Data Sharing Fundamentals: Characteristics and Definition

Conference Paper · January 2023

CITATIONS
2

READS
1,026

5 authors, including:

Ilka Jussen
Technische Universität Dortmund
7 PUBLICATIONS **8** CITATIONS
SEE PROFILE

Julia Schweihoff
Technische Universität Dortmund
6 PUBLICATIONS **4** CITATIONS
SEE PROFILE

Valentin Dahms
Technische Universität Dortmund
4 PUBLICATIONS **1** CITATION
SEE PROFILE

Frederik Möller
Technische Universität Braunschweig
60 PUBLICATIONS **580** CITATIONS
SEE PROFILE

COVER SHEET

This is the author's version of the following source:

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

-Author's Accepted Manuscript-

Data Sharing Fundamentals: Definition and Characteristics

Ilka Jussen
TU Dortmund University
ilka.jussen@tu-dortmund.de

Julia Schweihoff
TU Dortmund University
julia.schweihoff@tu-dortmund.de

Valentin Dahms
TU Dortmund University
valentin.dahms@tu-dortmund.de

Frederik Möller
TU Dortmund University
and Fraunhofer ISST
frederik.moeller@tu-dortmund.de

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

Abstract

The importance of data as a key resource is a universal theme dominating social and business life. In this regard, inter-organizational data sharing shines in a new light prompting businesses to leverage the potential of their own data. However, it is still unclear what data sharing actually entails, i.e., what it means, what its potentials are, and what barriers organizations must overcome. In short, it lacks conceptual clarity and a clear description of its characteristics. The conceptual ambiguity and the synonymous use with data exchange in the literature are particularly problematic, which prevents a targeted conceptualization and use. The paper starts precisely at this point as it proposes a unifying definition and characteristics of data sharing. We report on a systematic literature review characterizing data sharing and delineating it from data exchange.

Keywords: Data Sharing, Data Exchange, Definitions, Characteristics, Literature Review

1. Introduction

The importance of data as a resource and data sharing is currently an omnipresent topic spurring a variety of European data-sharing initiatives in many industries, such as *automotive* (e.g., Catena X (2022), *mobility* (e.g., Mobility Data Space (2022), or *manufacturing* (e.g., SCSN (2022)). Although it is well known that data sharing can boost business success, many organizations still avoid giving others access to their data and maintain their data silos (Gartner Inc., 2021; Gelhaar et al., 2021). A study from the European Commission illustrates this problem. On the one hand, the volume of data within the EU increased from 33 zettabytes in 2018 to a forecast of 175 zettabytes in 2025. On the other hand, 80% of industrially generated data is currently left untapped, clearly indicating the need for investigating this issue (European Commission, 2022). Another study commissioned by the European Commission adds that data sharing between different companies can increase the economic opportunities of companies in terms of their products or services and internal efficiency. Given the growing number of emerging data suppliers and data users, leveraging the potential of data sharing is paramount (Arnaut et al., 2018).

Many examples show how collecting and sharing data is increasingly important, such as public transport, healthcare, or logistics (Becker et al., 2021; Compton, 2020; Foraker et al., 2020). Thus, sharing data is intertwined with economic interests and generating value for the public good (Susha et al., 2017). Data sharing between individuals, companies, and governments, being of great importance, is not an entirely new concept, but it has been well known for many years (Zhong et al., 2011). The potential of data sharing is relevant for businesses regardless of size because different players can use the same data resource for other applications (Snowflake, 2022).

By looking at the issue of data from a political perspective, it is becoming clear that this is not just an economic issue. For example, there has been active work on corresponding legislation in recent years, such as the *Data Governance Act* and the *Data Act*, which provide a fair legal framework for sharing data (European Commission, 2022; Mildebrath, 2021). The importance of these legal frameworks is also shown by the quote from Margrethe Vestager, *Executive Vice-President for a Europe fit for the Digital Age* (European Commission, 2022):

> "*We want to give consumers and companies even more control over what can be done with their data, clarifying who can access data and on what terms.*"

However, research on data sharing is so far in its infancy, resulting in a lack of a clear understanding of the term and related activities. In particular, the conceptual distinction from *data exchange* makes a

shared understanding and targeted research regarding the potential of data sharing difficult. In our paper, we tackle the issue of conceptual blurriness of data sharing based on a systematic literature review. To the best of our knowledge, there is no publication characterizing data sharing and delineating it from data exchange. Because of the above, we pursue the following research questions:

Research Question 1: *How to define and characterize data sharing and exchange?*

Research Question 2: *How to distinguish data sharing and data exchange?*

To answer these research questions, the paper is structured as follows. First, we will briefly explain the theoretical background for this paper, focusing on the importance of the term data sharing as well as the often-synonymous use of the term data exchange. Following this, Section 3 explains the design of the literature review in detail to answer the research questions above. Section 4 then presents the summarized results of the literature review to develop the definitions and characteristics of data sharing and data exchange and distinguish them from each other specifically. The last two sections, 5 and 6, discuss the results, limitations, and an outlook for further research.

2. Data Sharing

The increasing importance of data in the business context is also seen in the shift towards data as a core resource in business models (Engelbrecht et al., 2016; Hartmann et al., 2014). Data is an asset in most companies (Parmar et al., 2014). However, this data is often only used to a limited extent or frequently only in an internal company context. Data sharing as part of data-driven business models is often associated with knowledge risks (Fruhwirth et al., 2019). For these reasons, among others, data silos occur, which on the other hand, is also legally prescribed by politics in the case of mergers of companies (Kraemer et al., 2021). Sharing data with others can ensure that companies can extract more information from data than companies can do on their own (Richter & Slowinski, 2019; van den Broek & van Veenstra, 2015). Factors such as the improved possibilities in predictive maintenance or, in the specific application area of transportation, the savings in emissions or costs must also be considered (Nokkala et al., 2019). Other reasons that can make sharing data between different actors lucrative include increasing data quality and productivity (de Corbière & Rowe, 2010). In general, sharing data can improve existing services and develop new services and business models (Richter & Slowinski, 2019).

