Electronic Markets (2024) 34:47
https://doi.org/10.1007/s12525-024-00732-0

RESEARCH PAPER

Data sharing practices: The interplay of data, organizational structures, and network dynamics

Marcel Fassnacht[1] · Jannis Leimstoll[1] · Carina Benz[1] · Daniel Heinz[1] · Gerhard Satzger[1]

Received: 22 March 2024 / Accepted: 2 September 2024

Abstract
With the proliferation of data and advanced analytics, organizations are increasingly recognizing the potential value of sharing data across organizational boundaries. However, there is a lack of empirical evidence and systematic frameworks to guide the design of effective data sharing practices. Realizing the full potential of data sharing requires the effective design and implementation of data sharing practices by considering the interplay of data, organizational structures, and network dynamics. This study presents an empirically and theoretically grounded taxonomy of data sharing practices drawing on existing literature and real-world data sharing cases. The subsequent cluster analysis identifies four generic archetypes of data sharing practices, differing in their primary orientation toward compliance, efficiency, revenue, or society. From a theoretical perspective, our work conceptualizes data sharing practices as a foundation for a more systematic and detailed exploration in future research. At the practitioner level, we enable organizations to strategically develop and scale data sharing practices to effectively leverage data as a strategic asset.

Keywords Data sharing · Data ecosystems · Taxonomy · Archetypes · Cluster analysis

JEL classification L8 · O3

Introduction

The global expansion of digitalization and connectivity has resulted in data becoming a fundamental part of business activities—reflected in the 'data economy,' which is expected to be worth €550 billion by 2025 (Cattaneo et al., 2020). As data collection continues to surge, organizations increasingly recognize and treat data as a pivotal strategic asset (Gelhaar et al., 2021a; Holstein et al., 2023; Schüritz et al., 2017). Private and public organizations increasingly seek to integrate external data sources through data sharing and leveraging their own internal data (Abbas et al., 2021; Gelhaar et al., 2021a). While not a fundamentally new concept, related concepts such as data ecosystems (Jussen et al., 2023) and data marketplaces (Lindner et al., 2021) are accelerating and driving organizational efforts to engage in data sharing (Cichy et al., 2021; S. Spiekermann et al., 2015). A prominent example is the 'Japan Data Space,' which aims to provide a cross-industry data space to share data for logistics status monitoring or developing services such as live maps of the public transportation net in Tokyo (Koshizuka, 2023).

In recent years, the possibility of making internal data assets accessible for external use has attracted increasing interest in the academic literature (Jussen et al., 2023; Schweihoff et al., 2023a). Despite this growing attention, the practical implementation of data sharing is still in its infancy, primarily focused on application in government or scientific contexts with only modest but rapidly emerging adoption in the private sector (Janssen et al., 2012; Krotova et al., 2020). This may be due to the strategic complexity associated with data sharing, which requires intensive planning and decision-making efforts before it can be economically viable (Bastiaansen et al., 2020; Fassnacht et al., 2023a; Jagals & Karger, 2021). To thoroughly analyze real-world settings, data sharing must be understood as a multidimensional concept considering the nature of the

Responsible Editor: Thorsten Schoormann

Marcel Fassnacht
marcel.fassnacht@kit.edu

[1] Institute of Information Systems and Marketing (IISM), Karlsruhe Institute of Technology (KIT), Kaiserstraße 12, Karlsruhe 76131, Germany

data itself as a shareable entity (e.g., domain or processing maturity), the organizational structure (e.g., motivation to share or revenue models), and the dynamics of the broader network that provides the overall context and setting for data sharing (e.g., interoperability or access coordination) (Dreller, 2018; Fassnacht et al., 2023b; Gelhaar & Otto, 2020). Consequently, organizations face challenges characterizing potential data sharing practices (Jussen et al., 2023) and difficulties in designing processes and artifacts for data sharing within their specific context.

Therefore, conceptualizing data sharing practices stands as a crucial focus in information systems (IS) research to understand how data sources can be effectively shared for value co-creation and innovation (Gelhaar et al., 2021b; Schüritz et al., 2019). This is particularly relevant in electronic markets, which today typically take the form of (de-)centralized digital platforms (Alt, 2020a), where effective data sharing can lead to the emergence of new digital business models that leverage emerging digital technologies such as the Internet of Things or AI (Alt, 2020b). As a result, many scholars (e.g., Abbas et al., 2021; Jussen et al., 2023; van de Ven et al., 2021) call for defining and conceptualizing the fundamental element in this context: the *data sharing practice*, which is described as the set of underlying motives, activities, and mechanisms, as well as the configurational setting for data sharing between organizations or individuals. While existing scholarly work examines specific areas such as data sharing business models (Schweihoff et al., 2023a; van de Ven et al., 2021), data governance (Lis & Otto, 2021), and incentive mechanisms (Gelhaar et al., 2021b), there is still no widespread agreement on the definitions, models, or theories of data sharing and data sharing practices. To address this issue, this work aims to bridge this gap by developing a taxonomy of data sharing practices and identifying archetypes. Our methodological approach is structured into two subsequent phases.

First, foundational research is imperative to acquire integral insights into potentially affected (inter-)organizational facets (i.e., dimensions) and corresponding design options (i.e., characteristics). Thereby, we focus on the dimensions and characteristics that are essential, or 'key,' to define and conceptualize data sharing practices (e.g., data domain, motivation, or reward) and therefore neglect subordinate, specific facets (e.g., industry-specificities), and design options (i.e., individual technological implementation of data exchange). Hence, the first research question states:

RQ 1: What are the key dimensions and characteristics of data sharing practices?

Second, examining and presenting archetypes of data sharing in real-world settings can offer critical insights into how organizations can effectively design and establish data sharing practices. Thereby, we can identify potential relationships and the interplay of data-, organizational-, and network-related dimensions and characteristics (Azkan et al., 2020; Schweihoff et al., 2023a). Following the call of Schweihoff et al. (2023a), examining archetypes of these practices in terms of incentives, benefits achieved, or data governance frameworks applied is likely to inspire organizations that have encountered challenges in utilizing data sharing for innovation. Thus, the second research question reads:

RQ 2: What are the archetypes of data sharing practices?

To address the outlined research questions, we followed a sequential research design consisting of two phases. First, we constructed a taxonomy of data sharing practices following the taxonomy development method proposed by Nickerson et al. (2013). In the second phase, we identified archetypes of data sharing practices using cluster analysis and interpretation (Kaufman & Rousseeuw, 1990).

Our work contributes to both research and practice. For research, we present a taxonomy and archetypes that extend the body of scientific knowledge and enable the establishment of a common, unified understanding for conceptualizing, analyzing, and designing data sharing practices and categorizing existing knowledge in this field. It then assists researchers in positioning their research within this emerging field and facilitates a more structured and nuanced exploration of data sharing practices. For practitioners, the taxonomy and archetypes provide initial guidance and a valuable tool for strategically designing data sharing practices, evaluating the resulting design options, and ultimately effectively using data as a strategic asset.

The paper is structured as follows: First, we provide an overview of related work, focusing on associated concepts and existing taxonomies and systematizations of data sharing. Next, we describe our methodological approach. Subsequently, we present a taxonomy of data sharing practices and the derived archetypes. In the following, we discuss our findings, outline scientific and managerial implications, and present limitations and future research opportunities. Finally, we summarize and conclude our work.

Background and related work

In this section, we address two foundational components essential to our research. First, we analyze the existing scientific literature on the conceptualization and exploitation of data sharing and associated concepts. Second, we examine existing related approaches to categorizing and systematizing data sharing practices.

Data sharing and associated concepts

While data sharing still lacks a generally accepted definition in the literature, this article refers to data sharing as 'the domain-independent process of giving third parties access to the data sets of others' (Jussen et al., 2023, p. 4). Subsequently, existing literature (e.g., Dreller (2018), Gelhaar et al. (2021b), or Jussen et al. (2023)) has characterized data sharing practices as the set of underlying motives, activities, and mechanisms, as well as the configurational setting for the data exchange between organizations or individuals. These data sharing practices include various aspects such as data management processes, technological infrastructure, motivational frameworks, reward models, and legal implications (Fassnacht et al., 2023b; Kitchin, 2014). To successfully establish data sharing, the engagement of actors is mainly dependent on the pursued and perceived benefits for each actor, which makes creating mutually beneficial solutions a decisive criterion (Enders et al., 2022). In this realm, Enders et al. (2022) investigate the benefits of data providers revealing data openly along three dimensions: *innovation driver*, *internal improvement*, and *visibility and participation*. The plethora of benefits extending beyond the narrow scope of financial compensation becomes evident in the work of Kawashita et al. (2022), who explore the benefits of open government data that outlines political and social, economic and financial, and operational and technical benefits for data providers. In contrast, barriers and challenges of data sharing are more extensively addressed in scientific literature, being described as the foundation for developing solutions to overcome these barriers towards successfully designing data sharing practices (Jussen et al., 2024b). Particularly Jussen et al. (2024b), Kajüter et al. (2022), and Fassnacht et al. (2023a) emphasize the multidimensionality of challenges for data sharing such as legal, financial, institutional, or technological challenges (Kajüter et al, 2022) and along aspects of motivation, ecosystem generation, design and operationalization, as well as intensification and scalability (Jussen et al., 2024b). This underpins the necessity of foundational work on data sharing design to extract value from data sharing by overcoming these challenges (Vesselkov et al. 2019). To date, research has not yet linked these challenges with potential design solutions due to the missing foundational work on the dimensions and characteristics that affect data sharing practices.

Further, the current body of literature often discusses data sharing practices only indirectly within the framework of associated concepts that include data sharing as a constituent activity (Dreller, 2018; Heinz et al., 2022; Jussen et al., 2024a; Susha et al., 2017). For instance, *data trading* and *data exchange* are often used synonymously for data sharing but significantly differ by definition (Jussen et al., 2023). *Data exchange* describes the technical transmission of data. It thus neglects consideration of legal aspects, cultural elements, and (inter-)organizational practices of sharing data, reflecting an essential but more narrow sub-concept of data sharing (Awada and Kiringa, 2015; Jussen et al., 2023). Similarly, data trading focuses on sharing data for commercial purposes, thus limiting data sharing practices toward sole practices with financial interests, reflecting another essential sub-concept of data sharing (Liang et al., 2018). *Open data*, in contrast, entails the non-commercial sharing of data accessible to all for free use and redistribution (Enders et al., 2020). *Data ecosystems* are characterized by the formation of a multilateral set of actors around a shared value proposition, with data serving as the primary resource and data sharing constituting a fundamental practice within the ecosystem (Heinz et al., 2022; Oliveira et al., 2019). Emerging to provide a secure and reliable infrastructure, *data spaces* are defined as 'a federated, open infrastructure for sovereign data sharing, based on common policies, rules, and standards' (Reiberg et al., 2022, p. 11) with data sharing as an inherent practice. Data spaces are characterized by an open infrastructure that is freely accessible, which is not required for data ecosystems (Gieß et al., 2023; Otto & Jarke, 2019). *Data marketplaces* represent a specialized setting for data sharing practices with commercial objectives, defined as 'third-party platforms acting as neutral intermediaries and allowing others to sell standardized data products' (Sterk et al., 2022, p. 3). Data marketplaces aim to provide a unified platform for domain-specific commercial data sharing between independent data providers and consumers (Jussen et al., 2024a; Sterk et al., 2022; van de Ven et al., 2021). The concept of *data trusts* is described as 'trusted intermediaries that enable data sharing through a confident and sovereign infrastructure and standardized processes' (Lauf et al., 2023, p. 2), emphasizing the role of a neutral intermediary that aims to enable trust in data sharing practices (Jussen et al. 2024a; Lauf et al., 2023). In the realm of addressing societal challenges, Susha (2017) introduces the concept of *data collaboratives*, describing the sharing of data across organizations in specific partnerships to address societal challenges.

The emergence of these promising concepts, summarized in Table 1, which all revolve around data sharing as their constituent activity, further reinforces the need for a fundamental understanding and conceptualization of the underlying concept of data sharing practices.

