>Use cases</td><td>Connector repository[4]</td><td>n=23</td><td rowspan="4">●</td><td rowspan="4">-</td><td rowspan="4">-</td></tr>
<tr><td rowspan="3">Other</td><td>Video data</td><td>n=7 (⌀ 58:30 min)</td></tr>
<tr><td>Whitepaper</td><td>n=3</td></tr>
<tr><td>GitHub repositories</td><td>n=4</td></tr>
<tr><td>Intermed. feedback</td><td>Data space expert</td><td>n=1</td><td>●</td><td>-</td><td>-</td></tr>
<tr><td rowspan="5">Scientific literature</td><td>ACM Digital</td><td>n=6</td><td rowspan="5">-</td><td rowspan="5">●</td><td rowspan="5">-</td></tr>
<tr><td>AIS eLibrary</td><td>n=7</td></tr>
<tr><td>IEEE Xplore</td><td>n=1</td></tr>
<tr><td>Scopus</td><td>n=7</td></tr>
<tr><td>Backward search</td><td>n=4</td></tr>
<tr><td>Workshop</td><td>Experts (3 developers, 2 product owners)</td><td>n=1 (90 min)</td><td>-</td><td>-</td><td>●</td></tr>
</table>

Intermediary feedback (evaluation of initial taxonomy). We had the chance to present our intermediary results to an expert from a data space umbrella organization. By doing this, we ensured that what we found in publicly available sources (e.g., websites and grey literature) is mirrored in the 'real' world of data space connector design. We derived new avenues to analyze the data and sharpened concepts and their wording from this feedback.

Second Iteration (conceptual-to-empirical). In the second iteration, we used the existing body of literature to obtain a comprehensive theoretical overview of data space connectors. Therefore, we reviewed the existing literature on data space connectors by adapting the established guidelines of (vom Brocke et al., 2009) and Webster and Watson (2002). 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, and Scopus). In the second step, we provided working definitions of data space connectors for further research. In the third step, the literature search provided 87 papers with the search string 'data space' AND 'connector'. Since a uniform notation of data space has not yet been established, we have also integrated another notation into our search string.

Upon gathering potentially relevant publications, we proceeded to the fourth phase, analyzing, and filtering the literature based on the exclusion criteria. First, the term data space and connector should either be a keyword, part of the title, or mentioned in the abstract. Next, all duplicates and papers which were not accessible as well as not yet published were excluded. Additionally, in this step, we applied several quality aspects as proposed by vom Brocke et al. (2009) and Cooper (1988). For example, a publication must be methodologically consistent and argue comprehensibly. The final literature corpus consisted of 29 papers, which we considered for detailed analysis. After applying all exclusion criteria, we identified 21 papers as relevant. Through conducting a backward search, an additional four papers became relevant. In total, we analyzed 25 papers.

Third Iteration (empirical-to-conceptual). Using the taxonomy from the second iteration as a foundation, we held a 90-minute group discussion with five experts (i.e., three senior developers and two product owners) working on several connectors during the third iteration to ensure that our taxonomy reflects the reality of data space connector design options. With the participant selection, we intended to get a comprehensive perspective on various data space connector facets. The comments we got confirmed our preliminary findings and assisted us in concretizing the traits for each dimension. To make the taxonomy clearer and self-explanatory, we specifically modified the language of a few traits. Because there were only minor changes in this iteration and we were able to meet all 13 ending conditions (e.g., each cell is unique and not repeated; all objects can be classified, Nickerson et al., 2013), we finalized the taxonomy (see Section 4).

3.4. Taxonomy Demonstration and Evaluation

In the final stage, we had the chance to present our results to an expert (i.e., product owner) to ensure that

[3] Following principles of Open Science: Full list available upon request.

[4] As listed in the IDSA Data Connector Report by Giussani and Steinbuss (2023)

our taxonomy addresses the right target group, namely product owners, and to classify their characteristics in our dimensions to ensure that our taxonomy reflects the reality of data space connector design options (see Section 5). The comments we got confirmed our preliminary findings. As part of the evaluation, we have carried out a promote-step of Kundisch et al. (2022). To be precise, we have moved the former characteristics of access control and GUI to a higher level of abstraction, the dimensions.