Nevertheless, in addition to the potential of data sharing, a variety of barriers currently prevent companies from sharing data with other stakeholders or within their ecosystem (e.g., Gelhaar et al., 2021). Some companies believe their processes are working without sharing data; accordingly, they assume they do not need to share data (Parvinen et al., 2020). In part, this also goes hand in hand with a lack of knowledge about the opportunities offered by data sharing (van den Broek & van Veenstra, 2015). Furthermore, factors such as a lack of trust in data security or the fear of losing control over their sensitive data are the main reasons why many companies are reluctant to share data (Bastiaansen et al., 2019; Gelhaar & Otto, 2020; Knol et al., 2014). The lack of knowledge of how to value one's data financially, since no one wants to give their data away for free, as well as the dependency on the data quality that the data owner makes available to the user, significantly influences the further use of the data and causes further confusion among companies (Knol et al., 2014; Mosch et al., 2020; Parvinen et al., 2020) The motivation to collect, analyze and share data is well known in various application areas. In this context, data sharing is defined with different emphases, elaborated on in more detail in Section 4.1. This implies the understanding that data sharing includes all possible data flows between companies in a B2B context or also with customers (Richter & Slowinski, 2019).

In the literature, data sharing and exchange are included under the generic term data collaboratives, which are cross-sector initiatives that collect, use, and process data together to solve public problems (Gasco et al., 2018; Susha et al., 2017).

3. Literature Review

This paper reports on a systematic literature review uncovering the fundamentals of data sharing (definitions and characteristics). We followed the well-accepted guidelines for literature reviews by Webster and Watson (2002). Our first task **(1)** was to search for adequate literature. We decided on two databases. First, we opted to use AISeL as the premier database for Information Systems (IS) research hosting some journals and the top IS conferences (e.g., ECIS, HICSS, and ICIS). Second, we complemented the database by using Scopus to search for leading IS journals based on their ranking in VHB (A+B), complementing what was available in AISeL.

In the second task **(2)**, we identified keywords that would produce results fitting our research question.

An initial finding from the literature review (and our experience) is the synonymous and overlapping use of the terms *data sharing* and *data exchange*. Since one of our goals is to differentiate or classify the two terms, we searched for both keywords. **Table 1** lists the keywords and the results they produced for each database. We used the Title and Abstract search in both databases, producing 154 (AISeL) and 106 (Scopus) papers.

Table 1. Literature Review - keywords and results

Database	Keywords	Results
AiSeL	title: "Data Sharing" OR abstract: "Data Sharing" OR title: "Data Exchange" OR abstract: "Data Exchange"	154
Scopus	ABS ("Data Sharing") OR TITLE ("Data Sharing") OR ABS ("Data Exchange") OR TITLE ("Data Exchange")	106
3 Iterations	AiSeL & Scopus	155

In a third step **(3)**, we reviewed and filtered the papers in three iterations. First, we assessed the thematic fit of the paper, excluding papers that do not focus on data sharing or data exchange. We also excluded papers not available in English or accessible to us. In the second iteration, the papers were reviewed in their entirety and roughly skimmed to see if they addressed the thematically relevant content. This reduced the number of relevant papers drastically to 57 papers at *AiSeL* and 78 papers at *Scopus*. Based on the result of the second iteration, duplicate papers were removed in the third iteration, and relevant papers were added through *forward and backward searches*. The result was 155 papers, which were further investigated regarding definitions and characteristics of data sharing and data exchange. These 155 papers were all analyzed, and 60 papers of them were classified as helpful in terms of naming definitions or characteristics for further evaluation. Papers that only marginally touched on the topic and did not show any relevant content were removed. Correspondingly, papers that were too specific to a particular software, such as the findings cited were not classified as generally valid in the context of data sharing and data exchange, were also removed.

The publications used for this paper have been sorted according to their publication date in **Figure 1**. Here, it is clear from the literature examined that data sharing has been gaining more and more traction over the last few years compared to data exchange.

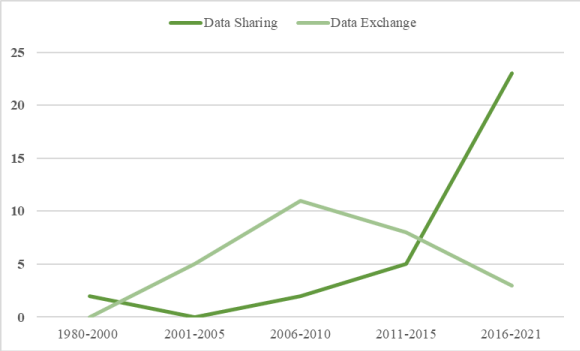

Figure 1. Timeline of used publications

Furthermore, this figure shows that data sharing is not an entirely new topic but has been known in the literature since the 1980s. To analyze the literature regarding the definitional bases and characteristics of data sharing and data exchange, a concept matrix was developed according to Webster and Watson (2002).

4. Findings

4.1 Definition of Data Sharing

Table 2 shows some example definitions of data sharing. When analyzing the existing definitions in the literature, it was evident that the term data sharing is defined with different focuses.