Taxonomies and classification of data sharing

Classification approaches, such as taxonomies, serve as crucial instruments for both researchers and practitioners, helping to understand, analyze, and organize knowledge within emerging research domains by identifying shared characteristics within a coherent conceptual framework (Hunke et al., 2021; Nickerson et al., 2013). However, existing efforts to

Table 1 Overview of data sharing and related concepts

Concept	Definition	Contributing literature
Data Sharing	Data sharing is the domain-independent process of giving third parties access to the data sets of others.	Dreller, 2018; Jussen et al., 2023; Richter & Slowinski, 2019; Schweihoff et al., 2023a
Data Trading	Data trading is a commercial form of sharing data seeking to gain financial profits from data sharing.	Liang et al., 2018; Muschalle et al., 2012; Spiekermann et al., 2015
Open Data	Open data describes the non-commercial sharing of data, which is available to anyone for use and redistribution free of any charge.	Enders et al., 2020; Susha et al., 2017; Wilms et al., 2018
Data Ecosystems	Data ecosystems are the formation of a multilateral set of actors around a shared value proposition, with data serving as the primary resource and data sharing constituting a fundamental practice within the ecosystem.	Azkan et al., 2020; Gelhaar et al., 2021a; Heinz et al., 2022; Lis & Otto, 2021; Oliveira et al., 2019
Data Spaces	A data space is defined as an open and decentralized infrastructure for sovereign data sharing, which incorporates common standards, rules, and policies.	Gieß et al., 2023; Otto & Jarke, 2019; Reiberg et al., 2022
Data Marketplaces	A data marketplace is a third-party platform acting as neutral intermediary and allowing others to sell standardized data products for commercial purposes.	Abbas et al., 2021; Spiekermann, 2019; Sterk et al., 2022; van de Ven et al., 2021
Data Trusts	A data trust is a form of an intermediary that enables trusted data sharing through a sovereign and confident infrastructure and standardized processes	Arlinghaus et al., 2021; Czech et al., 2023; Lauf et al., 2023; Schweihoff et al., 2023b
Data Collaborative	Data collaboratives represent a type of partnership spanning across sectors and involving both public and private organizations to share data for the purpose of tackling societal issues.	Susha et al., 2017; Susha & Gil-Garcia, 2019

systematize data sharing and related concepts tend to focus on specific facets of data sharing, neglecting the integral structuring of data sharing practices. For instance, Schweihoff et al. (2023a) focus on delineating design options and criteria for data sharing business models. Other contributions in this domain include the work of Hartmann et al. (2016) and van de Ven et al. (2021), who develop taxonomies to systematically examine data-driven business models used by startups and business models tailored for data marketplaces. In the realm of data ecosystems, prevailing classification approaches emphasize understanding the concept and incentive mechanisms for data sharing within data ecosystems (Azkan et al., 2020; Gelhaar et al., 2021a, 2021b). Moreover, several studies elaborate on the design of associated concepts of data sharing, such as Gieß et al. (2023) presenting design options for data spaces or Susha et al. (2017) exploring data collaboratives as a variant of data sharing prevalent in public-private collaborations.

In terms of identifying archetypes related to data sharing practices, two contributions are noteworthy. First, Scheider et al. (2023) focus on deriving design options for data marketplaces for trading personal data and derive archetypes to characterize these specific data markets. However, this reflects a focus on one particular aspect of data sharing and thus neglects the applicability to the broader context of data sharing practices in general, which may not involve personal data. Second, Lauf et al. (2023) develop a taxonomy and derive archetypes for data trustees in healthcare, thereby exploring the design characteristics for a specific stakeholder within a data sharing practice (the intermediary), and thus also lacking generalizability to the underlying data sharing practice.

In summary, our review of existing studies highlights the lack of a concise conceptualization of data sharing practices. Despite the steady expansion of the data sharing literature, the term *data sharing* remains an elusive concept with a lack of widely accepted definitions, models, and theories (Jussen et al., 2023; Lindner et al., 2021; Richter & Slowinski, 2019). Given the central role of data sharing within the concepts mentioned above, it is imperative to conceptualize data sharing practices to understand and characterize the forms of application to streamline research and guide practice in future data sharing endeavors. Furthermore, existing research is predominantly grounded in specific forms or concepts of data sharing with a particular purpose, leaving a gap in comprehensive coverage. To our knowledge, there are currently no taxonomies or archetypes that address data sharing practices. We contend that the burgeoning field of data sharing research would benefit from a more general systematization of knowledge about data sharing. This will foster a shared understanding of data sharing, facilitate the materialization of data sharing ideas and considerations, and provide initial guidance and a valuable tool for systematically designing data sharing practices in organizations.

Research design

Our research design consists of two sequential phases, as illustrated in Fig. 1. In the first phase, we addressed the first research question and developed and evaluated a taxonomy that includes the key dimensions and characteristics of data sharing practices. In doing so, we followed the taxonomy development methodology of Nickerson et al. (2013), complemented by the suggested evaluation criteria of Kundisch et al. (2022). In the second phase, we drew upon the results of the first phase and addressed the second research question. Thereby, we extended the case database and verified the coding scheme, conducted a cluster analysis (Kaufman & Rousseeuw, 1990), performed a cross-table analysis, interpreted the resulting clusters, and calculated the silhouette coefficient (Rousseeuw, 1987) to identify and evaluate archetypes of data sharing practices.

Phase 1: taxonomy development

Iterative taxonomy development process

Following the approach of Nickerson et al. (2013), we first defined the primary objective of our taxonomy, the meta-characteristic. Our taxonomy aims to improve the ability to conceptualize and design data sharing practices and provide practical guidance for structurally characterizing data sharing practices. Accordingly, we defined 'key dimensions and characteristics of data sharing practices' as the meta-characteristic, which guides our evaluation of potential dimensions and characteristics throughout the iterative development process. Given our goal of developing a taxonomy of data sharing practices to support the conceptualization and design of such practices, we refrained from focusing on an organization's particular role in data sharing (e.g., data provider, data consumer, or intermediary as essential roles proposed by Oliveira et al. (2019)) but aimed to develop a taxonomy that ensures applicability from any role perspective. Independent from the role perspective, the taxonomy's applicability is pursued in the design and decision phase of data sharing practices, providing mainly two target groups or users of the taxonomy (Kundisch et al., 2022) with a valuable framework. Researchers in the field of data sharing can gain a foundational understanding of data sharing practices and build and contextualize upon the findings. In practice, data sharing project leaders and decision-makers (e.g., chief information officers, chief data officers, data stewards, and data sharing specialists) are provided with a valuable tool for actionable guidance on structurally configuring data sharing practices, comparing cases, and deriving suitable design options to operationalize strategic approaches to data sharing, applicable at both the ideation and implementation stages. In practice, the users of the taxonomy must have either individually or collectively a cross-functional (business, data, information technology, and legal) understanding of data sharing and corresponding decision rights and authority on implementing a data sharing project to successfully design and establish data sharing practices. Further, the taxonomy aims to provide a compact conceptualization of data sharing practices, capable of classifying various forms and manifestations of such practices (Glass & Vessey, 1995).

A review of the existing literature revealed the following overarching meta-dimensions: the data itself as a shareable entity, the organizations involved (e.g., data providers, consumers, intermediaries), and the encompassing network in which these organizational actors interact (Arenas et al., 2019; Dreller, 2018; Gascó et al., 2018; Gelhaar & Otto, 2020). The first meta-dimension, *data*, could be retrieved from multiple dimensions and meta-dimensions characterizing the data as the shareable entity such as from the service domain (*timeframe*) or the technology domain (*data*

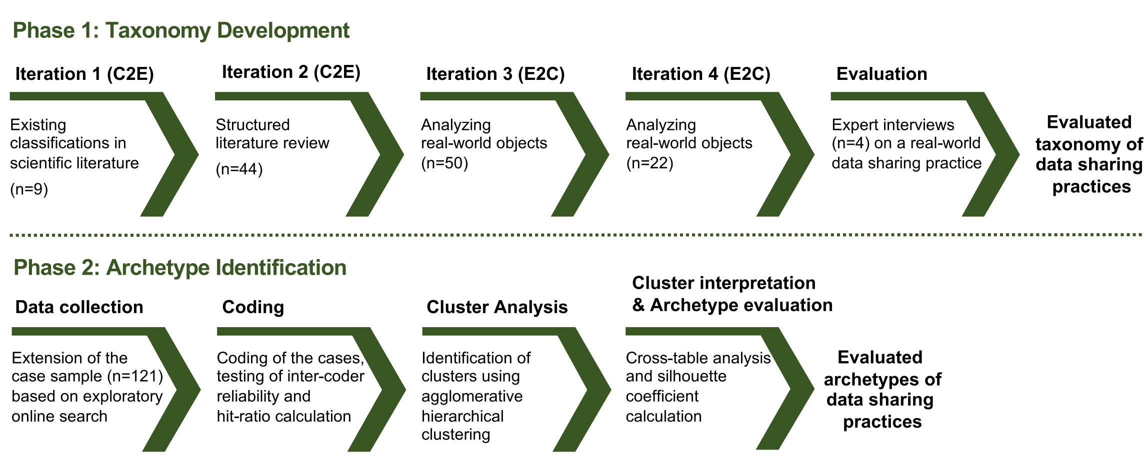

Fig. 1 Research design of the two successive phases

source) of van de Ven et al. (2021), Gelhaar, Gürpinar et al. (2021), and Azkan et al. (2020) outlining data (assets) as a relevant meta-dimension to characterize the resources (data objects and assets) as a necessary entity of data sharing. The second meta-dimension *organization* could be derived from the micro-perspective of one actor within the data sharing practice, such as Gelhaar, Gürpinar et al. (2023) and van de Ven (2021), seeing this as the (organizational) foundational structures for engaging in a data sharing practice, including the financial perspective on value capturing, and Schweihoff et al. (2023a) designing their entire taxonomy to be applied from a micro-perspective. In contrast, the meta-dimension *network* reflects the macro-perspective of data sharing practices, such as Schäffer and Stelzer (2017) and Lis and Otto (2021) refer to the dynamics and relationship between organizations, coordination structures and mechanisms, or interdependencies as well as the technical medium for sharing data between organizations. Thus, we referred to these three elements—data, organization, and network—as meta-dimensions to structure our taxonomy. These meta-dimensions served as a conceptual lens for organizing both conceptual and empirical characteristics.

We defined ending conditions against which the fulfillment of the taxonomy was measured at each iteration of the development process by adopting the eight objective (E1 to E8) and five subjective (E9 to E13) ending conditions described by Nickerson et al. (2013) and Kundisch et al. (2022). The definitions of the eight objective and five subjective ending conditions are shown in Fig. 2. In terms of taxonomy design, the taxonomy development process allows for two different design approaches: conceptual-to-empirical (C2E) to extract dimensions and characteristics from existing literature and empirical-to-conceptual (E2C) to derive them from empirical objects (Nickerson et al., 2013). These approaches were performed iteratively, starting with the C2E approach and verifying the findings with the E2C approach, until all ending conditions were satisfied, as recommended when the existing knowledge base contains relevant insights about the phenomenon under consideration (Kundisch et al. 2022). The evolution of the dimensions during the iterations is shown in Fig. 2.

Iteration 1 (C2E)

First, we extracted dimensions and characteristics from existing classifications in scientific literature. Eight different taxonomies (Gelhaar et al., 2021a, 2021b; Hartmann et al., 2016; Lis & Otto, 2021; Schäffer & Stelzer, 2017; Schweihoff et al., 2023a; Susha et al., 2017; van de Ven et al., 2021) and one morphology (Azkan et al., 2020) were examined, which cover specific aspects (e.g., business models, incentive mechanisms) or focus on associated concepts (e.g., data ecosystems, data marketplaces) of data sharing. An open coding approach was used to derive relevant dimensions and corresponding characteristics. Two independent researchers coded and assessed the relevance of these characteristics to the objectives of the work. While these classifications provided a solid foundation, no objects have yet been classified under any of the characteristics (E1, E3), and all dimensions and characteristics were newly created (E4).

Iteration 2 (C2E)

In the second iteration, we again used the C2E approach to refine the taxonomy based on the scientific literature. We conducted a systematic literature review following the methodology proposed by Webster and Watson (2002). We applied the search string ‘data sharing’ OR ‘data exchange’ OR ‘data trading’ OR ‘shared data’ across the Senior Scholars’ Basket of Eight[1] and the AIS eLibrary. For the search string, we chose the term ‘data sharing’ as the broad concept under investigation, and we included the terms ‘data exchange’ as well as ‘data trading’ as they reflect essential sub-concepts of data sharing, mainly focusing on the technical transmission of data or the commercial purpose of data sharing. Initially, 223 relevant articles were identified and cleaned by removing six duplicates. Two independent researchers reviewed the sample in a two-step approach: first, title and abstract screening resulted in 46 relevant articles, followed by a full-text screening of 28 out of the 46 papers, which resulted in removing 17 papers from the sample, which ultimately led to 29 articles considered relevant. Thereby, we applied the following inclusion and exclusion criteria: (1) contains at least one dimension or characteristic relevant to our study, (2) no pure technical focus on the data transfer, (3) written in English. We referred in exclusion criteria (2) solely on excluding papers that mainly investigate a specific technology and its application for technical data exchange (i.e., data protocols or database configuration) or specific mechanisms for data security and privacy (i.e., anonymization and pseudonymization mechanisms). We aim to analyze data sharing practices focusing on the interplay of the meta-dimensions (data, organizational structures, and network dynamics) to inform the conceptualization and design of data sharing practices rather than their technical implementation. Thus, we neglected to include papers on the specific technological mechanisms and configurations of technical architectures to be consistent regarding the

[1] The literature search was performed before the AIS Senior Scholars’ List of Premier Journals was updated. The considered journals are European Journal of Information Systems, Information Systems Journal, Information Systems Research, Journal of the Association for Information Systems, Journal of Information Technology Journal of Management Information Systems, Journal of Strategic Information Systems & MIS Quarterly.