4. Results

This section presents the final taxonomy (see Table 3) in detail, which results from the taxonomy development process. The taxonomy serves as an answer to the research question of this study as it identifies the key dimensions and characteristics of data space connectors. The taxonomy consists of 9 dimensions with 25 characteristics. In addition to the individual dimensions and the corresponding characteristics, the right column shows whether a characteristic is exclusive (E) or non-exclusive (N). We visualize the taxonomy as a morphological box, as this is a common type of taxonomy visualization (Möller, Stachon, et al., 2021; Szopinski et al., 2020).

The **license (D_1)** (i.e., the availability of code) of a connector can be open source **(C_{11})** (e.g., Apache 2.0) (Altendeitering et al., 2022; Pampus et al., 2022), or closed source **(C_{14})**. We further distinguish these characteristics with copyleft **(C_{12})** as a subset for open source and extendable **(C_{13})** as a subset for closed source. Open source consists of data connector frameworks and generic solutions. Data connector frameworks address software developers to build their connector, like the EDC Connector available on GitHub (Eclipse Foundation, 2023). Generic open-source solutions offer companies opportunities for new businesses, such as the further development of the code into off-the-shelf solutions and turning it closed source. Closed source is mainly used for inherent solutions like the trusted supplier connector from German Edge Cloud (Juschkat, 2020) or the WeTech Smart Data Connector (Giussani & Steinbuss, 2023). Closed-source solutions can also be generic, which allows for configuration, add-ins, and individualization like the Nicos GAIAboX Connector (Giussani & Steinbuss, 2023). Furthermore, closed-source solutions can be off-the-shelf. For example, tritom connects Finnish food industry operators from agriculture to retailers and consumers (Hyyrönmäki et al., 2022).

A connector can be **deployed (D_2)** in different environments, such as on the edge **(C_{21})** (Brost et al., 2018), on-premises **(C_{22})** (e.g., local server), or in the cloud **(C_{23})** (Firdausy et al., 2022b). Figure 2 gives an overview of the different deployment options. Deploying a connector on-premises needs the connector provider to manage the connector on their own server, although a different organization can provide the software for the connector. With the advances of infrastructure as a service and data being stored in the cloud, the deployment of a connector can also be done via a cloud infrastructure, instead of on-premises (Firdausy et al., 2022a). Connectors can also be deployed on the edge (e.g., in manufacturing units),

Table 3. Design options for data space connectors (Note: E = Exclusive, N = Non-exclusive).

<table>
<tr><th colspan="2">Dimension</th><th colspan="4">Characteristics</th><th>E/N</th></tr>
<tr><td colspan="2">License</td><td>Open source</td><td>Open source (copyleft)</td><td>Closed source (extendable)</td><td>Closed source</td><td>E</td></tr>
<tr><td colspan="2">Deployment options</td><td>Edge</td><td colspan="2">On-premises</td><td>Cloud</td><td>N</td></tr>
<tr><td colspan="2">Service level</td><td>Connector as a service</td><td colspan="2">Platform as a service</td><td>Self-service</td><td>E</td></tr>
<tr><td colspan="2">Portability</td><td colspan="2">Agnostic</td><td colspan="2">Specific</td><td>E</td></tr>
<tr><td colspan="2">Access control</td><td>OAuth</td><td colspan="2">Basic auth</td><td>API key</td><td>E</td></tr>
<tr><td rowspan="2">Interoperability</td><td>Communication protocol</td><td colspan="2">IDS protocol (Multipart / IDSCP)</td><td colspan="2">Data space protocol (HTTPS)</td><td>N</td></tr>
<tr><td>Transfer protocol</td><td colspan="2">Determined</td><td colspan="2">Undetermined</td><td>E</td></tr>
<tr><td colspan="2">Graphical user interface</td><td colspan="2">Available</td><td colspan="2">Not available</td><td>E</td></tr>
<tr><td colspan="2">Supported data space specifications</td><td>Identity management</td><td>Catalog</td><td>Vocabulary</td><td>…</td><td>N</td></tr>
</table>

to allow for data exchange from hardware or sensor arrays (Brost et al., 2018).

Figure 2. Deployment options of data space connectors.