Table 2. Definitions of data sharing

Source	Definition
Vesselkov et al. (2019, p.303)	"**Data sharing** can allow organizations to access complementary data sources and help them develop innovative applications and services. […] We considered **data sharing** as either third parties' opening the data they own or their consuming the data from other providers."
Brathwaite (1983, p.40)	"In a database where **data sharing** is the norm of the owner of the data […] must delegate responsibilities […] with sole purpose of using data to fulfill the goals of the corporation."
Nokkala et al. (2019, p.2)	"… we define **data sharing** as an exchange of data between different stakeholders, and when happening on a platform, as lateral data exchange in a network."

Several definitions (Chowdhury et al., 2018; Vesselkov et al., 2019) explain the aspect of permitted "access" to datasets as a part of the definitions. Some authors also address the different roles involved in data sharing. Examples are data owners (Brathwaite, 1983) or the data custodian (Chowdhury et al., 2018). In addition, individual definitions also deal with the

application of data sharing, such as the joint development of new and innovative applications and services and their use in support of operations and analysis (Sarathy & Muralidhar, 2006; Vesselkov et al., 2019). Some definitions, like Nokkala et al. (2019)'s, even use the term *data exchange* directly to explain the term *data sharing* and indicate that *data exchange* is a sub-process of *data sharing*. Some authors do not define the term data sharing generally but instead in terms of an application area. For example, Hao et al. (2020) describe government data sharing as a process of passing data from one department to another via a platform. This is done in compliance with legal requirements to exploit the value of the data fully.

Various forms of data sharing are used in the literature. While some forms are named after the technologies used (e.g., blockchain-based data sharing, cloud data sharing), other forms refer to the scope of application or the actors involved in data sharing. These include *business data sharing*, *government data sharing*, *personal data sharing*, *peer-to-peer data sharing*, *inter-organizational data sharing*, or *cross-domain data sharing*. This multitude of forms of data sharing shows that this topic is becoming increasingly relevant in the most varied fields of application and is influenced by the interests of a wide range of researchers and stakeholders.

Based on the literature review, we present a synthesized definition bringing together all relevant aspects of the existing literature:

> *Data sharing is the domain-independent process of giving third parties access to the data sets of others. These third parties may be other companies (usually not direct competitors), individuals, or public institutions. The shared data is often used to develop new applications and services. The expectation is to be compensated financially or through other benefits (e.g., receiving data) for providing the data. What the data may be used for and how it is made available is determined within the framework of the (legal) agreements between the data providers, data consumers, and other roles, depending on the use case.*

4.2 Characteristics of Data Sharing

Complementing the definitory part of the systematic literature view, we also extracted a set of characteristics to understand data sharing more in-depth. That is also a necessary step to distinguish it distinctly from data exchange. In the following, we focus on data sharing characteristics. **Table 3** lists the seven data sharing characteristics that we have extracted from the literature corpus. We differentiate between *characteristics* and more detailed *sub-characteristics.*

Table 3. Characteristics of Data Sharing

Characteristics	Sub-Characteristics
Infrastructure	• Data Marketplace • Data Sharing Platform
Technologies	• Blockchain • Cloud computing
Legal Concepts	• Data Sharing Agreements *(Terms of Use, Usage Contracts…)*
Services	• Match-making • Privacy-as-a-Service • Analytics
Customer groups	• B2B • C2B/B2C • G2G/G2B/G2C
Actors	• Producer • Data Buyers/User • Data Owner • Data Consumer • Data Provider • Data Prosumer • Data Custodian
Data types	• Anonymous personal data • Metadata • Aggregated data

First, papers imply the necessity for data sharing **infrastructure**. In our analysis, we identified two relevant dimensions, i.e., *data marketplace* and *data sharing platform*. For example, Abbas et al. (2021) and Agahari et al. (2021) focus on the need for data marketplaces to mediate when sharing data to different domains (Pujol Priego & Wareham, 2020). Other authors like Vesselkov et al. (2019) and Kalkman et al. (2019) mention platforms or data sharing platforms as necessary infrastructure. Concerning infrastructure, the terms marketplaces and platform are sometimes used synonymously in the literature (Abbas et al., 2021; Agahari et al., 2021).

Another characteristic is using specific **technologies** such as *blockchain* or *cloud technologies* to enable data sharing. Authors such as (Abbas et al., 2021; Coelho et al., 2021; Karger et al., 2021) recommend using blockchain, for example, in the form of blockchain-based marketplaces, to create added value for users (Abbas et al., 2021). With the help of blockchain, processes are presented transparently so that participants can openly see the interactions within their network (Coelho et al., 2021). Blockchain enables the secure transmission of data and, as a decentralized procedure, represents an alternative to previous centralized and established storage of data

(Karger et al., 2021). Other technologies, such as cloud computing, are described as very cost-efficient and flexible in data sharing (Samanthula et al., 2015).

The characteristics also include **legal concepts** required to facilitate data sharing. Therefore, the so-called *data sharing agreements* and *usage contracts* are mentioned in the literature. Developing a data sharing agreement implies sub-processes of defining terms of use, usage, and access controls and negotiating commercial and legal agreements (Bastiaansen et al., 2019). They are needed to create trust between the different actors when sharing data (D'Hauwers & Walravens, 2021). The usage contracts, just like the terms of use, represent partial aspects of the data sharing agreement. In the usage contract itself, access control policies and usage control policies are combined to answer questions such as who may access the data or how the data may be used (Bastiaansen et al., 2020).

In addition, the literature mentions various **services** that are enabled as part of data sharing. While Abbas et al. (2021) mention the service of *match-making* between data providers and data buyers, Agahari et al. (2021) discuss the possibility of offering *privacy-as-a-service* via marketplaces or platforms as a business model. In addition, Sarathy and Muralidhar (2006) cite the possibility of carrying out precise data *analyses* based on data sharing to derive meaningful information from this knowledge.