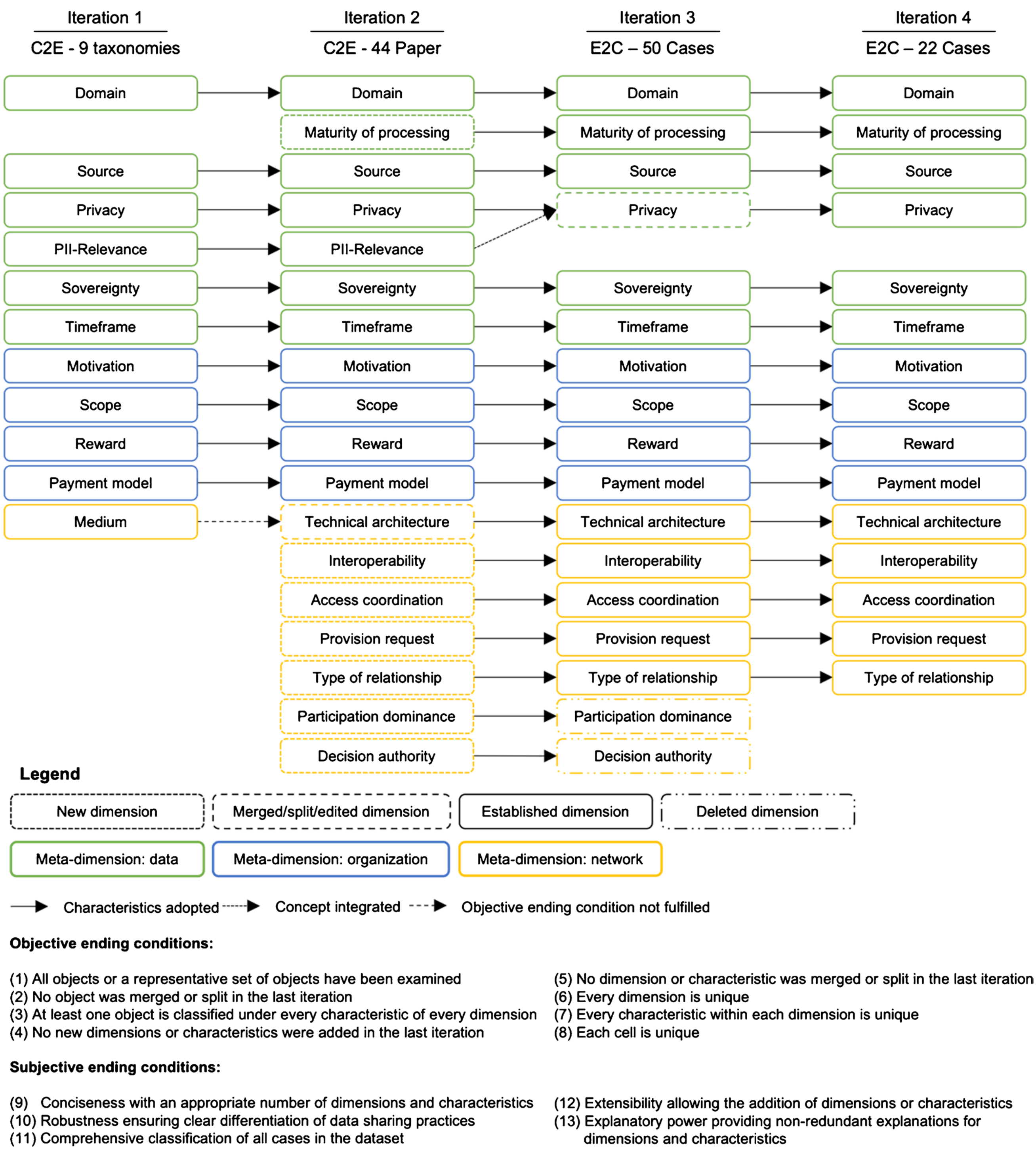

Fig. 2 Overview of the iterative taxonomy development process

level of detail and the purpose of the taxonomy. Instead, we only included papers with technical aspects that refer to the conceptual characterization and design of data sharing practices from a technical perspective (such as overall technical architecture, technical interoperability, or access coordination). Further, 15 additional articles were obtained through forward and backward searches, resulting in a final sample of 44 articles relevant to our study. To extract relevant dimensions and characteristics, we used an iterative coding approach with two coding cycles (Saldaña, 2015). First, we applied open coding to identify characteristics of data sharing practices. Then, we applied axial coding to aggregate these characteristics into dimensions and aligned them with the pre-defined meta-dimensions. The coding was conducted independently by two researchers to ensure consistency. Table 2 provides examples of the open and axial coding procedures. The resulting coding was aligned, and discrepancies were mutually resolved through discussion among the authors. This iteration added new dimensions (e.g., *maturity of processing*, *interoperability*) and edited one dimension (*scope*) regarding its characteristics. So far, no empirical objects have been studied, thus violated ending conditions E1, E3, and E4, which required a third iteration.

Iteration 3 (E2C)

In this iteration, we applied the E2C approach, focusing on empirical evidence from publicly available data sharing practices. To ensure suitability with our study, we established three criteria for case selection: (1) written in English, (2) describing a data sharing practice, and (3) for each meta-dimension, at least one characteristic can be derived and assigned from the case description. We compiled 72 real cases from a variety of reliable, publicly available, secondary sources, including business reports of organizations or initiatives (e.g., the Open Data Institute), practice-oriented journals and magazines (e.g., Harvard Business Review), corporate press releases by organizations, or blog posts. In total, we collected 43 data sources that comprised 72 cases relevant to our study. These cases spanned a variety of industries, including agriculture, oil and gas, transportation, automotive, or finance and insurance, and range from large organizations to small start-ups. The sample included data sharing cases where organizations made data publicly available ('open data'), reciprocal systems where actors in an ecosystem shared data, organizations traded data through data marketplaces, and bilateral collaborations in which organizations shared data for specific purposes (e.g.,, reducing carbon emissions, achieving process improvements, implementing new data-driven services). We analyzed 50 randomly selected cases to identify novel dimensions and characteristics and to empirically validate results from previous iterations. In this phase, we merged two dimensions into one (*privacy* and *personally identifiable information* (*PII*)-*relevance*) because organizations typically classify sensitive data as PII-relevant. Since not all the objects have been examined and some dimensions have been merged, some final conditions were still not met (E1, E5).

Iteration 4 (E2C)

Following the E2C approach, we analyzed the remaining 22 cases and consequently eliminated two dimensions: *decision authority* and *participation dominance*. For *decision authority*, complexity arose from intricate decision-making processes involving multiple stakeholders (e.g., legal counsel, or data owner, steward, and management). Similarly, *participation dominance* was not clearly assignable because participation in data sharing practices is subject to negotiation. Finally, all objective ending conditions were met, with subjective conditions having been discussed among the author team. We considered the taxonomy to be concise by having broken down the enormous complexity into 15 distinct dimensions (E9). Robustness

Table 2 Coding examples and aggregation to the corresponding dimensions

Coding examples (open coding)	Dimension (axial coding)
'Actors in a data ecosystem have various motivations and interests for the willingness to share their data.' (Gelhaar et al., 2021b, p. 3)	Motivation
'The national funding bodies are also enforcing policy for sharing datasets by making data sharing a requirement by default.' (Sayogo & Pardo, 2011, p. 4)	
'previous research assumed […] a benefit for the community rather than personal benefits in data exchange.' (Wilms et al., 2018, p.7)	
'[…] access can be granted to others free of charge or in exchange for compensation (e.g., monetary or reciprocal data).' (Schweihoff et al. 2023a, p. 2)	Reward
'Besides a direct payment as a reward, there are also incentive mechanisms that use bartering [of data or services].' (Gelhaar et al. 2021b, p. 8)	
'[…] government agencies, like the UNO or the World Bank, which provide statistics free of charge.' (Muschalle et al., 2012, p.5)	
'An approach [..] is the new European Interoperability Framework […]. It distinguishes four interoperability levels that must be implemented.' (Bastiaansen et al., 2020, p. 3).	Interoperability
'we must consider aspects such as interoperability, privacy, and trust in shared data.' (Coelho et al., 2021, p. 1)	
'[…] the private sector can provide various degrees of access to their data – ranging from making available select insights from data to select users to making data available to anyone by publishing it as open data.' (Susha et al., 2017, p. 6)	Access coordination
'Many interviewees discussed data access rights.' (Nokkala et al. 2019, p. 6)	

(E10) and comprehensiveness (E11) were confirmed by the applicability to the 22 objects in this iteration. The independence of dimensions and characteristics ensured extensibility (E12). Non-redundant explanations were provided to characterize data sharing practices, satisfying the explanatory (E13) ending condition. With all ending conditions having been met, the taxonomy development process was terminated.

Taxonomy evaluation

To evaluate our taxonomy, we performed an in-depth analysis of an appropriate real-world data sharing case to demonstrate our taxonomy, as Szopinski et al. (2019) propose demonstrations as a suitable and frequently used evaluation approach. For this, we conducted four semi-structured interviews with experts involved in one specific data sharing practice in the chemical industry that is not part of the previous case sample. For the expert sample, we followed a purposeful sampling approach according to Palinkas et al. (2015) based on the following criteria: (a) the experts must have in-depth knowledge of data sharing in general, and (b) the experts must be directly involved in the case under study. The experts' roles within the organization that served as the data provider were (I1) global data steward, (I2) head of data stewardship, (I3) exposure modeling specialist, and (I4) global project leader. The interviews lasted between 43 and 57 min and were divided into three parts. First, the taxonomy was presented and described. Second, we asked the experts to apply our taxonomy to characterize the case under study. Third, we asked the experts to evaluate the taxonomy regarding the criteria conciseness, robustness, comprehensiveness, extensibility, and explanatory (Nickerson et al., 2013) to verify the subjective ending conditions. Additionally, we evaluated the taxonomy regarding its applicability and usefulness, as suggested by Szopinski et al. (2020), to ensure that the designed artifact can support stakeholders in fulfilling their purpose of utilizing the artifact. For this, we observed the application of the taxonomy by the experts on a real-world case followed by asking questions on the applicability and usefulness. All interviews were recorded, transcribed, and coded with open coding independently by two researchers using MAXQDA software to extract the case characterization with our taxonomy and corresponding explanations. A second coding cycle was conducted by one researcher following a deductive coding approach aiming to extract the evaluation of the subjective ending conditions. The results are depicted in the results section (Fig. 4) and elaborated on in detail in Fassnacht et al. (2024). The evaluation of the subjective ending conditions and the taxonomy's applicability and usefulness is outlined in the results section after presenting the taxonomy.

Phase 2: archetype identification

Data collection and preparation

We expanded our initial sample of 72 cases for the second phase of the study to enable a robust and meaningful identification of archetypes based on a sufficient number of cases. Thus, we collected additional cases from different data sources. First, we searched for additional business reports that focus on empirical practices in the field of data sharing (e.g., Lindner et al. (2021)), from which we retrieved 21 cases. Second, we conducted an online search for trustful archival data on data sharing practices involving public and/or private sector organizations, which added 28 cases to our sample. In total, we expanded our sample to 121 data sharing practices (listed in Table 6 in the Appendix), which served as the case base for the subsequent cluster analysis.

Case coding of data sharing practices

In the subsequent step, we applied provisional coding (Saldaña, 2015) to code each case in the sample using the developed taxonomy as a codebook. A single author conducted the coding process and assessed each case according to the dimensions outlined in the taxonomy. Subsequently, a random subset of 20% (n = 24) cases of the case base was subjected to a second independent coding by a second researcher to assess inter-coder reliability, as recommended by O'Connor and Joffe (2020). The comparison of the two independent coding results yielded a percentage agreement of 89.76% and a Cohen's Kappa of 0.68 (Cohen, 1960) as inter-coder reliability. This demonstrated 'substantial agreement' between the coders (Landis & Koch, 1977). Consequently, the initial coding was considered a reliable ground truth for further analysis. To further validate the robustness of our dataset, ten IS scholars independently coded a new random sample (n = 10) extracted from the case database. By comparing their results with the ground truth (i.e., the coding results from the initial coding), we calculated the hit ratio for each taxonomy dimension. Table 3 presents a summary of these results, with higher values indicating greater agreement and lower values indicating disagreement. The resulting values ranged from 71.28 to 93.25%, with an average of 78.41%, indicating a high level of agreement and supporting the validity of the coding (cf. Jonas et al., 2022; Moore & Benbasat, 1991).