Connectors can be provided on different **service levels (D_3)**. First, a connector can be provided similarly to a software as a service offering, here referred to as connector as a service **(C_{31})**. This can be used as a plug-and-play solution that can be embedded into any form of application deployment (Sovity, 2023). Within the service, there can also be different service levels. For instance, the service level agreement can set a maximum response time or the type of communication (e.g., via ticket portal, e-mail, or telephone). Second, the supplier can offer a platform as a service (PaaS) **(C_{32})** solution, where the cloud environment is provided, and connectors realize peer-to-peer data sharing (e.g., Data Intelligence Hub (T-Systems, 2023)). Lastly, the user can develop a connector independently, using an established framework like the EDC (Eclipse Foundation, 2023), a so-called self-service **(C_{33})**. Competencies in the respective programming language are required if a framework is used. For illustration, to use the EDC, Java is demanded. However, self-service toolboxes also enable low-code application creation, making it possible for novice or non-technical users to connect (such as Apache StreamPipes) (Yallıç et al., 2022).

There is also a difference in the **portability (D_4)** of connectors. Connectors can be agnostic **(C_{41})** or specific **(C_{42})** in various contexts. For example, connectors can be platform agnostic, meaning the connector can be hosted on various cloud environments. An illustration is the EDC, which can be deployed in various environments. Gaia-X also prioritizes the development of universal standards and a technology-agnostic architecture (Braud et al., 2021). Portability can also be distinguished on the technology level. For instance, a connector can be agnostic or specific regarding the operating system (e.g., Linux or Mac OS). This dimension can also relate to the deployment option dimension, as choosing an edge device might affect the operating system, or choosing a cloud environment might affect the cloud provider (e.g., working with AWS or Azure).

The connector can intercept data processing applications and prohibit access (Altendeitering et al., 2022). This **access control (D_5)** can be realized via different frameworks. First, OAuth (i.e., open authorization) **(C_{51})** can implement the authentication (Sarabia-Jacome et al., 2019; Solmaz et al., 2022), which is a delegated authorization standard/ framework for REST/APIs. Another option is Basic Auth **(C_{52})** which is a basic access authorization method providing a username and password when making a request. Lastly, the API key **(C_{53})** uses identity and access management to manage access through a unique code for the programming interface. For example, the Dataspace Connector can natively establish a connection via HTTPS and optionally add Basic Auth or API key (Opriel et al., 2022).

Another dimension is **interoperability ($D_{6\&7}$)**, which focuses on the protocols. Implementing protocols is especially important to achieve more flexibility in general data transfer and enable functions like streaming (Ettl & Diemer, 2022). The interoperability dimension is divided into **communication (D_6)** and **transfer protocols (D_7)**. Communication protocols encompass a collection of rules and guidelines enabling the connection between two electronic devices for data exchange. In the data space context, a distinction can be made between the IDS protocol **(C_{61})** (e.g., Multipart or IDSCP) (Pampus et al., 2022) and the data space protocol **(C_{62})** (e.g., HTTPS) (Olms et al., 2020). The transfer protocol is a mechanism used to index and retrieve data from different databases (Boukhers et al., 2023). Within the taxonomy, we only refer to if a transfer protocol is determined **(C_{71})** or undetermined **(C_{72})**.

For human-computer interaction, a **graphical user interface (D_8)** (GUI) can enable the application of a data space connector. Using a GUI **(C_{82})** has advantages such as enabling direct system usage, supporting companies with limited digital capabilities, and enforcing strict usage policies (Opriel et al., 2021). Therefore, it is a very easy form to work with a connector via a desktop. This can be realized through a GUI to manage data assets within a connector (Sovity, 2023). Ultimately, there can be no user interface **(C_{82})** offered. This can lead to two options, one being that the connector can only be used via the command line, and two (Schütte et al., 2023), that the operator needs to develop their GUI. Lastly, a connector can provide features on top of the simple data exchange through **supported data space specifications (D_9).** An illustration is identity management **(C_{91})** which focuses on participant information based on organizational assessment (Landolfi et al., 2019). Identity management can be centralized or decentralized. Exemplary for centralized systems is X.509 (Pampus et al., 2022) or did:web (Kanal, 2022) for decentralized systems. Further examples of supported data space specifications were the catalog **(C_{92})** for data providers to showcase and explain their data resources, along with the terms and conditions governing their usage (Otto, 2022b) or vocabulary **(C_{93}),** which enables the storage of recognized ontologies that can be linked together to describe the data exchanged (Solmaz et al., 2022). As there are many other possibilities, such as the broker or clearing house component (Jarke et al., 2019), we also provide space for further supported data space specifications **(C_{94})**. However, logging of data transactions and enforceability of terms are only possible in applications that run within the connector. Neither can be guaranteed if the data leaves a system chain of the connector implementations (Otto & Burmann 2021).