The literature finds different **customer groups** between which data sharing can occur. For example, data sharing in a *business-to-business (B2B)* context between different companies may be subject to different requirements or challenges than data sharing between individuals (*customers*) and companies (*B2C or C2B*) (Abbas et al., 2021; Agahari et al., 2021). In addition, the literature reviewed also mentions the possibility of data sharing between governments (G2G), between governments and businesses (G2B), or individuals (G2C) as a potential implementation of data sharing (Hao et al., 2020; Omar et al., 2014; Susha et al., 2017). This shows that many actors and segments can engage in data sharing across domains (Gelhaar & Otto, 2020; Pujol Priego & Wareham, 2020; Zhong et al., 2011).

In addition, the various customer groups that cooperate in data sharing are relevant to different **actors**. The literature specifically identifies the *data producer* (Fisher & Fortmann, 2010), *data buyers/user* (Agahari et al., 2021; Fisher & Fortmann, 2010), *data owner* (Bastiaansen et al., 2020; Brathwaite, 1983; Samanthula et al., 2015), *data provider* (Agahari et al., 2021; Bastiaansen et al., 2019; Gelhaar et al., 2021; Vesselkov et al., 2019), *data consumer* (Chowdhury et al., 2018; Samanthula et al., 2015; Vesselkov et al., 2019) and *data prosumer* (Vesselkov et al., 2019). The role of the data prosumer is assumed by actors who act both as data consumers and data providers (Vesselkov et al., 2019). Some publications also refer to an external trust organization, also called a *data custodian*, which is responsible for the implementation of data sharing but is not itself involved in the data sharing process (Bastiaansen et al., 2019; Chowdhury et al., 2018).

The last and final characteristics are the relevant **data types** for data sharing. In the context of data sharing, we have identified the data types of *anonymous personal data*, *meta data*, and *aggregated data* in the literature. For example, anonymized personal data are sold or transferred from individuals to companies in C2B data trading (Abbas et al., 2021). According to (Bastiaansen et al., 2019), metadata is data generated during the execution of processes supporting data sharing. This data includes, for example, data sharing agreements, terms of use, or logging or provenance data from certain processes (Bastiaansen et al., 2019). Aggregated data can include, for example, data collected via sensors that are subsequently further analyzed (Nokkala et al., 2019). The above characteristics illustrate the many facets of data sharing that would all merit deeper analysis. What has become apparent is that there is no one way to tackle data sharing. Instead, there are many ways to share data and various issues to be considered.

4.3 Definition of Data Exchange

When analyzing the definition of *data exchange*, it was evident that the definitions are very similar in terms of content, in contrast to the definitions of *data sharing*. **Table 4** lists some definitional approaches. Across the different definitional approaches, data exchange implies the technical transformation process of data. In the context of this transformation process, data available under the so-called source scheme are transferred or transformed into another scheme, also called the target scheme (Afrati & Kolaitis, 2008; Cate & Kolaitis, 2010; Fagin et al., 2005; Fagin et al., 2011; Golshanara & Chomicki, 2020; Kolaitis, 2005; Libkin, 2006). In addition to the definitions, the literature also mentions some relevant terms that need to be considered in the context of data exchange. For example, the term *'data translation'* is equated with data exchange by authors like Kolaitis (2005).

Data exchange is also distinguished from *'data integration'* in several papers because data integration is a sub-process of data exchange. The link between the processes of data sharing and data exchange will be discussed below (Amano et al., 2014; Arenas et al., 2014; Cate & Kolaitis, 2010; de Corbiere & Rowe,

2010; Fagin et al., 2005; Kolaitis, 2005). Further terms are *'data exchange setting'* and *'data exchange problem'*. However, these terms are only used to differentiate the general definition of data exchange in a technical context and are therefore not presented in more detail in this paper (Fagin et al., 2005).

Table 4. Definitions of data exchange

Source	Definition
Afrati and Kolaitis (2008, p. 129)	"**Data exchange**, also known as data translation, can be succinctly described as the problem of transforming data structured under one schema, called the source schema, into data structured under a different schema, called the target schema […]. **Data exchange** is typically formalized using schema mappings between the source schema and the target schema."
Golshanara and Chomicki (2020, p. 1)	"**Data exchange** is the problem of transforming data that is structured under a source schema into data structured under another schema, called the target schema, so that both the source and target data satisfy the relationship between the schemas. "

Like data sharing, the term data exchange is used in different forms in the literature. It becomes clear that the forms of data exchange focus on various technical possibilities for implementation (e.g., *electronic data exchange, automated data exchange, web-based data exchange, XML data exchange*) (Arenas & Libkin, 2008; de Corbiere & Rowe, 2010; Elgarah et al., 2005; Meyer et al., 2015; Nicolaou et al., 2013). Other authors (Kuerschner et al., 2009; Link et al., 2017; Liu et al., 2001) focus on the actors involved in the forms of data exchange they use. Kuerschner et al. (2009) describe *'External Data Exchange'* as a relationship in which data is exchanged between different companies. Link et al. (2017) discuss the term *'Open Community Data Exchange'* to explain the exchange of data within a group, in this case, researchers. Another focus show describes the *'Fair Data Exchange'* as a form of data exchange in which special attention is paid to fairness and trust between the actors (Liu et al., 2001). For example, a trusted third party should be integrated in addition to the actors directly involved in the data exchange.

4.4 Aligning data sharing and data exchange

To distinguish the two terms data sharing and data exchange, not only with the help of their definitions, we will now also look at the associated processes of the two terms (see **Figure 2**). This developed process view was derived from the findings of several publications on relevant processes of data sharing (Abbas et al., 2021; Bastiaansen et al., 2019; Fisher & Fortmann, 2010; Samanthula et al., 2015; Sarathy & Muralidhar, 2006) and data exchange (Amano et al., 2014; Arenas et al., 2014; de Corbiere & Rowe, 2010; de Corbière & Rowe, 2013; Fagin et al., 2005; Kolaitis, 2005).