Cluster analysis

Next, we performed a cluster analysis to identify distinct archetypes of data sharing practices. Cluster analysis involves statistical methods for categorizing objects according to their similarity, seeking to minimize the distance

Table 3 Hit ratio for each taxonomy dimension

Meta-dimension: data		Meta-dimension: organization		Meta-dimension: network	
Dimension	Hit ratio	Dimension	Hit ratio	Dimension	Hit ratio
Domain	81.40%	Motivation	93.25%	Technical architecture	75.63%
Maturity of processing	71.87%	Scope	85.00%	Interoperability	74.25%
Source	81.06%	Reward	80.36%	Access coordination	71.28%
Privacy	72.67%	Payment model	83.32%	Provision request	72.69%
Sovereignty	81.44%			Type of relationship	75.00%
Timeframe	76.92%				
Average overall	**78.41%**				

Table 4 Agglomerative coefficients for different clustering methods

Clustering method	Agglomerative coefficient
Ward	0.93
Complete	0.79
McQuitty	0.73
Average	0.70
Single	0.67

between objects within the same cluster while maximizing the divergence between clusters (Han et al., 2012; Michener & Sokal, 1957). It is particularly useful for inductive, open-ended research that aims to describe the objects under study along a wide range of variables and to identify thematic patterns (Miller, 2018). This technique allowed us to cluster cases according to their similarities and differences in characterizations along the taxonomy. The cluster analysis was performed using the statistical computing language *R*.

To assess the distance between cases, we converted the coding into a dichotomous table. Each row represented a case from the dataset, and each column corresponded to a characteristic of the taxonomy. We assigned a value of 1 to objects where the characteristic is present and a value of 0 to objects where the characteristic is absent. For each pair of cases, we computed the Manhattan distance (i.e., the L^1 distance), a distance metric well-suited to the dichotomous structure of the dataset. This distance metric quantified the distance between two observations as the number of characteristics on which they differ.

To determine the most appropriate clustering algorithm, we compared different approaches of agglomerative hierarchical clustering (Backhaus et al., 2006). We evaluated the agglomerative coefficient (Rousseeuw, 1986) for single, complete, average, McQuitte, and Ward clustering methods. This coefficient, ranging from 0 to 1, indicates the balance and robustness of a clustering structure (Kaufman & Rousseeuw, 1990). Table 4 displays the resulting agglomerative coefficients for the compared methods. Ward's method (Ward, 1963) yielded the highest agglomerative coefficient (c = 0.93), indicating superior balance and robustness in the clustering results. This method is commonly used in cluster analyses to identify archetypes (Hunke et al., 2021; Möller et al., 2019) and is particularly effective with the Manhattan distance measure (Strauss & von Maltitz, 2017).

To identify a sufficient number of clusters, we chose hierarchical clustering for its ability to accommodate a random number of clusters without prior specification. This iterative process initially assigned each case to its own cluster, which were then iteratively merged, based on proximity, until all cases were consolidated into a single cluster. Consequently, the algorithm generated potential clustering solutions for different numbers of clusters that were feasible for the dataset under consideration. The dendrogram (Fig. 6 in the Appendix) illustrates the potential cluster configurations and the merging process. The dendrogram helped to visually examine the clustering results and provided a guideline for selecting an optimal number of clusters. Identifying the most appropriate number of clusters is a complex task addressed in the literature, offering various mathematical and empirical methods, each with advantages and disadvantages (Hardy, 1996; Milligan & Cooper, 1985). To ensure the logical validity and practical relevance of the clustering structure, an interpretive approach was adopted, following recent recommendations to focus on clusters as real-world phenomena and emphasizing a comprehensive and consistent cluster characterization (Mirkin, 2011). The qualitative evaluation was performed by drawing a vertical line starting at the right side of the dendrogram and gradually moving this line to the left. The number of branches this cutoff line crosses in the dendrogram corresponded to the number of clusters resulting from the hierarchical clustering at that level. Several iterations comparing the number of resulting clusters in terms of possible interpretations and practical significance led to a consensus on four distinct clusters. The final set of clusters balances manageability and cluster homogeneity (Milligan & Cooper, 1985) while remaining consistent with Hambrick's (1984) sample size-based suggestion.

Cluster interpretation and evaluation of archetypes

After determining an appropriate cluster solution, the clusters were evaluated using both qualitative and quantitative measures. Qualitatively, the case descriptions were reviewed to identify the primary and distinctive characteristics of each cluster. The cross-table analysis, which was performed to calculate the frequency distribution of each characteristic per cluster (see Fig. 5), supported this evaluation step (Hambrick, 1984). Consequently, the following archetypes were derived from the clusters: (I) compliance-oriented, (II) efficiency-oriented, (III) revenue-oriented, and (IV) society-oriented data sharing practices.

To quantitatively evaluate the cluster solution, we calculated the silhouette widths for each cluster, which indicate cluster validity regarding their structural strength on a range from − 1 to 1, with higher values indicating more substantial and robust results (Rousseeuw, 1987). All four clusters yielded an average silhouette value above the recommended threshold of 0.25 (Kaufman & Rousseeuw, 1990). The analysis revealed an existing but weak clustering structure, with an average silhouette value of s = 0.28 among the clusters. The values ranged from $s_I = 0.27$ (Cluster I) to $s_{III} = 0.31$ (Cluster III). Clusters II and IV had a silhouette value of $s_{II} = s_{IV} = 0.28$. These results were consistent with common findings in social science data such as this dataset, where natural groups are rarely strongly represented (Hambrick, 1984). Figure 7 in the Appendix illustrates the silhouette values for the cases within each cluster. This approach of cluster interpretation and evaluation of archetypes was also applied by Sterk et al. (2024) in their recently published paper in Electronic Markets.

Results

In the following, we present a scientifically and empirically grounded taxonomy of data sharing practices and four corresponding archetypes of data sharing practices as our results. The taxonomy results from previously conducted research to systematize data sharing practices, which introduces a taxonomy identifying key dimensions and characteristics of data sharing practices and is communicated in detail in a conference paper (Fassnacht et al., 2024). Therefore, this results section focuses on the elaboration of the archetypes and provides a consolidated presentation of the taxonomy.

Taxonomy of data sharing practices

The taxonomy of data sharing practices comprises three meta-dimensions (data, organization, network) encompassing 15 dimensions and their respective characteristics (Fig. 3). The last column denotes whether a dimension is mutually exclusive (E), i.e., accommodating a single

	Dimension	Characteristics						E/N
Data	Domain	**Process data (52)**	Product data (23)	Environment data (24)	Customer data (20)	Meta-data (8)		N
Data	Maturity of processing	**Raw data (36)**	Modified data (16)	Enriched data (2)	Processed data (31)			N
Data	Source	**Generated (51)**	Acquired (17)	Customer-provided (21)	Free available (4)			N
Data	Privacy	Public (3)	**Internal (37)**	Confidential (11)	Sensitive (21)			E
Data	Sovereignty	Individual (10)	**Organizational (58)**	Shared (4)				E
Data	Timeframe	Static (4)	**Up-to-Date (44)**	(Near-) real-time (27)				N
Organization	Motivation	**Economic (51)**	Social & Environmental (42)	Legal (5)	Cultural (13)			N
Organization	Scope	Intra-organizational (2)	**Same industry (56)**	Cross-industry (14)				E
Organization	Reward	Financial (21)	Virtual assets (2)	Data (13)	**Service (61)**	Reputation (22)		N
Organization	Payment model	**Free (59)**	Fixed (3)	Subscription-based (3)	Usage-based (2)	Revenue-based (0)	Hybrid (5)	E
Network	Technical architecture	Closed API (21)	**Platform (48)**	Open API (3)				E
Network	Interoperability	Legal (24)	Organizational (48)	Semantic (10)	**Technical (58)**			N
Network	Access coordination	**Agreement-based (58)**	Application-based (3)	Trust-based (2)	Open (9)			E
Network	Provision request	**On demand (43)**	Event-based (7)	Continuous (22)				E
Network	Type of relationship	One-to-one (12)	One-to-many (6)	Many-to-one (15)	**Many-to-many (39)**			E

Fig. 3 Taxonomy of data sharing practices

applicable characteristic, or non-exclusive (N), i.e., allowing for multiple characteristics to apply to a data sharing practice. The number in brackets behind each characteristic indicates the number of cases from the sample ($n = 72$) to which the characteristic applies.

Meta-dimension: Data

Domain: A data domain describes categories of data with shared characteristics or contexts (Azkan et al., 2020). *Process data* pertains to any data associated with organizational processes, including, e.g., production or research and development processes (Azkan et al., 2020). *Product data* encompasses data collected by smart products or sensors and is often referred to as big data (Azkan et al., 2020; Cichy et al., 2021). *Environment data* covers natural environment-related data, like geographic and climate data (Sayogo & Pardo, 2011). *Customer data* includes data specifically related to organizational customers (Ackermann et al., 2022; Lawrenz & Rausch, 2021; Mollick, 2016). *Meta-data* refers to descriptive data such as structure, format, content, and context of actual data (Azkan et al., 2020).

Maturity of processing: The maturity of processing describes the progression of data through various stages of modification, enrichment, and refinement (Nokkala et al., 2019). *Raw data* represents data in its original form, directly captured from sources like sensors or mobile devices, without any alterations (Muschalle et al., 2012; Susha et al., 2017). *Modified data* undergoes data processing, such as cleaning, structuring, or labeling (Kuk, 2011; Nokkala et al., 2019). Data may be *enriched* with other data, such as combining product data with sales or customer data (de Corbière, 2009; Otto & Jarke, 2019). *Processed data* may be analyzed using algorithms or machine learning models (M. Spiekermann, 2019).

Source: The data source describes the origin of the data (Azkan et al., 2020). *Generated data* is produced directly by organizations, e.g., through sensors or web applications (Azkan et al., 2020; van de Ven et al., 2021). *Acquired data* is captured from previous data sharing and thus generated by other organizations (Hartmann et al., 2016; van de Ven et al., 2021). *Customer-provided data* encompasses data directly collected from customers (Lawrenz & Rausch, 2021; Treiblmaier & Pollach, 2007), with legal compliance and customer consent being inevitable (Choi & Kröschel, 2015; Rupasinghe et al., 2019). Further, organizations may capture and share *freely available data sources* (e.g., climate or satellite data) (Enders et al., 2020).

Privacy: Data can be categorized into four privacy classes based on factors like business criticality or sensitivity (Ackermann et al., 2022; Lindner et al., 2021). *Public data* can be shared without restriction (Gascó et al., 2018; Sayogo & Pardo, 2012). *Internal data* is business-relevant data shareable within the organization (Schäffer & Stelzer, 2018; Xiao et al., 2013), while *confidential data*, typically indicating business-critical data, is restricted and thus can be accessed and used only by authorized personnel (Azkan et al., 2020; Schäffer & Stelzer, 2018). *Sensitive data* includes personally identifiable data and mandates legal compliance or anonymization or pseudonymization when being shared (Ackermann et al., 2022; Treiblmaier & Pollach, 2007).

Sovereignty: Data sharing practices are legally defined by data sovereignty, reflecting control and authority over specific data (Bastiaansen et al., 2020; Lis & Otto, 2021). *Individual sovereignty* grants control over data to specific individuals, necessitating explicit consent when being captured or shared (Opriel et al., 2021; S. Spiekermann et al., 2015). *Organizational sovereignty* involves an organization's control over data, such as data from its production line sensors (Azkan et al., 2020; Schäffer & Stelzer, 2018). *Shared sovereignty* occurs when data is freely available (Gelhaar & Otto, 2020; Lis & Otto, 2021).

Timeframe: The timeframe dimension delineates if data sharing practices demand continuous (*(near-)real-time*) or frequent (*up-to-date*) sharing of data to ensure relevance or if the dataset shared is a static snapshot (Schäffer & Stelzer, 2018; van de Ven et al., 2021).

Meta-dimension: Organization

Motivation: The motivation describes an actor's intended purpose and motive for data sharing engagement (Gelhaar et al., 2021b). *Economic motivation* revolves around commercial interests like developing new services or business models (Choi & Kröschel, 2015; Steudner et al., 2019). *Social and environmental motivation* entails data sharing for social good, such as sustainability or corporate social responsibility (Sayogo & Pardo, 2011, 2012; Wilms et al., 2018). *Legal motivation* arises from obligations mandated by national or international regulations (Gascó et al., 2018; Grace, 2020). *Cultural motivation* involves the intrinsic or extrinsic embedding of data sharing within the organizational culture (Gelhaar et al., 2021b; Schäffer & Stelzer, 2018).

Scope: The scope defines the range of a data sharing practice (Lis & Otto, 2021). Data can be shared *intra-organizationally*, within the *same industry* (e.g., data being shared between suppliers and manufacturers), or *cross-industry* (e.g., sharing weather data between agriculture and aerospace companies) (Bastiaansen et al., 2020; de Corbiere & Rowe, 2013; Zaheer & Trkman, 2017).

Reward: Organizations share data for various reward purposes (Schweihoff et al., 2023a). *Financial rewards* indicate direct monetary returns (Agahari et al., 2021; Cichy et al., 2021; Spiekermann, 2019). Organizations can receive *virtual assets* such as cryptocurrency in return (Gelhaar et al., 2021b). Besides monetary rewards, organizations can receive *data* in return (Agahari et al., 2021). Shared data can lead to the development of novel *services* that are returned (Cichy et al., 2021; Kuk, 2011) or to increased *reputation*, e.g., through transparency or increased brand attractiveness (Lawrenz & Rausch, 2021; Nokkala et al., 2019).