5. Demonstration: Illustrative Application

In the following section, we showcase the practical applicability of the taxonomy by presenting two illustrative examples, as proposed by Kundisch et al. (2022) (see Figure 3). For each case, we had the advantage of accessing both the public data used to construct the taxonomy and direct communication with the product owner of the connectors. By collaborating with the product owner, who acted as our informant for both examples, we were able to effectively populate the taxonomy based on their extensive expertise and knowledge.

The two example cases are from sovity[5], a data space provider and operating company that enables data sharing and data-driven business models through a holistic software-as-a-service solution to connect companies in data ecosystems. The first case is the sovity community edition connector. It extends the EDC Connector's functionality and adds enterprise-ready managed services like out-of-the-box fully configured DAPS, and integration to already existing other data space technologies. The second case is the *connector as a service* solution from sovity, which also enables technically unsophisticated users to use connectors. While there are some similarities, such as the provision of a GUI to allow users to easily use the complex technology or the integration of data space components such as identity management, catalog, broker, or clearing house, there are also some differences between the versions. While the first connector, the community edition, is a free version the user must configure, the second connector is available for a monthly fee but additionally offers support and services not included in the free version. Another exemplary difference is the deployment location and portability, as the user of the community edition can freely choose where to deploy the connector, while the connector as a service solution runs in the cloud.

The taxonomy's application demonstrates that we can distinguish between various data space connector design possibilities inside the taxonomy, proving that it serves its intended function (Peffers et al., 2012). Additionally, the taxonomy demonstrates that we can identify key design features of data space connectors since they have diverse qualities and are created in various ways. As the figure shows how the taxonomy is applied, we have effectively executed our meta-characteristic: key characteristics of data space connectors.

6. Discussion and Conclusion

This paper developed a conceptual and empirical grounded taxonomy by including use cases, literature, and expert knowledge to capture and organize design options for data space connectors. As connectors

Demonstration case 1				
License	Open source	Open source (copyleft)	Closed source (extendable)	Closed source
Deployment options	Edge	On-premises	Cloud	
Service level	Connector as a service	Platform as a service	Self-service	
Portability	Agnostic	Specific		
Access control	OAuth	Basic auth	API key	
Communication protocol	IDS protocol (Multipart / IDSCP)	Data space protocol (HTTPS)		
Transfer protocol	Determined	Undetermined		
Graphical user interface	Yes	No		
Supported data space specifications	Identity management	Catalog	Vocabulary	…

Demonstration case 2				
License	Open source	Open source (copyleft)	Closed source (extendable)	Closed source
Deployment options	Edge	On-premises	Cloud	
Service level	Connector as a service	Platform as a service	Self-service	
Portability	Agnostic	Specific		
Access control	OAuth	Basic auth	API key	
Communication protocol	IDS protocol (Multipart / IDSCP)	Data space protocol (HTTPS)		
Transfer protocol	Determined	Undetermined		
Graphical user interface	Yes	No		
Supported data space specifications	Identity management	Catalog	Vocabulary	…

Figure 3. An illustrative application of the taxonomy of data space connectors, including the sovity community edition connector (left), and the sovity connector as a service solution (right).

[5] https://sovity.de/ [Accessed 02.06.2023].

enable cross-company data sharing in a sovereign way, they are an emerging artifact that is able to boost data-driven solutions that typically require access to large amounts of data, such as for business intelligence applications and reporting, or for fusing consumer health data with health records.

In terms of **scientific contributions**, our study contributes to the fast-expanding and mostly untapped field of data spaces and its components research. Our taxonomy's specific goals include using conceptual and empirical insights to add to the body of knowledge already available on design options for data spaces as well as to help define a common understanding of this complicated subject by focusing on the connector component and showing options on how to design a data space connector (e.g., Gieß et al., 2023; Otto et al., 2022). The taxonomy is a system for storing the information we have gathered from the corpus of literature and the analyzed connectors. Given the array of data space research projects (e.g., Mobility Data Space, 2021 or Catena-X, 2022), systematizing the knowledge about one of its key components is highly relevant for research. Naturally, these data space connectors are not uniform across projects but differ based on the context of the data space, making our research a relevant starting point for others to go deeper in their respective fields and contribute to sharpening the artifact.