The process of data sharing starts with the preparation of a data set. The sub-processes of data generation and consolidation can be assigned to this process. Finally, in this step, the data is published, for example, on a platform. The next step is establishing a data-sharing agreement between the data provider and the potential data consumer. Part of this process step is to clarify the relevant legal issues. Topics such as ensuring that only authorized actors have access to the data are defined, and the usage policy specifies the further use of the data.

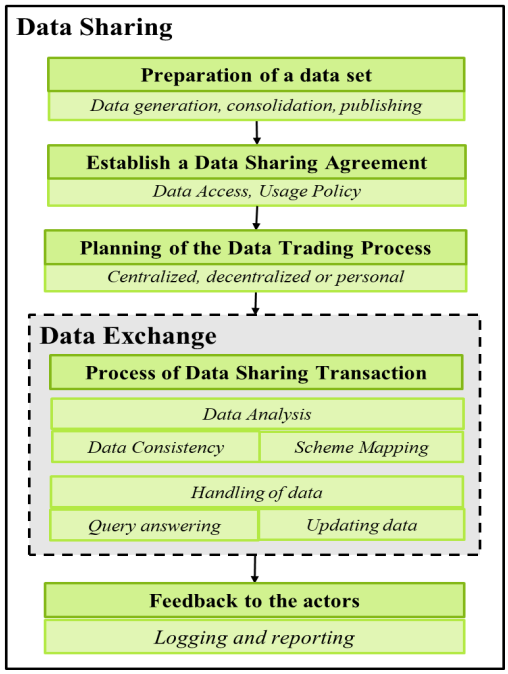

Figure 2. A process perspective on data sharing and data exchange

Following this, the next step is to start **planning the data trading process**. In the literature, a distinction is made between *centralized, decentralized*, and *personal* data trading, depending on the exact design of the data sharing and the involved actors (Abbas et al., 2021). For example, in centralized data trading, the data is shared via a platform. The situation is similar to decentralized data trading, where the architecture is decentralized. Personal data trading is different, where individuals sell their data to companies.

With the transition to the next process step, the processes overlap with *data exchange*. This includes the sub-processes that contribute to the higher-level **process of the data sharing transaction**. After reviewing the literature, the tasks of this process step can be divided into the topics of *data analysis* and *handling of data*. *Data analysis* includes the tasks of checking the *consistency of the data* and carrying out the *scheme mapping*. Scheme mapping is the process of converting the source schema into the target schema. This sub-process often provides the core of the definitions of data exchange. The *handling of data* process step includes the tasks of *query answering* and *updating data*. Ultimately, both tasks serve to check whether the data still has the same information after transfer to the target schema as it did in the source schema.

After the data exchange has also been carried out, the technical process and **feedback** are finally sent to the actors in the data sharing process. This implies precise *logging* and subsequent *reporting* to the actors about the completed data sharing process. This completes the last step of the data sharing process, and new data sharing activities can be carried out following a similar process flow.

5. Discussion

Taking a detailed look at our findings, we identified great differences between data sharing and data exchange. While sometimes used synonymously, both concepts vary greatly in their use in the literature corpus and what content they aim to convey. Data exchange is used to refer to the technical conversion process of data, while data sharing is 'the big picture, of granting access to data. Subsequently, while data exchange is technical, *data sharing* has distinct organizational components, as it intersects with broader themes (e.g., business models, usage policies, or services, actor-to-actor transactions). Subsequently, data sharing is the superordinate term that requires, as one process, technical data exchange (see **Figure 2**).

The fact that data sharing has been used more and more frequently in the literature, especially in recent years, shows its emergent importance. On the contrary, publications dealing with data exchange have decreased constantly. Referring to our introduction section, we highlighted large-scale projects dealing with data sharing, which could offer an explanation. We can also see this in the many application areas that address data sharing in more detail. From large and general areas of application such as *research* (Coelho et al., 2021; Köster et al., 2020; Pujol Priego & Wareham, 2020), *education* (Chowdhury et al., 2018), *health* (Kalkman et al., 2019; Vesselkov et al., 2019; Zhong et al., 2011), supply chain (Knol et al., 2014), or the *public domain* (Vesselkov et al., 2019), but also for specific areas of application data sharing is becoming increasingly important. More specific application areas include *future transport* (Becker et al., 2021), *biotechnology* (Fisher & Fortmann, 2010), *food traceability* (Gasco et al., 2018), and *additive manufacturing* (Morar & Kemper, 2016). However, after the analysis, it can also be stated that the two terms are firmly linked thematically and in terms of process. Nevertheless, the definitions of the terms should not be used synonymously, as the scope of the definitions, as presented in the process view, differs considerably.

6. Contributions, Limitations, Outlook

To conclude our paper, we would like to summarize our research's contribution. We have synthesized a definition of data sharing based on the literature reviewed and, in contrast to previous publications, brought together different perspectives and aspects of data sharing. The paper has multiple contributions to research and practice. In this paper we have come to the conclusion that data exchange can be a part of data sharing. However, the implementation in practice might also show that the concepts can stand alone.