Payment model: The payment model outlines how organizations are compensated for sharing data. Organizations may opt to provide their data *free of charge* (Enders et al., 2020; Muschalle et al., 2012). If payment is subject to the practice, organizations may receive a *fixed* payment or apply a *subscription-based* model, e.g., payment based on the duration of data access (Hartmann et al., 2016; M. Spiekermann, 2019). *Usage-based* models determine payment according to data consumption or access frequency (Gelhaar et al., 2021b), and *revenue-based* models are grounded on calculating compensation based on the revenue generated from shared data (M. Spiekermann, 2019). These payment models can be *hybrid*, combining different payment structures like freemium or mixing fixed and usage-based models (Gelhaar et al., 2021b).

Meta-dimension: network

Technical Architecture: Technical architecture can be provided in various architectural settings. *Closed application programming interfaces* (APIs) restrict access to authorized parties (Coelho et al., 2021; Xiao et al., 2013). Various forms of *platforms* like Otonomo or Dawex allow connecting multiple systems to a platform, storing, processing, and maintaining data on the platform, managing access to data, and enabling commercial data exchange (Nokkala et al., 2019; Otto & Jarke, 2019; van de Ven et al., 2021). *Open APIs* offer unrestricted access to data (de Corbière & Rowe, 2011; Enders et al., 2020; Kuk, 2011).

Interoperability: According to the European Interoperability Framework by the European Commission, in the realm of data sharing, multiple levels of interoperability are distinguished that are necessary when data is shared across two or multiple information systems (Bastiaansen et al., 2020). Thereby, the term interoperability describes a common language, protocols, standards, and mechanisms that enable seamless sharing and usage of data across disparate information systems (Bastiaansen et al., 2020; Coelho et al., 2021). Orchestrating data sharing practices can necessitate different types of interoperability (Nokkala et al., 2019). *Legal interoperability* describes the necessity of compatible legal frameworks, standards, and regulations essential for sharing data across different jurisdictions (Bastiaansen et al., 2020; Grace, 2020). *Organizational interoperability*, such as aligned processes, objectives, and resources, is vital for seamless data sharing and utilization of data across organizational boundaries (Bastiaansen et al., 2020). *Semantic interoperability* ensures consistency, interpretability, and accuracy of data across diverse sources and information systems (Coelho et al., 2021). *Technical interoperability* may be required to enable seamless technical data exchange between information systems (Bastiaansen et al., 2020; M. Spiekermann, 2019).

Access coordination: Access coordination defines the mechanism ensuring controlled data access (Nokkala et al., 2019). *Agreement-based* access relies on data sharing agreements between providers and consumers defining access terms (Rukanova et al., 2020; Schäffer & Stelzer, 2017). *Application-based* access regulates data access through application-specific access rights (Susha et al., 2017). *Trust-based* access coordination grants access based on trust between provider and consumer, omitting coordination or legal liabilities (Cichy et al., 2021; Susha et al., 2017). No access coordination refers to being *open* (Sayogo & Pardo, 2012; Wilms et al., 2018).

Provision request: The provision request signifies the event that triggers the data sharing practice (Susha et al., 2017). This trigger can be characterized by *on-demand* initiation, prompted by a data consumer's request, *event-based*, such as accessing a system, or *continuous* provision, involving automatic data sharing in real-time or at regular intervals (Cichy et al., 2021; Liu & Kumar, 2003).

Type of relationship: The relationship type is dictated by the count of providers and consumers. Two types of relationships involve a single data provider sharing with one data consumer (*one-to-one*) or with multiple data consumers (*one-to-many*) (Schäffer & Stelzer, 2017; van de Ven et al., 2021). The other two types of relationships entail multiple data providers sharing either with one data consumer (*many-to-one*) or with multiple consumers (*many-to-many*), such as sharing data in data ecosystems (van den Broek & van Veenstra, 2015).

Evaluation of the taxonomy on a real-world data sharing practice

To evaluate our taxonomy, we demonstrate the taxonomy on a real-world data sharing practice of an organization (anonymized as 'Alpha') in the chemical industry. Alpha offers products for various sectors, including agriculture. Further, Alpha collects field trial data on the growing condition of agricultural fields treated with Alpha's products, such as fertilizers. In the case discussed with the interviewees, Alpha received a data request from the company 'Beta,' aiming to enhance the efficiency and accuracy of its risk assessment service for agricultural product applications and to build a Europe-wide database. For this evaluation, four experts apply the taxonomy from Alpha's viewpoint as the data provider. The results are depicted in Fig. 4.

The assessment of the case by demonstrating the taxonomy and the experts' explanations are outlined in detail in the previously conducted and communicated research (Fassnacht et al., 2024). In the following, we focus on presenting the evaluation of the subjective ending conditions by the experts. The experts emphasize the applicability and usefulness of the taxonomy in practice, as Alpha lacks a structured approach to characterizing and conceptualizing data sharing practices (I1, I2). I1 states: '[…] we perform each data sharing case individually by a designated project team, and each team documents it differently in PowerPoint and Excel files.' The experts highlight the complexity of the topic ('there is nearly every business unit involved, legal, business, IT guys, we as data stewards, and it is highly complex to find consensus.' (I2)) and stress the need for a dynamic taxonomy to adapt to diverse data sharing practices (I1, I2). Currently, Alpha relies on *privacy* as the primary criterion for data-sharing decisions as I1 mentions, '[…] we only consider data shareable if they are labeled as internal or public, we do not share any confidential data per se across legal entities' (I1, I4). Successful practices within Alpha are seen as isolated individual solutions lacking a systematic design (I1, I3). Regarding the taxonomy's conciseness (E9) and robustness (E10), I1 and I4 stress the necessity of multiple dimensions and the variety of characteristics to capture the diverse aspects of data sharing practices ('[the taxonomy] could be even more detailed, but then we lose ourselves in detail again. It is enough to create a common understanding of a case and find similarities and differences.' (I4)). Further, I2 and I3 emphasize the comprehensiveness of the characteristics that correspond to the multitude of design options of data sharing practices (E11) ('due to so many facets, legal, technical, business perspectives, and the various options and characteristics to consider for each case, it needs this variety […] to describe a case' (I2)). They further suggest initial ideas for extending the taxonomy to include industry- or company-specific dimensions (e.g., decision authority or data formats) (I2, I4), indicating that it

	Dimension	Characteristics						E/N
Data	Domain	**Process data**	Product data	**Environment data**	Customer data	Meta-data		N
	Maturity of processing	Raw data	**Modified data**	Enriched data	Processed data			N
	Source	**Generated**	Acquired	Customer-provided	Free available			N
	Privacy	Public	**Internal**	Confidential	Sensitive			E
	Sovereignty	Individual	**Organizational**	Shared				E
	Timeframe	**Static**	Up-to-Date	(Near-) real-time				N
Organization	Motivation	**Economic**	**Social & Environmental**	**Legal**	Cultural			N
	Scope	Intra-organizational	**Same industry**	Cross-industry				E
	Reward	Financial	Virtual assets	**Data**	**Service**	**Reputation**		N
	Payment model	**Free**	Fixed	Subscription-based	Usage-based	Revenue-based	Hybrid	E
Network	Technical architecture	Closed API	**Platform**	Open API				E
	Interoperability	**Legal**	Organizational	**Semantic**	Technical			N
	Access coordination	**Agreement-based**	Application-based	Trust-based	Open			E
	Provision request	**On demand**	Event-based	Continuous				E
	Type of relationship	**One-to-one**	One-to-many	Many-to-one	Many-to-many			E

Fig. 4 Evaluation of the taxonomy by demonstration on a real-world data sharing practice

is easily extendable (E12). Additionally, they note the clarity in the explanations of dimensions and characteristics without redundancies (E13) as I3 states, '[…] the dimensions and characteristics are self-explaining.'

Archetypes of data sharing practices

In this subsection, we present the results of the cluster analysis and its interpretation, which revealed four distinct archetypes of data sharing practices. The clusters comprise 19 to 47 cases and expose varying emphases across the taxonomy's characteristics for each cluster. Figure 5 illustrates the characteristics' frequency distribution across the clusters, highlighting prominent and delineating characteristics. Through careful examination of the cases and distinct focal points, we delineated the four archetypes of data sharing practices: (I) Compliance-oriented, (II) efficiency-oriented, (III) revenue-oriented, and (IV) society-oriented data sharing practices.

In the following, we outline the delineating characteristics for each archetype and provide exemplary data sharing practices from the analyzed sample, summarized in Table 5.

Archetype I: compliance-oriented data sharing practices

Archetype I includes data sharing practices that prioritize regulatory compliance in cases where organizations must proactively ensure data security and prevent unauthorized access or breaches.

One example of this archetype is the *Gaia-X-Medical-Records* case (UC117), where medical data such as diagnoses, decision protocols, and personal data is gathered, structured, and stored in a centralized repository of patient data. This centralized approach to storing sensitive medical data necessitates a delicate balance between the value generated (in this case, medical benefits) and legal obligations. Without assurances of privacy safeguards before disclosing their information, customers may be reluctant to contribute data for such purposes. Consequently, such data sharing practices may fail to yield the insights they could provide. Hence, in the case of Gaia-X-MedicalRecords, patients retain ownership of their data, affording them transparency and control over its utilization. Access to the data is only granted through a blockchain-secured communication channel, ensuring compliance with the sensitive nature of the data shared.

Another example is the *Informes De Movilidad* case (UC015). During the initial stages of the COVID-19 pandemic, Telefónica Chile shared telecommunication data records obtained from their customers with the Data Science Institute of the National University for Development. Through analysis of mobility patterns using this dataset, the Chilean government was able to avoid nationwide shutdowns and instead implement localized, temporary mobility restrictions. Integration of available data for service delivery empowered the Chilean government to augment its existing dataset and extract more meaningful insights. Informes De Movilidad employed closed API technology to meet regulatory standards for the transfer of anonymized telecommunications data between stakeholders.

Safeguarding data security is of particular importance in such cases as the majority of data sharing practices in archetype I involve handling sensitive data (customer data: 97%, sensitive data: 100%), predominantly sourced directly from customers (customer-provided data: 90%). The timeliness of data is maintained through regular updates (up-to-date data: 90%), ensuring its relevance for both regulatory compliance and organizational utility.

Further, cases of archetype I primarily address socially motivated data sharing practices (social and environmental motivation: 77%), wherein the shared data is offered to third parties free of charge (free data: 87%). Rather than monetary reimbursement, data providers receive services in return (service reward: 80%), which they can integrate into their business activities.

Interoperability among data-sharing entities prioritizes legal considerations (legal interoperability: 83%), emphasizing the sensitive nature of the data and its regulatory implications. A slightly smaller proportion of cases address technical aspects as influential factors (technical interoperability: 77%), reflecting the stringent security requirements for sensitive data. Participating organizations ensure stringent access control to shared data by regulating data access through mutually agreed-upon legal contracts (agreement-based access control: 80%). Such a regulatory framework is crucial for compliance with on-demand data requests (on-demand provision request: 93%) and defining clear access rights embodied in the technical solution.

Archetype II: efficiency-oriented data sharing practices

Archetype II encompasses data sharing practices wherein organizations share data to leverage data-driven services to enhance industrial processes. These processes may include boosting machine productivity, streamlining supply chain operations, or reducing resource consumption. In contrast to archetype I, which prioritizes regulatory compliance, archetype II is driven by financial objectives.