This study's findings also have various **contributions for practitioners**. For instance, companies taking part in data space projects like Catena-X must be able to participate in data sharing at some point. To do this, they need a data space connector. Therefore, this taxonomy can highlight different design options and, as a result, help guide the choice of a suitable connector, regardless of whether the connector is built in-house or purchased as a software-as-a-service solution, as we gained promising results from the evaluation with practical partners. The taxonomy provides practitioners with a valuable framework for addressing the complex design challenges of choosing and developing connectors in the real world, and the associated opportunity for active participation in data spaces enables organizations to take advantage of cross-organizational data sharing effectively.

To fully understand the implications of our study, it is important to acknowledge the **limitations**. First, the reliance on grey literature as a source of information may introduce biases and inaccuracies due to the lack of oversight from commercial publishers. Second, while expert opinions were sought to supplement the information from grey literature, the potential for subjectivity cannot be entirely ruled out. Thus, other researchers may derive other features based on their influences, preferences, and biases. This is amplified because Iterations 1 and 2 were each conducted by one researcher and were not double-coded. Third, the taxonomy is formulated at a high level based on the target group, people who want to build products (i.e., future product owners). Fourth, the data space connector experts chosen for the workshop were also part of the selected use cases in the first iteration. Nevertheless, we have gained deeper insights through the workshop that were not publicly available via the website. Lastly, due to the rapid pace of technological change, the taxonomy should be seen as a snapshot in time. New dimensions and characteristics may become more relevant in the future, or previous ones may be omitted.

The limitations generally indicate possible **future research** paths. Initially, the taxonomy should be evaluated and further developed with other connector operating companies, such as Cofinity-X. The compatibility of the different data space connectors, so-called interoperability, is becoming important, which leads to future research questions. Issues such as scalability, load balancing, and availability may also become more important when choosing a connector in the future. Furthermore, different connectors can have different security levels, which wasn't investigated here. Another important future research question may be about which connector will prevail, as several companies are currently developing different connectors based on different frameworks. Besides, other and new connector approaches from the broader context of data sharing, data infusion, and data integration can be taken into account.

7. Acknowledgement

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

Agri-gaia. (2022). *Ein agrarwirtschaftliches KI-Ökosystem für die Agrar- und Ernährungswirtschaft*. https://www.agri-gaia.de/

Altendeitering, M., Pampus, J., Larrinaga, F., Legaristi, J., & Howar, F. (2022). Data sovereignty for AI pipelines. In L. E. Lwakatare (Ed.), *1st International Conference on AI Engineering - Software Engineering for AI: CAIN 2022 : 16-17 May 2022, Pittsburgh, Pennsylvania : proceedings* (pp. 193–204). IEEE.

Benzies, K. M., Premji, S., Hayden, K. A., & Serrett, K. (2006). State-of-the-evidence reviews: Advantages and challenges of including grey literature. *Worldviews on Evidence-Based Nursing*, *3*(2), 55–61.

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

Boukhers, Z., Lange, C., & Beyan, O. (2023). *Enhancing Data Space Semantic Interoperability through Machine Learning: A Visionary Perspective*.

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

Brost, G. S., Huber, M., Weiß, M., Protsenko, M., Schütte, J., & Wessel, S. (2018). An Ecosystem and IoT Device Architecture for Building Trust in the Industrial Data Space. In D. Gollmann (Ed.), *ACM Conferences, Proceedings of the 4th ACM Workshop on Cyber-Physical System Security* (pp. 39–50). ACM.

Catena-X. (2022). *Catena-X Automotive Network - Die Auffahrt zur Digital Economy*. https://catena-x.net/de/

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

Donald, A., Galanopoulos, A., Curry, E., Muñoz, E., Ullah, I., Waskow, M. A., Dabrowski, M., & Kalra, M. (2023). Towards a Semantic Approach for Linked Dataspace, Model and Data Cards. In Y. Ding, J. Tang, J. Sequeda, L. Aroyo, C. Castillo, & G.-J. Houben (Eds.), *Companion Proceedings of the ACM Web Conference 2023* (pp. 1468–1473). ACM.