In terms of **research contributions**, our work is one of the first studies investigating the conceptual nature of data sharing. Against the background of the societal and economic importance of data sharing, we see this as a significant contribution to establishing the 'playing field' for research. Through the definition, we enable future research to draw from a shared understanding, clearly distinguishing data sharing from data exchange. The characteristics would allow researchers to select specific areas of data sharing that they can analyze in-depth. Furthermore, we have also clearly distinguished the terms data sharing and data exchange from each other via their definitions. The presentation of relevant characteristics also underlined the comprehensive meaning of the term data sharing. In addition, we have put both terms into a concrete context with the help of the process perspective. In this way, we have clearly worked out that data exchange is the technical part of the higher-level and more organizational-focused data sharing term.

In terms of **practical contributions,** we shed light on an issue that offers potential for companies like few others at this point. At the very least, we highlight the necessity to consider data sharing a key activity for companies. We also provide characteristics managers can use as a sort of 'checklist' that supports understanding what data sharing means for their

business and helps them not to oversee important issues. For example, contrary to a simple provider and receiver relationship, we outline multiple actor roles prevalent in the literature that could benefit actors (e.g., data buyer or seller).

The paper is subject to **limitations**. First, even though we strive for a comprehensive review, a systematic literature review, naturally, is prone to missing some articles. Second, filtering and assessing the suitability of papers and their inclusion in the literature sample requires judgment. Both the chosen databases and the used search terms significantly influence this paper's developed results (definitions and characteristics). In addition, the authors affected the further procedure by their personal classification of whether a paper is thematically suitable or not.

The paper offers multiple ways to influence **further research**. First, we identified generic data sharing characteristics that would merit deeper analysis with various research methods. For example, deep-diving into the roles of data sharing with a qualitative interview study or a questionnaire could result in highly beneficial research. However, there is currently a lack of tools or guidelines that would reduce the complexity of the topic for the end-user and thus make the beginnings of data sharing more accessible. While the legal framework is being pushed more and more forward, especially by political leaders, there are still questions regarding data monetization. According to our work, the current obstacles that prevent companies from participating in data sharing must be solved through the more differentiated elaboration of incentives for data sharing.

Acknowledgement

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

References

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

Afrati, F., & Kolaitis, P. G. (2008). Answering aggregate queries in data exchange. *Proceedings of the Twenty-Seventh ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Vancouver, Canada*, 129–138. https://doi.org/10.1145/1376916.1376936

Agahari, W., Dolci, R., & Reuver, M. de (2021). Business Model Implications of privacy-preserving technologies in data marketplaces: The case of multi-party computation. In *Proceedings of the 29th European Conference on Information Systems,* A Virtual AIS Conference.

Amano, S., David, C., Libkin, L., & Murlak, F. (2014). XML Schema Mappings. *Journal of the ACM*, *61*(2), 1–48. https://doi.org/10.1145/2590773

Arenas, M., Barceló, P., Libkin, L., & Murlak, F. (2014). *Foundations of Data Exchange*. Cambridge University Press.

Arenas, M., & Libkin, L. (2008). XML data exchange. *Journal of the ACM*, *55*(2), 1–72. https://doi.org/10.1145/1346330.1346332

Arnaut, C., Pont, M., Scaria, E., Berghmans, A., & Leconte, S. (2018). Study on data sharing between companies in Europe: Final report : A study prepared for the European Commission DG Communications Networks, Content & Technology. Advance online publication. https://doi.org/10.2759/354943

Bastiaansen, H., Dalmolen, S., Kollenstart, M., & Punter, M. (2019). Infrastructural Sovereignty over Agreement and Transaction Data ('Metadata') in an Open Network-Model for Multilateral Sharing of Sensitive Data. In *Proceedings of the 40th International Conference on Information Systems,* Munich: Germany.

Bastiaansen, H., Dalmolen, S., Kollenstart, M., & van Engers, T. M. (2020). User-Centric Network-Model for Data Control with Interoperable Legal Data Sharing Artefacts. In *Proceedings of the 24th Pacific Asia Conference on Information Systems,* Dubai: UAE.

Becker, I., Posner, R., Islam, T., Ekblom, P., Borrion, H., McGuire, M., & Li, S. (2021). Privacy in Transport? Exploring Perceptions of Location Privacy Through User Segmentation. In *Proceedings of the 54th Hawaii International Conference on System Sciences,* Hawaii: USA.

Brathwaite, K. S. (1983). Resolution of conflicts in data ownership and sharing in a corporate environment. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *15*(1), 37–42. https://doi.org/10.1145/1113500.1113505

Cate, B. ten, & Kolaitis, P. G. (2010). Structural characterizations of schema-mapping languages. *Communications of the ACM*, *53*(1), 101–110. https://doi.org/10.1145/1629175.1629201

Catena-X. (2022). https://catena-x.net/en/

Chowdhury, M. J. M., Colman, A., Han, J., & Ashad, M. (2018). A Policy Framework for Subject-Driven Data Sharing. In *Proceedings of the 51st Hawaii International Conference on System Sciences,* Hawaii: USA.

Coelho, R., Braga, R., David, J. M., Dantas, M., Stroele, V., & Campos, F. (2021). Integrating Blockchain for Data Sharing and Collaboration Support in Scientific

Ecosystem Platform. In *Proceedings of the 54th Hawaii International Conference on System Sciences,* Hawaii: USA.