Siemens' *Railigent* platform (UC067) exemplifies a case related to archetype II within the railway sector. Railigent functions as a platform, as it records, processes, and analyzes data provided by railway organizations and offers data-driven services in return. For instance, the Renfe Spanish Rail Company shares sensor data on critical train components through the platform. This enables the development

Fig. 5 Frequency distribution of characteristics along the four identified archetypes

Dimension	Characteristic	Archetype: I. Compliance-Oriented Data Sharing Practices	II. Efficiency-Oriented Data Sharing Practices	III. Revenue-Oriented Data Sharing Practices	IV. Society-Oriented Data Sharing Practices
Number of cases per cluster		30	47	19	25
Domain	Process data	10 (33%)	35 (74%)	13 (68%)	19 (76%)
	Product data	1 (3%)	28 (60%)	14 (74%)	5 (20%)
	Environment data	0 (0%)	16 (34%)	4 (21%)	18 (72%)
	Customer data	29 (97%)	3 (6%)	6 (32%)	0 (0%)
	Meta-data	2 (7%)	5 (11%)	1 (5%)	3 (12%)
Maturity of processing	Raw data	5 (17%)	35 (74%)	16 (84%)	20 (80%)
	Modified data	22 (73%)	6 (13%)	8 (42%)	2 (8%)
	Enriched data	2 (7%)	0 (0%)	5 (26%)	0 (0%)
	Processed data	12 (40%)	19 (40%)	6 (32%)	5 (20%)
Source	Generated	14 (47%)	44 (94%)	18 (95%)	16 (64%)
	Acquired	6 (20%)	4 (9%)	3 (16%)	14 (56%)
	Customer-provided	27 (90%)	8 (17%)	7 (37%)	0 (0%)
	Free available	1 (3%)	5 (11%)	1 (5%)	2 (8%)
Privacy	Public	0 (0%)	0 (0%)	0 (0%)	4 (16%)
	Internal	0 (0%)	23 (49%)	18 (95%)	20 (80%)
	Confidential	0 (0%)	21 (45%)	0 (0%)	1 (4%)
	Sensitive	30 (100%)	3 (6%)	1 (5%)	0 (0%)
Sovereignty	Individual	16 (53%)	4 (9%)	0 (0%)	1 (4%)
	Organizational	14 (47%)	41 (87%)	19 (100%)	22 (88%)
	Shared	0 (0%)	2 (4%)	0 (0%)	2 (8%)
Timeframe	Static	2 (7%)	0 (0%)	5 (26%)	1 (4%)
	Up-to-date	27 (90%)	7 (15%)	12 (63%)	24 (96%)
	(Near-) real-time	3 (10%)	43 (91%)	4 (21%)	2 (8%)
Motivation	Economic	11 (37%)	46 (98%)	19 (100%)	9 (36%)
	Social & Environmental	23 (77%)	21 (45%)	0 (0%)	25 (100%)
	Legal	4 (13%)	4 (9%)	0 (0%)	1 (4%)
	Cultural	11 (37%)	7 (15%)	0 (0%)	12 (48%)
Scope	Intra-Organizational	0 (0%)	1 (2%)	1 (5%)	0 (0%)
	Same industry	22 (73%)	38 (81%)	9 (47%)	20 (80%)
	Cross-industry	8 (27%)	8 (17%)	9 (47%)	5 (20%)
Reward	Financial	6 (20%)	4 (9%)	16 (84%)	6 (24%)
	Virual assets	0 (0%)	1 (2%)	1 (5%)	0 (0%)
	Data	4 (13%)	11 (23%)	4 (21%)	3 (12%)
	Service	24 (80%)	46 (98%)	10 (53%)	24 (96%)
	Reputation	4 (13%)	2 (4%)	1 (5%)	19 (76%)
Payment model	Free	26 (87%)	44 (94%)	2 (11%)	25 (100%)
	Fixed	1 (3%)	0 (0%)	0 (0%)	0 (0%)
	Subscription-based	0 (0%)	1 (2%)	7 (37%)	0 (0%)
	Usage-based	0 (0%)	0 (0%)	4 (21%)	0 (0%)
	Revenue-based	0 (0%)	0 (0%)	1 (5%)	0 (0%)
	Hybrid	3 (10%)	2 (4%)	5 (26%)	0 (0%)
Technical architecture	Closed API	15 (50%)	9 (19%)	0 (0%)	3 (12%)
	Platform	14 (47%)	37 (79%)	19 (100%)	21 (84%)
	Open API	1 (3%)	1 (2%)	0 (0%)	1 (4%)
Interoperability	Legal	25 (83%)	9 (19%)	3 (16%)	11 (44%)
	Organizational	16 (53%)	40 (85%)	17 (89%)	19 (76%)
	Semantic	9 (30%)	11 (23%)	3 (16%)	2 (8%)
	Technical	23 (77%)	45 (96%)	18 (95%)	17 (68%)
Access coordination	Agreement-based	24 (80%)	42 (89%)	0 (0%)	19 (76%)
	Application-based	2 (7%)	2 (4%)	8 (42%)	0 (0%)
	Trust-based	0 (0%)	1 (2%)	0 (0%)	1 (4%)
	Open	4 (13%)	2 (4%)	11 (58%)	5 (20%)
Provision request	On demand	28 (93%)	4 (9%)	19 (100%)	21 (84%)
	Event-based	1 (3%)	3 (6%)	0 (0%)	4 (16%)
	Continous	1 (3%)	40 (85%)	0 (0%)	0 (0%)
Type of relationship	One-to-one	6 (20%)	5 (11%)	0 (0%)	2 (8%)
	One-to-many	3 (10%)	1 (2%)	1 (5%)	2 (8%)
	Many-to-one	4 (13%)	16 (34%)	0 (0%)	4 (16%)
	Many-to-many	17 (57%)	25 (53%)	18 (95%)	17 (68%)

0% - 20% | 21% - 50% | 51% - 80% | 81% - 100%

Table 5 Overview of the identified archetypes of data sharing practices

No.	Archetype	Delineating characteristics	Application examples
(I)	Compliance-oriented data sharing practices	• Sensitive customer data • Data provision in exchange for services • Importance of complying with legal regulations • High IT security and data privacy demands	Medical records, customer analysis
(II)	Efficiency-oriented data sharing practices	• Data provision in exchange for services • Indirect economic incentives • Continuous provision of real-time data • Organizational and technical interoperability	Production process optimization, supply chain optimization, IoT applications
(III)	Revenue-oriented data sharing practices	• Monetization strategy • Data trading via marketplaces • Organizational and technical interoperability	IoT data marketplaces, selling of data assets, data standardization
(IV)	Society-oriented data sharing practices	• Social or environmental focus • Data provision free of charge • Combination of multiple data sources • Use case-based agreements	Development aid, social services, ecologic applications

and delivery of predictive maintenance services aimed at minimizing technical failures. The ability to detect signs of wear and prevent damage underscores the significance of sharing real-time data. Consequently, the data sharing practice contributes to significantly improving the data-driven services that ensure trains operated by the Spanish Rail Company arrive on time.

In the VW-*SupplyChain* case (UC111), German car manufacturer Volkswagen collaborates with its supplier ThyssenKrupp in a data sharing practice to optimize supply chain operations. This involves sharing data concerning production forecasts, plant capacities, and supply contracts for upcoming car models. Since safeguarding confidential information on production planning and fostering a significant level of trust is a challenge in this collaboration, organizational characteristics such as supply contracts and trustworthiness are pivotal.

Archetype II cases primarily involve data derived from industrial processes or smart products (process data: 74%, product data: 60%) generated by the data provider (generated data: 94%). This data serves data consumers as a foundation for developing services primarily to enhance operational efficiency. Since shared data is often claimed to be internal and confidential (internal: 49%, confidential: 45%), the organizations providing the data typically maintain control and sovereignty over their data (organizational sovereignty: 87%). Notably, the data is often made available in real-time (91%), which is another distinctive characteristic of archetype II cases.

Economic incentives are prominent in archetype II cases (economic motivation: 98%). Organizations engage in data sharing to enhance their operational efficiency, aiming to increase productivity or reduce resource consumption. Due to this indirect compensation, they mostly opt for non-monetary rewards, resulting in a payment model free of charge instead of direct trading (free payment model: 94%), thus enabling external development of data-based services, which they receive in return (service reward: 98%). Data sharing practices within archetype II primarily center on sharing data within the same industry (same industry scope: 81%).

Archetype II cases display a preference for platform-based data sharing (platform architecture: 79%). This necessitates robust technical integration among involved actors to harmonize data formats, transmission protocols, and service delivery channels, characterizing archetype II (technical interoperability: 96%). Additionally, integration between organizations is pivotal (organizational interoperability: 85%). Most archetype II practices rely on agreements among actors to monitor and coordinate access restrictions as needed (agreement-based access coordination: 89%). With a predominant focus on continuous data provision (85%), timely service applications and decision-making support are facilitated.

Archetype III: revenue-oriented data sharing practices

Archetype III comprises data sharing practices, whose primary objective is the generation of direct revenue streams through monetizing data as an asset.

An example of an archetype III data sharing practice is *Terbine* (UC092), a platform enabling the monetization of machine data. Through the Terbine platform, organizations can make their data available for purchase, allowing customers to acquire and utilize the data according to their specific needs. Terbine primarily caters to organizations utilizing IoT devices to generate data, offering meta-data harmonization mechanisms and a data catalog for browsing available data products.

Another illustration is the *CARUSO* data marketplace (UC074), a platform on which organizations in the automotive and mobility sector generate direct revenue streams from trading data via the CARUSO. Further, CARUSO utilizes a hybrid pricing model, combining regular membership fees with an additional pay-per-use system. CARUSO also offers a development portal to facilitate technical data integration for clients.

Product and process data are both common in archetype III (product data: 74%, process data: 68%). These practices typically utilize infrastructure to handle machine-generated data from industrial processes in its raw form, ensuring versatility in its usage (raw data: 84%, generated data: 95%). To prevent the leakage of sensitive information, data in archetype III practices is devoid of any sensitive or confidential data (internal data: 95%).

Archetype III data sharing practices are primarily driven by financial incentives, with 100% exhibiting economic motivation. A defining characteristic of archetype III is the direct monetization of data. This is reflected in the substantial financial rewards organizations pursue (financial reward: 84%). Payment models in archetype III vary considerably. While 89% of practices adopt a monetization approach, the specific payment models differ (subscription-based: 37%, usage-based: 21%, hybrid: 26%). Notably, some cases within this archetype lack a defined payment structure (e.g., UC058, UC080). This may suggest plans to implement pricing strategies later while initially focusing on expanding their customer base.

Data sharing platforms are the sole technical architecture evident in all archetype III cases (platform architecture: 100%), enabling organizations to offer their data to a broad spectrum of customers. These platforms also facilitate the seamless technical integration of data streams between data-sharing entities and data consumers, another delineating characteristic of archetype III (technical interoperability: 95%). Alongside technical considerations, organizational interoperability plays a pivotal role (89%) in incorporating external data into business processes. This becomes pertinent when numerous organizations are involved in data sharing practices (many-to-many relationship type: 95%). Data consumers usually request data on demand (on-demand provision request: 100%).

Archetype IV: society-oriented data sharing practices

In contrast to archetypes II and III, archetype IV encompasses data sharing practices motivated by social or environmental concerns rather than seeking economic benefits as their primary objective.

An example is the *CABI* case (UC004), where agricultural data concerning soil and crop quality is shared to aid smallholder farmers in sub-Saharan Africa. Collaborating with the Bill & Melinda Gates Foundation, they address the effects of climate change on productivity and its associated social ramifications. Data from diverse sources is combined and shared with numerous smallholder farmers to assist them in enhancing environmental, economic, and social factors. This case emphasizes the challenge of organizational interoperability, as structures within supporting organizations like the Bill & Melinda Gates Foundation must align with operational processes at the smallholder farmers' application sites.

Another example is *RNLI*'s open marine data project (UC036), which focuses on sharing data acquired from various sources to enhance maritime safety. Enriching data from coastguard agencies and the Royal National Lifeboat Institute with geographic data like satellite imagery or aerial footage results in an extensive database. This facilitates the creation of more efficient and reliable lifeguard services. Enhancing reputation may serve as a means for spreading awareness about the project and attracting new donations to expand its scope and finance further advancements.

Data sharing practices clustered in archetype IV focus on sharing process data (76%) and environmental data (72%). The shared data is predominantly unprocessed (raw data: 80%), allowing for diverse analytics applications and meaningful insights generation. In most cases, the organization sharing the data maintains control and sovereignty (organizational sovereignty: 88%). Similar to archetype III, archetype IV data sharing practices emphasize timely data (up-to-date timeframe: 96%) to ensure relevance and prevent data obsolescence.

Data sharing practices in archetype IV exhibit a strong social or ecological motivation (social and environmental motivation: 100%). In these cases, data is shared among organizations to pursue common objectives. Consequently, the shared data typically remains within a single industry (same industry scope: 80%). Similar to archetypes I and II, the primary benefit for engaging organizations is derived from data-based services (service reward: 96%). The desire to enhance reputation serves as an additional incentive for data sharing (reputation reward: 76%), delineating archetype IV significantly. The altruistic nature of these incentives is further demonstrated by the free availability of shared data (free payment model: 100%), aiming to facilitate data accessibility and its utilization for respective purposes.

Cases in archetype IV predominantly engage in data sharing practices through data sharing platforms serving as central hubs (platform architecture: 84%). This enables organizations to aggregate and complement data from diverse sources and formats, allowing them to derive more profound insights. Data sharing practices within this archetype necessitate both organizational (76%) and technical (68%) interoperability ensured through actors utilizing agreements to coordinate data access (agreement-based access coordination: 76%). The provision of data in response to specific application demands aligns with the use case-oriented nature of archetype IV instances (on-demand provision request: 84%).