Dong, X., Guo, B., Shen, Y., Duan, X., & Zhang, H. (2019). A Self-Controllable and Balanced Data Sharing Model. *IEEE Access (7)*, 103275–103290.

Eclipse Foundation. (2023). *Eclipse Dataspace Components*. https://github.com/eclipse-edc/Connector

Ettl, S., & Diemer, J. (2022). *Erweiterte Funktionalität durch den Eclipse Data Space Connector (EDC)*. https://catena-x.net/de/angebote/edc-die-zentrale-komponente-fuer-die-navigation

Firdausy, D., Silva, P., van Sinderen, M., & Iacob, M. (2022a). Semantic Discovery and Selection of Data Connectors in International Data Spaces: Proceedings of the Workshop of I-ESA'22. *I-ESA '22*.

Firdausy, D., Silva, P., van Sinderen, M., & Iacob, M. (2022b). Towards a Reference Enterprise Architecture to enforce Digital Sovereignty in International Data Spaces. In *2022 IEEE 24th Conference on Business Informatics (CBI)* (pp. 117–125). IEEE.

Gargani, J., & Donaldson, S. I. (2011). What works for whom, where, why, for what, and when? Using evaluation evidence to take action in local contexts. *New Directions for Evaluation*, *2011*(130), 17–30.

Garousi, V., Felderer, M., & Mäntylä, M. V. (2019). Guidelines for including grey literature and conducting multivocal literature reviews in software engineering. *Information and Software Technology*, *106*, 101–121.

Gieß, A., Möller, F., Schoormann, T., & Otto, B. (2023). Design Options for Data Spaces. In *Proceedings of the 31st European Conference on Information Systems*.

Giussani, G., & Steinbuss, S. (May 2023). *Data Connector Report*. Berlin, Germany.

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

Hellmeier, M., & von Scherenberg, F. (2023). A Delimitation of Data Sovereignty from Digital and Technological Sovereignty. In *Proceedings of the 31st European Conference on Information Systems*.

Hyyrönmäki, J., Kivelä, A., Järvinen, K., & Luoma-Kyyny, J. (2022). *DataSpace Europe*. https://www.dataspace.fi/en/homepage

IDSA. (2023). *Data Space Radar: Faster IDS breakthroughs are within range*. https://internationaldataspaces.org/adopt/data-space-radar/

Jarke, M., Otto, B., & Ram, S. (2019). Data Sovereignty and Data Space Ecosystems. *Business & Information Systems Engineering*, *61*(5), 549–550.

Jiang, Z., Thieullent, A., Jones, S., Perhirin, V., Baerd, M., Shagrithaya, P., Cecconi, G., Isaac-Dognin, L., Buvat, J., Khadikar, A., Khemka, Y., & Nath, S. (2021). *Data Sharing Masters: How smart organizations use data ecosystems to gain an unbeatable competitive edge*. Capgemini Research Institute. https://www.capgemini.com/de-de/insights/research/data-sharing-masters/

Juschkat, K. (2020). *German Edge Cloud ist IDS-ready*. https://www.elektrotechnik.vogel.de/german-edge-cloud-ist-ids-ready-a-952427/

Jussen, I., Schweihoff, J., Dahms, V., Möller, F., & Otto, B. (2023). Data Sharing Fundamentals: Characteristics and Definition. In *Proceedings of the 56th Annual Hawaii International Conference on System Sciences: January 3-6, 2023*.

Kanal, M. (2022). *Der Eclipse Dataspace Connector (EDC) – Architektur und Nutzen des Frameworks*. https://blog.doubleslash.de/der-eclipse-dataspace-connector-edc-architektur-und-nutzen-des-frameworks

Kolodziej, M., & Vazquez, P. (2023). *Enabling data sharing through data spaces and AWS*. https://aws.amazon.com/fr/blogs/publicsector/enabling-data-sharing-through-data-spaces-aws/

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

Landolfi, G., Barni, A., Izzo, G., Fontana, A., & Bettoni, A. (2019). A MaaS platform architecture supporting data sovereignty in sustainability assessment of manufacturing systems. *Procedia Manufacturing*, *38*, 548–555.