Compton, J. (2020). *Data Sharing In The Covid Era*. https://www.forbes.com/sites/tableau/2020/12/16/data-sharing-in-the-covid-era/?sh=7d50e1dd3db2

D'Hauwers, R., & Walravens, N. (2021). Do You Trust Me? Value and Governance in Data Sharing Business Models. *Proceedings of Sixth International Congress on Information and Communication Technology*, 217–225.

de Corbiere, F., & Rowe, F. (2010). Understanding the Diversity of Interconnections between IS: Towards a New Typology of IOS. In *Proceedings of the 18th European Conference on Information Systems,* Pretoria: South Africa.

de Corbière, F., & Rowe, F. (2010). Adoption Factors of Electronic Data Exchanges and Technologies to Improve Data Synchronization in BtoB Relationships: Are they similar? *DIGIT 2010 Proceedings*.

de Corbière, F., & Rowe, F. (2013). From Ideal Data Synchronization to Hybrid Forms of Interconnections: Architectures, Processes, and Data. *Journal of the Association for Information Systems*, *14*(10), 550–584. https://doi.org/10.17705/1jais.00345

Elgarah, W., Falaleeva, N., Saunders, C. C., Ilie, V., Shim, J. T., & Courtney, J. F. (2005). Data exchange in interorganizational relationships. *ACM SIGMIS Database: The DATABASE for Advances in Information Systems*, *36*(1), 8–29. https://doi.org/10.1145/1047070.1047073

Engelbrecht, A., Gerlach, J., & Widjaja, T. (2016). Understanding the anatomy of data-driven business models: Towards an empirical taxonomy. In *Proceedings 24th European Conference on Information Systems,* Istanbul: Turkey.

European Commission (Ed.). (2022). *Data Act: Commission proposes measures for a fair and innovative data economy*. https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113

Fagin, R., Kimelfeld, B., & Kolaitis, P. G. (2011). Probabilistic data exchange. *Journal of the ACM*, *58*(4), 1–55. https://doi.org/10.1145/1989727.1989729

Fagin, R., Kolaitis, P. G., Miller, R. J., & Popa, L. (2005). Data exchange: semantics and query answering. *Theoretical Computer Science*, *336*(1), 89–124. https://doi.org/10.1016/j.tcs.2004.10.033

Fisher, J. B., & Fortmann, L. (2010). Governing the data commons: Policy, practice, and the advancement of science. *Information & Management*, *47*(4), 237–245. https://doi.org/10.1016/j.im.2010.04.001

Foraker, R. E., Lai, A. M., Kannampallil, T. G., Woeltje, K. F., Trolard, A. M., & Payne, P. R. O. (2020). Transmission dynamics: Data sharing in the COVID-19 era. *Learning Health Systems*, 1-8. https://doi.org/10.1002/lrh2.10235

Fruhwirth, M., Pammer-Schindler, V., & Thalmann, S. (2019). To Sell or Not to Sell: Knowledge Risks in Data-Driven Business Models. *Proceedings of the 2019 Pre-ICIS SIGDSA Symposium*, *11*, 1–9.

Gartner Inc. (Ed.). (2021). *Data Sharing is a Business Necessity to Accelerate Digital Business*. https://www.gartner.com/smarterwithgartner/data-sharing-is-a-business-necessity-to-accelerate-digital-business

Gasco, M., Feng, W., & Gil-Garcia, J. R. (2018). Providing Public Value through Data Sharing: Understanding Critical Factors of Food Traceability for Local Farms and Institutional Buyers. In *Proceedings of the 51st Hawaii International Conference on System Sciences,* Hawaii: USA.

Gelhaar, J., Gürpinar, T., Henke, M., & Otto, B. (2021). Towards a taxonomy of incentive mechanisms for data sharing in data ecosystems. In *Proceedings of the 25th Pacific Asia Conference on Information Systems,* Dubai: UAE.

Gelhaar, J., & Otto, B. (2020). Challenges in the Emergence of Data Ecosystems. In *Proceedings of the 23rd Pacific Asia Conference on Information Systems,* Dubai: UAE.

Golshanara, L., & Chomicki, J. (2020). Temporal data exchange. *Information Systems*(87), 101414. https://doi.org/10.1016/j.is.2019.07.004

Hao, Y., Piao, C., Zhao, Y., & Jiang, X. (2020). Privacy Preserving Government Data Sharing Based on Hyperledger Blockchain. In K.-M. Chao, L. Jiang, O. K. Hussain, S.-P. Ma, & X. Fei (Eds.), *Advances in E-Business Engineering for Ubiquitous Computing* (pp. 373–388). Springer International Publishing.

Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2014). *Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms.* University of Cambridge, Cambridge.

Kalkman, S., Mostert, M., Gerlinger, C., van Delden, J. J. M., & van Thiel, G. J. M. W. (2019). Responsible data sharing in international health research: A systematic review of principles and norms. *BMC Medical Ethics*, *20*(21), 1–13. https://doi.org/10.1186/s12910-019-0359-9

Karger, E., Jagals, M., & Ahlemann, F. (2021). Blockchain for AI Data – State of the Art and Open Research. In *Proceedings of the 42nd International Conference on Information Systems,* Austin: USA.

Knol, A., Klievink, B., & Tan, Y.-H. (2014). Data Sharing Issues and Potential Solutions for Adoption of Information Infrastructures: Evidence from a Data Pipeline Project in the Global Supply Chain over Sea. *BLED 2014 Proceedings*.

Kolaitis, P. G. (2005). Schema mappings, data exchange, and metadata management. In G. Gottlob & F. Afrati (Eds.), *Proceedings of the twenty-fourth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems - PODS '05* (p. 61). ACM Press. https://doi.org/10.1145/1065167.1065176

Köster, A., Baumann, A., Krasnova, H., Avital, M., Lyytinen, K., & Rossi, M. (2020). PANEL 1: TO SHARE OR NOT TO SHARE: SHOULD IS RESEARCHERS SHARE OR HOARD THEIR PRECIOUS DATA? In *Proceedings of the 28th European Conference on Information Systems,* Marrakesh: Morocco.

Kraemer, J., Shekhar, S., & Hofmann, J. (2021). Regulating Algorithmic Learning in Digital Platform Ecosystems through Data Sharing and Data Siloing: Consequences for Innovation and Welfare. In *Proceedings of the 41st International Conference on Information Systems,* Austin: USA.