Discussion

Our research conceptualizes data sharing practices in interorganizational settings regarding their defining dimensions and characteristics. For this purpose, we developed a taxonomy and identified four distinct archetypes that represent conceptual abstractions of related literature and real-world data sharing practices. Our research enables organizations to extract and unlock the potential of data gathered for innovation, monetization, and value creation. Although studying the phenomenon of data sharing is increasingly gaining momentum in the IS field (Cattaneo et al., 2020; Gelhaar et al., 2021b; Jussen et al., 2023; Krotova et al., 2020; Richter & Slowinski, 2019), existing research so far only focused on specific aspects and activities of data sharing practices, such as developing business models (Arenas et al., 2019; van de Ven et al., 2021), designing incentive mechanisms (Gelhaar et al., 2021b), and focusing on data privacy and security (Bastiaansen et al., 2020; Cichy et al., 2021). To the best of our knowledge, there was no conceptual work and empirical evidence to conceptualize the practice of data sharing in general, which inhibited further holistic efforts to support the design of data sharing practices. Our work resolves this issue by proposing a taxonomy that captures the peculiarities of data sharing by combining dimensions that focus on the data as the shareable entity (e.g., *data domain, privacy*), the organization as a key actor (e.g., *motivation*, *reward*), and the entire configurational dynamics and setting of the network (e.g., *interoperability*, or *type of relationship*).

As a second contribution, we empirically derived four distinct archetypes and underscored their relevance by demonstrating their distinct focus through real-world cases of data sharing practices. This set of archetypes distinguishes different types of data sharing practices built on the proposed taxonomy to characterize each of these real-world objects. The resulting archetypes of data sharing practices are interpreted as differentiated primarily based on the core motivation for data sharing of the actors involved, which is often the ultimate determinant of whether data is shared or not (Gelhaar et al., 2021b; Müller et al., 2020). Further, the archetypes reveal the interplay of dimensions and characteristics along all three meta-dimensions—data, organizational structures, and network dynamics. As such, archetype I is primarily concerned with regulatory compliance, whereas archetype II is oriented towards efficiency and technological integration. Archetype III is characterized by an emphasis on sellable and interoperable data, while archetype IV is oriented towards sharing data for social and ecological well-being. These core motivations often unfold in structural and architectural consequences that lead to a correlated differentiation in the characteristics of data sharing practices among the archetypes, which we were able to uncover through our taxonomy-based coding and cluster analysis.

Archetype analysis and comparison

Archetype I reflects a strong orientation toward sharing data compliant with existing regulations, such as the General Data Protection Regulations or the Data Governance Act (Cichy et al., 2021; Mollick, 2016; S. Spiekermann et al., 2015). It includes data sharing practices that rely on legal interoperability and typically involve sharing sensitive personal data about individuals. The importance of this archetype is underpinned by recent and future efforts of regulatory bodies to develop and establish policies, standards, and regulations, such as the Supply Chain Act. In this context, data trusts are gaining momentum as independent intermediaries that ensure privacy and security for data sharing and add algorithmic solutions for safeguarding data and its usage across organizational boundaries (Czech et al., 2023; Lauf et al., 2023; Schweihoff et al., 2023b). While existing literature (e.g., Schweihoff et al., 2023b) focuses on the design and characteristics of intermediaries, they outline the necessity to transcend the descriptive domain toward a more prescriptive domain and interdisciplinary research. Hence, they call for elaborating on design solutions for sharing data since recent regulatory endeavors such as the Data Governance Act and DA might mandate specific functions and practices (Schweihoff et al., 2023b) to which we contribute with archetype I.

Archetypes II and III are rooted in the overall aspiration of organizations for continuous improvement and economic growth. Our findings are closely tied to related emerging streams of research on data monetization, data ecosystems, data trading, and data marketplaces. These fuel the endeavors of organizations to share data beyond organizational boundaries, e.g., for improving production efficiency, service innovation (e.g., predictive maintenance), value co-creation (novel joint business models), and financial revenue (e.g., selling trial data) (Abbas et al., 2021; Gelhaar et al., 2021b; M. Spiekermann, 2019; Sterk et al., 2022). With these archetypes, we contribute to the call of, i.e., Jussen et al. (2024b), outlining the need for research on understanding data sharing operationalization, the critical considerations in the practice, and its implications for internal optimization and value co-creation as an essential prerequisite for practitioners' future decision-making processes on data sharing.

In contrast, archetype IV addresses data sharing practices that reflect a growing interest in sustainability and corporate social responsibility of organizations to contribute to the overall well-being of society (Enders et al., 2022; Janssen et al., 2012; Sayogo & Pardo, 2012; Susha & Gil-Garcia, 2019). The emerging concept of data collaboratives, which focuses on sharing data in partnerships between organizations for a common good (Susha et al., 2017; Susha & Gil-Garcia, 2019) or releasing data publicly (Enders et al., 2022; Janssen et al., 2012; Zuiderwijk et al., 2014), is particularly pursued by organizations to contribute to a better environment and life for society. With archetype IV, we can respond to various calls in research on understanding the determinants of sharing data in collaborative networks to encourage enthusiasts (Sayogo & Pardo, 2012) to support in structurally describing different forms of data sharing to allow the derivation of advantages and disadvantages (Susha et al., 2017), and the call for additional research to deepen our understanding of the potential impact of data sharing by organizations to engage in societal value creation (Enders et al., 2022).

In comparing the archetypes, two underlying perspectives were recognized. The first relates to the different roles platforms can take in data sharing practices. In archetype III, the platform is the only technological architecture present, being implemented as a data marketplace that facilitates the trading of data assets by matching supply and demand (Abbas et al., 2021; Richter & Slowinski, 2019). Archetypes II and IV, on the other hand, do not exclusively employ platforms. Rather than focusing on market interactions, they serve as a common space to exchange data from different domains and provide adjacent services such as query searches or data harmonization (Agahari et al., 2021; Schweihoff et al., 2023a). This allows organizations to engage and co-create value through these data sharing practices (Dreller, 2018). While data privacy-preserving services may encourage the use of data platforms to share sensitive data, as in archetype I (van den Broek & van Veenstra, 2015), our analysis reveals that only half of the sampled practices use data platforms. Closed APIs seem to provide even greater reliability and data security, and the connection of multiple different parties plays a subordinate role.

Second, differences in the perception of value in and from data are highlighted through the archetypes. In the question of what value is assigned to data, the archetypes differ in that archetype IV is characterized by the view that data is a public good. This contrasts in particular with archetype III, which tends to view data as an internal asset. These different perceptions consequently translate into various approaches to generating value from data. Concerning the type of value, archetypes I and IV describe use cases for sharing data where primarily social or environmental value is created, while archetypes II and III are clearly focused on economic value generation. A sharp distinction may however not always be applicable, as social or environmental motivations can lead to indirect financial effects (e.g., by improving a company's reputation, which consequently leads to increases in sales). Regarding the immediateness of value manifestation, the archetypes represent a continuum, with archetype III representing direct monetary value creation through the selling of data and archetype IV representing indirect or long-term value creation triggered by potential spillover effects.

This comparison emphasizes the importance of assessing the unique requirements of a data sharing practice and designing it to meet its specificities, with the understanding of data value and, according to core motivation, often being the central starting point.

Theoretical implications

For researchers in the field of data sharing, the presented taxonomy provides a standardized framework for describing, classifying, and configuring data sharing practices and enables scholars to effectively contextualize their work. As an analytic theory, our study contributes to the organization of knowledge in the emerging field of data sharing within IS research and facilitates the systematic understanding and analysis of data sharing practices (Gregor, 2006). Moreover, the presented taxonomy allows researchers to triangulate research on related topics (e.g., data ecosystems, data marketplaces, data spaces), elucidating design options of data sharing practices as the constituent activity of these concepts and facilitating the development of rigorous theories to accumulate both descriptive and prescriptive knowledge in the field of data sharing (Gregor & Hevner, 2013). Consequently, the taxonomy and archetypes can build the foundation for studies pursuing design-oriented objectives such as designing a repository for concepts and best practices of data sharing practices, methods for individual and group work based on the taxonomy (e.g., cards with general descriptive information, supplemented with industry- or organization specific knowledge), or applying the taxonomy to real-world cases to build systematized repositories serving the development and application of computationally-intensive approaches (e.g., applying generative AI) to derive novel patterns. While research and practical endeavors on data ecosystems and data spaces are still in emergence, the taxonomy supports amalgamating aspects of these concepts, such as governance design, infrastructural design, and actor engagement strategies towards sustainably establishing data sharing practices within these networks. Our taxonomy can further serve as a foundational framework for developing rigorous theories, incorporating existing information systems (IS) theories. This includes the resource-based view, which aligns data sharing design options with organizational resources; actor-network theory, which explores the interactions and interdependencies among actors in data sharing within related concepts; and socio-technical systems theory, which examines the interdependent social and technical components of complex data sharing systems. Building on this conceptualization, the identified archetypes further develop a foundational understanding of data sharing practices in the real world, empirically contributing to the body of knowledge on configurational options for data sharing. The archetypes provide an in-depth characterization of data as the shareable entity, organizational structures, as well as interorganizational network dynamics, addressing recent calls for conceptualizing data sharing practices, elaborating on the interplay of the multiple dimensions and participation requirements (Azkan et al., 2020) by examining archetypes of data sharing (Schweihoff et al., 2023a). As another empirical implication, we offer a systematically analyzed dataset comprising 121 empirical cases of data sharing practices that illustrate various configurational options in Table 6 in the Appendix. Focusing only on publicly available sources ensures reproducibility and extendibility for future research, providing a valuable resource for further studies on data sharing practices.

Practical implications

For practitioners, particularly decision-makers, our findings enhance the ability to establish effective data sharing practices, thereby maximizing the potential of data sharing across organizational boundaries. Our research offers clear, actionable guidance on configuring data sharing practices and operationalizing strategic approaches to data sharing, applicable at both the ideation and implementation stages. In addition, our study provides a detailed, empirical overview of data sharing practices, presenting four ideal archetypes that aid decision-makers in configuring and evaluating practices to fit their organization's specific context. The derived archetypes and corresponding cases (see Table 5) serve as reference points and cases for comparison, facilitating this configuration process in practice and the systematic identification and communication of similarities and differences between real-world cases. This is further supported by the cross-table analysis (see Fig. 5), which offers insights into how market participants currently organize their data sharing practices. Consequently, the presented findings provide a robust framework for understanding configuration options and pathways to establish goal-oriented data sharing practices from a multidimensional perspective. Therefore, this framework can influence decision-making processes by providing a structured approach to understanding and navigating the complexities of data sharing. Decision-makers can leverage this framework to identify viable data sharing strategies, align them with organizational goals and existing networks, and anticipate potential challenges and opportunities.

Limitations and future research opportunities

Our research has limitations that point to opportunities for future research. Taxonomy-based research is inherently dynamic, capturing a moment in time (Nickerson et al., 2013), which also applies to our specific study and its findings. Given the evolving nature of the field, future research should revisit and expand upon our findings to maintain relevance. Current and future advancements in data ecosystems, data spaces, and legislation are expected to shape future practices and may result in changing configurations or other archetypes. Therefore, our findings may not encompass future innovations, as they are based solely on existing practices.

Our study primarily relies on publicly available data sharing practices and may miss emerging or not publicly communicated cases, dimensions, and characteristics (e.g., decision-making authority or funding). Future studies should address these gaps by directly engaging with companies to validate our results and augment the dataset and findings with novel insights. Further, as we aim to develop a taxonomy to inform the conceptualization and design of data sharing practices, we excluded, e.g., literature on the technical mechanisms that might lead to lacking dimensions and characteristics that focus on the realization and implementation of the technical exchange of data. Hence, completeness cannot be guaranteed, which offers potential for future research, such as extending the taxonomy towards including (sub-ordinate) dimensions and characteristics regarding the (technical) realization of data sharing practices or more differentiated elaboration of dimensions, e.g., the 'technical architecture.' Further, due to the complexity and multidimensionality of data sharing practices, the taxonomy could be expanded by, e.g., applying hierarchical taxonomy development (Nickerson et al., 2024) with multiple categories, dimensions, and corresponding characteristics to describe data sharing practices holistically at any hierarchical level. While our research provides an industry-agnostic taxonomy and archetypes for data sharing practices, future research should reveal insights into industry-specific nuances and peculiarities, such as regulations in the pharma and agricultural sectors or the additional complexity of competition and cartel rights in the automotive industry. To evaluate the archetypes, we used quantitative and qualitative measures. However, future research could qualitatively validate our findings through expert interviews that may provide additional insights and unveil interrelations among archetypes and critical strategic decision-making factors that guide companies in adopting different archetypes in the design phase of data sharing practices.