Marino, J. (2022). *EDC-Con 2022/01 - EDC Conceptual Overview and Architecture*. https://www.youtube.com/watch?v=IGd4oafLyAg

Martin-Jung, H. (2022). *Catena-X: Mehr Durchblick bei den Lieferketten*. https://www.sueddeutsche.de/wirtschaft/catena-x-bmw-sap-co2-fussabdruck-lieferkette-1.5605760

Mobility Data Space. (2021). *Mobility Data Space - Data Sharing Community*. https://mobility-dataspace.eu/de

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 S. Stieglitz, R. Schütte, & F. Ahlemann (Eds.), *Lecture Notes in Information Systems and Organisation. Innovation Through Information Systems: Volume I: A Collection of Latest Research on Domain Issues* (Vol. 46, pp. 523–538). Springer.

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

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.

Olms, C., Nissen, C., Schier, A., Leveling, J., Rademacher, R., & Hompel, M. ten. (2020). *Architektur einer adaptiven Plattform für unternehmens-übergreifende datenbasierte Dienste mit dem International Da-ta Spaces*.

Opriel, S., Berthold, T., & Silva, O. (2022). *Dataspace Connector*. https://github.com/International-Data-Spaces-Association/DataspaceConnector

Opriel, S., Möller, F., Burkhardt, U., & Otto, B. (2021). Requirements for Usage Control based Exchange of Sensitive Data in Automotive Supply Chains. In T. Bui (Ed.), *Proceedings of the Annual Hawaii International Conference on System Sciences, Proceedings of the 54th Hawaii International Conference on System Sciences.* Hawaii International Conference on System Sciences.

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

Otto, B. (2022b). A federated infrastructure for European data spaces. *Communications of the ACM*, *65*(4), 44–45.

Otto, B., & Burmann, A. (2021). Europäische Dateninfrastrukturen. *Informatik Spektrum*, *44*(4), 283–291.

Otto, B., Hompel, M. ten, & Wrobel, S. (Eds.). (2022). *Designing Data Spaces: The Ecosystem Approach to Competitive Advantage*. Springer International Publishing.

Pampus, J., Jahnke, B.-F., & Quensel, R. (2022). Evolving Data Space Technologies: Lessons Learned from an IDS Connector Reference Implementation. In T. Margaria-Steffen (Ed.), *Lecture Notes in Computer Science: Vol. 13704. Leveraging applications of formal methods, verification and validation: 11th international symposium on leveraging applications of formal methods, ISoLA 2022, Rhodes, Greece, October 22-30, 2022: proceedings* (Vol. 13704, pp. 366–381). Springer.

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

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.

Sarabia-Jacome, D., Lacalle, I., Palau, C. E., & Esteve, M. (2019). Enabling Industrial Data Space Architecture for Seaport Scenario. *IEEE 5th World Forum on Internet of Things (WF-IoT)*.

Schütte, J., Brost, G., & Wessel, S. (2023). *Datensouveränität im Internet der Dinge - Der Trusted Connector im Industrial Data Space*. München. https://www.aisec.fraunhofer.de/content/dam/aisec/Dokumente/Publikationen/Studien_TechReports/deutsch/IDS-Paper_Datensouveraenitaet.pdf

Schweihoff, J., Jussen, I., Dahms, V., Möller, F., & Otto, B. (2023). How to Share Data Online (fast) – A Taxonomy of Data Sharing Business Models. In *Proceedings of the 56th Hawaii International Conference on System Sciences*, Hawaii, USA.

Solmaz, G., Cirillo, F., Fürst, J., Jacobs, T., Bauer, M., Kovacs, E., Santana, J. R., & Sánchez, L. (2022). Enabling data spaces. In N. Laoutaris & M. Mellia (Eds.), *Proceedings of the 1st International Workshop on Data Economy* (pp. 42–48). ACM / Association for Computing Machinery.

Sovity. (2023). *EDC Connector*. https://github.com/sovity/EDC/blob/main/README.md

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

T-Systems. (2023). *Data Intelligence Hub*. https://dih.telekom.com/en/products

vom Brocke, J., Simons, A., Niehaves, B., Reimer, K., Plattfaut, R., & Cleven, A. (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*.

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

Yallıç, F., Albayrak, Ö., & Ünal, P. (2022). Asset Administration Shell Generation and Usage for Digital Twins: A Case Study for Non-destructive Testing. In *Proceedings of the 3rd International Conference on Innovative Intelligent Industrial Production and Logistics* (pp. 299–306). SCITEPRESS - Science and Technology Publications.

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