Kuerschner, C., Thiesse, F., & Fleisch, E. (2009). Factors Influencing Decentralized Strorage of RFID Tag Data: A Multi-Case Study. In *Proceedings of the 15th Amercias Conference on Information Systems,* San Francisco: USA. https://aisel.aisnet.org/amcis2009/144

Libkin, L. (2006). Data exchange and incomplete information. *Proceedings of the Twenty-Fifth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems, June 26-28, 2006, Chicago, Illinois, USA*, 60–69. https://doi.org/10.1145/1142351.1142360

Link, G., Lumbard, K., Germonprez, M., & Goggins, S. (2017). Advancing Open Online Community Science Practice: The Open Community Data Exchange. In *Proceedings of the 12th Midwest Association for Information Systems Conference,* Springfield: USA.

Liu, P., Ning, P., & Jajodia, S. (2001). Avoiding loss of fairness owing to failures in fair data exchange systems. *Decision Support Systems*, *31*(3), 337–350. https://doi.org/10.1016/S0167-9236(00)00141-X

Meyer, A., Pufahl, L., Batoulis, K., Fahland, D., & Weske, M. (2015). Automating data exchange in process choreographies. *Information Systems*, *53*, 296–329. https://doi.org/10.1016/j.is.2015.03.008

Mildebrath, H. (2021). *Data Governance Act: Briefing - EU Legislation in Progress*. https://www.europarl.europa.eu/RegData/etudes/BRIE/2021/690674/EPRS_BRI(2021)690674_EN.pdf

Mobility Data Space. (2022). https://mobility-dataspace.eu/de

Morar, D., & Kemper, H.-G. (2016). REQUIREMENTS OF INFORMATION SYSTEMS IN PRODUCT DEVELOPMENT AND PRODUCTION REGARDING ADDITIVE MANUFACTURING – A QUANTITATIVE EXPLORATION. In *Proceedings of the 24th European Conference on Information Systems,* Istanbul: Turkey.

Mosch, P., Schweikl, S., & Obermaier, R. (2020). Trapped in the supply chain: The digital Servitization of an industrial Supplier - out of the frying pan into the fire? In *Proceedings of the 28th European Conference on Information Systems,* Marrakesh: Morocco.

Nicolaou, A. I., Ibrahim, M., & van Heck, E. (2013). Information quality, trust, and risk perceptions in electronic data exchanges. *Decision Support Systems*, *54*(2), 986–996. https://doi.org/10.1016/j.dss.2012.10.024

Nokkala, T., Salmela, H., & Toivonen, J. (2019). Data Governance in Digital Platforms. *AMCIS 2019*, 1–10.

Omar, A., Bass, J. M., & Lowit, P. (2014). A Grounded Theory Of Open Government Data: A Case Study In The UK. *UK Academy for Information Systems Conference Proceedings 2014*, 1–22.

Parmar, R., Mackenzie, I., Cohn, D., & Gann, D. (2014). The New Patterns of Innovation. *Harvard Business Review*. https://hbr.org/2014/01/the-new-patterns-of-innovation

Parvinen, P., Pöyry, E., Gustafsson, R., Laitila, M., & Rossi, M. (2020). Advancing Data Monetization and the Creation of Data-based Business Models. *Communications of the Association for Information Systems*, *47*, 25–49. https://doi.org/10.17705/1CAIS.04702

Pujol Priego, L., & Wareham, J. (2020). The stickiness of scientific data. In *Proceedings of the 41st International Conference on Information Systems,* India.

Richter, H., & Slowinski, P. R. (2019). The Data Sharing Economy: On the Emergence of New Intermediaries. *IIC - International Review of Intellectual Property and Competition Law*, *50*, 4–29. https://doi.org/10.1007/s40319-018-00777-7

Samanthula, B. K., Elmehdwi, Y., Howser, G., & Madria, S. (2015). A secure data sharing and query processing framework via federation of cloud computing. *Information Systems*, *48*, 196–212. https://doi.org/10.1016/j.is.2013.08.004

Sarathy, R., & Muralidhar, K. (2006). Secure and useful data sharing. *Decision Support Systems*, *42*(1), 204–220. https://doi.org/10.1016/j.dss.2004.10.013

Smart Connected Supplier Network (SCSN). (2022). https://smart-connected.nl/en

Snowflake (Ed.). (2022). *What is Data Sharing?* https://www.snowflake.com/guides/what-data-sharing

Susha, I., Janssen, M., & Verhulst, S. (2017). Data Collaboratives as a New Frontier of Cross-Sector Partnerships in the Age of Open Data: Taxonomy Development. In *Proceedings of the 50th Hawaii International Conference on System Sciences,* Hawaii: USA.

van den Broek, T., & van Veenstra, A. F. (2015). Modes of Governance in Inter-Organizational Data Collaborations. In *Proceedings of the 23rd European Conference on Information Systems,* Münster: Germany.

Vesselkov, A., Hämmäinen, H., & Töyli, J. (2019). Design and Governance of mHealth Data Sharing. *Communications of the Association for Information Systems*, *45*(18), 299–321. https://doi.org/10.17705/1CAIS.04518

Webster, J., & Watson, R. T. (2002). Analyzing the Past to Prepare for the Future: Writing a Literature Review. *MIS Quarterly*, *26*(2), xiii–xxiii. http://dl.acm.org/citation.cfm?id=2017160.2017162

Zhong, J., Bertok, P., Mirchandani, V., & Tari, Z. (2011). Privacy-Aware Granular Data Access Control For Cross-Domain Data Sharing. In *Proceedings of the 15th Pacific Asia Conference on Information Systems,* Brisbane: Australia.