Conclusion

Our study conceptualizes the characteristics and design options for data sharing practices, resulting in two artifacts as its primary contributions: first, a theoretiscally grounded and empirically validated taxonomy that summarizes the dimensions and characteristics of data sharing practices, and second, four generic archetypes that represent recurring patterns across all characteristics. To ensure both scientific rigor and practical relevance, we developed our findings based on existing literature and 121 real-world data sharing practices. Our findings contribute to the descriptive and analytical knowledge in the field of data sharing by providing an integral view of data sharing practices and consolidating and organizing existing knowledge on data sharing. Consequently, we contribute to a fundamental understanding of data sharing with empirical, theoretical, and practical implications to guide and promote future research endeavors as well as the efforts of organizations willing to engage in data sharing practices.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12525-024-00732-0.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

Abbas, A., Agahari, W., van de Ven, M., Zuiderwijk, A., & Reuver, M. (2021). Business data sharing through data marketplaces: A systematic literature review. *Journal of Theoretical and Applied Electronic Commerce Research, 16*, 3321–3339. https://doi.org/10.3390/jtaer16070180

Ackermann, K. A., Burkhalter, L., Mildenberger, T., Frey, M., & Bearth, A. (2022). Willingness to share data: Contextual determinants of consumers' decisions to share private data with companies. *Journal of Consumer Behaviour, 21*(2), 375–386. https://doi.org/10.1002/CB.2012

Agahari, W., Dolci, R., & de Reuver, M. (2021). Business model implications of privacy-preserving technologies in data marketplaces: The case of multi-party computation. *ECIS 2021 Proceedings.*

Alt, R. (2020). Evolution and perspectives of electronic markets. *Electronic Markets, 30*(1), 1–13. https://doi.org/10.1007/s12525-020-00413-8

Alt, R. (2020). Electronic markets on business model development. *Electronic Markets, 30*(3), 405–411. https://doi.org/10.1007/s12525-020-00438-z

Arenas, A. E., Goh, J. M., & Matthews, B. (2019). Identifying the business model dimensions of data sharing: A value-based approach. *Journal of the Association for Information Science and Technology, 70*(10), 1047–1059. https://doi.org/10.1002/asi.24180

Arlinghaus, T., Kus, K., Kajüter, P., & Teuteberg, F. (2021). Datentreuhandstellen gestalten: Status quo und Perspektiven für Geschäftsmodelle. *HMD Praxis Der Wirtschaftsinformatik, 58*(3), 565–579. https://doi.org/10.1365/s40702-021-00727-x

Awada, R., & Kiringa, I. (2015). Data sharing and exchange: General data-mapping semantics. *International Conference BDAS 2015 Proceedings.*

Azkan, C., Möller, F., Meisel, L., & Otto, B. (2020). Service dominant logic perspective on data ecosystems—A case study based morphology. *ECIS 2020 Proceedings.*

Backhaus, K., Erichson, B., Plinke, W., & Weiber, R. (2006). *Multivariate Analysemethoden*. Springer, Berlin. https://doi.org/10.1007/978-3-642-14987-0

Bastiaansen, H. J. M., Kollenstart, M., Dalmolen, S., & van Engers, T. M. (2020). User-centric network-model for data control with interoperable legal data sharing artefacts. *PACIS 2020 Proceedings*.

Cattaneo, G., Micheletti, G., Glennon, M., La Croce, C., & Mitta, C. (2020). *The European data market monitoring tool: key facts & figures, first policy conclusions, data landscape and quantified stories*. European Commission Publications Office. https://data.europa.eu/doi/https://doi.org/10.2759/72084

Choi, S.-K. T., & Kröschel, I. (2015). Challenges of governing interorganizational value chains: Insights from a case study. *ECIS 2015 Proceedings.*

Cichy, P., Salge, T. O., & Kohli, R. (2021). Privacy concerns and data sharing in the internet of things: Mixed methods evidence from connected cars. *MIS Quarterly, 45*(4), 1863–1892. https://doi.org/10.25300/MISQ/2021/14165

Coelho, R., Braga, R., David, J. M. N., Dantas, M., Ströele, V., & Campos, F. (2021). Integrating blockchain for data sharing and collaboration support in scientific ecosystem platform. *HICSS 2021 Proceedings*.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Czech, A., Geenen, V., Breß, C., Popovski, M., Krauß, P., Riedel, T., & Gauterin, F. (2023). Designing a mobility data trustee (MDT): Findings from a multi-disciplinary analysis of requirements of an MDT. *CoRDI 2023 Proceedings*.

de Corbière, F. (2009). Data quality and interorganizational information systems: The role of electronic catalogues. *AMCIS 2009 Proceedings*.

de Corbière, F., & Rowe, F. (2011). Adoption factors of electronic data exchange and technology: Can we distinguish two phases?. *ECIS 2011 Proceedings*. https://aisel.aisnet.org/ecis2011/116

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), 2. https://doi.org/10.17705/1jais.00345
Dreller, A. (2018). Creating value from data sharing—Future-oriented business models in theory and practice. *Springer Vieweg Wiesbaden*. https://doi.org/10.1007/978-3-658-23276-4
Enders, T., Benz, C., Schüritz, R., & Lujan, P. (2020). How to implement an open data strategy? Analyzing organizational change processes to enable value creation by revealing data. *ECIS 2020 Proceedings*.
Enders, T., Satzger, G., Fassnacht, M. K., & Wolff, C. (2022). Why should i share? Exploring Benefits of Open Data for Private Sector Organizations. *PACIS 2022 Proceedings*.
Fassnacht, M., Benz, C., Heinz, D., Leimstoll, J. & Satzger G. (2023a). Barriers to data sharing among private sector organizations. *HICSS 2023 Proceedings*.
Fassnacht, M., Benz, C., Leimstoll, J. & Satzger G. (2023b). Is your organization ready to share? A framework of beneficial conditions for data sharing. *ICIS 2023 Proceedings*.
Fassnacht, M., Benz, C., Bode, J., Heinz, D. & Satzger, G. (2024). Systematizing data sharing practices: A taxonomy. *ECIS 2024 Proceedings*.
Gascó, 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. *HICSS 2018 Proceedings*.
Gelhaar, J., & Otto, B. (2020). Challenges in the emergence of data ecosystems. *PACIS 2020 Proceedings*.
Gelhaar, J., Groß, T., & Otto, B. (2021a). A taxonomy for data ecosystems. *HICSS 2021 Proceedings*.
Gelhaar, J., Gürpinar, T., Henke, M., & Otto, B. (2021b). Towards a taxonomy of incentive mechanisms for data sharing in data ecosystems. *PACIS 2021 Proceedings*.
Gieß, A., Möller, F., Schoormann, T., & Otto, B. (2023). Design Options for Data Spaces. *ECIS 2023 Proceedings*.
Glass, R., & Vessey, I. (1995). Contemporary application-domain taxonomies. *IEEE Software, 12*, 63–76. https://doi.org/10.1109/52.391837
Grace, K. (2020). Putting interoperability on health-information-systems' implementation agenda. *HICSS 2020 Proceedings.*
Gregor, S., & Hevner, A. (2013). Positioning and Presenting Design Science Research for Maximum Impact. *MIS Quarterly, 37*(2), 337–356. https://doi.org/10.25300/MISQ/2013/37.2.01.
Gregor, S. (2006). The nature of theory in information systems. *MIS Quarterly, 30*(3), 611–642. https://doi.org/10.2307/25148742
Hambrick, D. C. (1984). Taxonomic approaches to studying strategy: Some conceptual and methodological issues. *Journal of Management, 10*(1), 27–41. https://doi.org/10.1177/014920638401000104
Han, J., Kamber, M., & Pei, J. (2012). Data mining: Concepts and techniques (*3rd ed.).* Morgan Kaufmann Publishers. https://doi.org/10.1016/C2009-0-61819-5
Hardy, A. (1996). On the number of clusters. *Computational Statistics & Data Analysis, 23*(1), 83–96. https://doi.org/10.1016/S0167-9473(96)00022-9
Hartmann, P. M., Zaki, M., Feldmann, N., & Neely, A. (2016). Capturing value from big data—A taxonomy of data-driven business models used by start-up firms. *International Journal of Operations and Production Management, 36*(10), 1382–1406. https://doi.org/10.1108/IJOPM-02-2014-0098
Heinz, D., Benz, C., Fassnacht, M. & Satzger G. (2022). Past, present and future of data ecosystems research: A systematic literature review. *PACIS 2022 Proceedings*.
Holstein, J., Schemmer, M., Jakubik, J., Vössing, M., & Satzger, G. (2023). Sanitizing data for analysis: Designing systems for data understanding. *Electronic Markets, 33*, 52. https://doi.org/10.1007/s12525-023-00677-w
Hunke, F., Heinz, D., & Satzger, G. (2021). Creating customer value from data: foundations and archetypes of analytics-based services. *Electronic Markets, 32*(2), 503–521. https://doi.org/10.1007/s12525-021-00506-y
Jagals, M., & Karger, E. (2021). Inter-organizational data governance: A literature review. *ECIS 2021 Proceedings.*
Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management, 29*(4), 258–268. https://doi.org/10.1080/10580530.2012.716740
Jonas, C., Oberländer, A. M., Schmitt, K., & Wethmar, S. (2022). Demystifying industrial internet of things start-ups—A multi-layer taxonomy. *International Conference on Business Informatics 2022 Proceedings*.
Jussen, I., Schweihoff, J., Dahms, V., Möller, F., & Otto, B. (2023). Data sharing fundamentals: Characteristics and definition. *HICSS 2023 Proceedings*.
Jussen, I., Fassnacht, M., Schweihoff, J. & Möller F. (2024a). Reaching for the stars: Exploring value constellations in inter-organizational data sharing. *ECIS 2024 Proceedings*.
Jussen, I., Möller, F., Schweihoff, J., Gieß, A., Giussani, G., & Otto, B. (2024). Issues in inter-organizational data sharing: Findings from practice and research challenges. *Data & Knowledge Engineering, 150*, 102280. https://doi.org/10.1016/j.datak.2024.102280
Kajüter, P., Arlinghaus, T., Kus, K., & Teuteberg, F. (2022). Analysis of barriers to digital linking among healthcare stakeholders. *International Conference on Business Informatics 2022 Proceedings.*
Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster analysis. *Biometrics*. John Wiley & Sons, Inc. https://doi.org/10.2307/2532178
Kawashita, I., Baptista, A. A., & Soares, D. (2022). Open government data use by the public sector—An overview of its benefits, barriers, drivers, and enablers. *HICSS 2022 Proceedings.*
Kitchin, R. (2014). Big Data, New Epistemologies and Paradigm Shift. *Big Data & Society, 1*(1), 1–12. https://doi.org/10.1177/2053951714528481
Koshizuka, N. (2023). *DATA-EX and related activities in Japan*. Retrieved March 21, 2024, from https://filehub.admiralcloud.com/v5/deliverFile/fe360c55-77bd-45c7-a353-91c4b40da403
Krotova, A., Mertens, A., & Scheufen, M. (2020). *Open data and data sharing: An economic analysis*. IW-Policy Paper, No. 21/2020, German Economic Institute. Retrieved March 21, 2024, from https://hdl.handle.net/10419/225269
Kuk, G. (2011). The roles of agency and artifacts in assembling open data complementarities. *ICIS 2011 Proceedings.*
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*(4), 421–439. https://doi.org/10.1007/s12599-021-00723-x
Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174. https://doi.org/10.2307/2529310
Lauf, F., Scheider, S., Friese, J., Kilz, S., Radic, M., & Burmann, A. (2023). Exploring design characteristics of data trustees in healthcare—Taxonomy and archetypes. *ECIS 2023 Proceedings*.
Lawrenz, S., & Rausch, A. (2021). Don`t buy a pig in a poke a framework for checking consumer requirements in a data marketplace. *HICSS 2021 Proceedings*.
Liang, F., Yu, W., An, D., Yang, Q., Fu, X., & Zhao, W. (2018). A survey on big data market: Pricing, trading and protection. *IEEE Access, 6*, 15132–15154. https://doi.org/10.1109/ACCESS.2018.2806881
Lindner, S., Straub, S., & Kühne, B. (2021). *How to share data? Data sharing platforms for organizations*. Retrieved March 21, 2024, from https://www.digitale-technologien.de/DT/Redak

tion/EN/Downloads/Publikation/smartdata_%20Datasharing.pdf?__blob=publicationFile&v=19
Lis, D., & Otto, B. (2021). Towards a taxonomy of ecosystem data governance. *HICSS 2021 Proceedings*.
Liu, E., & Kumar, A. (2003). Leveraging Information Sharing to Increase Supply Chain Configurability. *ICIS 2003 Proceedings.*
Michener, C. D., & Sokal, R. R. (1957). A quantitative approach to a problem in classification. *Evolution, 11*(2), 130–162. https://doi.org/10.2307/2406046
Miller, D. (2018). Challenging trends in configuration research: Where are the configurations? *Strategic Organization, 16*(4), 453–469. https://doi.org/10.2307/26540168
Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika, 50*(2), 159–179. https://doi.org/10.1007/BF02294245
Mirkin, B. (2011). Choosing the number of clusters. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 1*(3), 252–260. https://doi.org/10.1002/WIDM.15
Möller, F., Bauhaus, H., Hoffmann, C., Niess, C., & Otto, B. (2019). Archetypes of digital business models in logistics start-ups. *ECIS 2019 Proceedings*.
Mollick, J. S. (2016). Privacy policies for E-commerce effects of information privacy policies on customers' willingness to transact with online vendors. *AMCIS 2016 Proceedings*.
Moore, G. C., & Benbasat, I. (1991). Development of an instrument to measure the perceptions of adopting an information technology innovation. *Information Systems Research, 2*(3), 192–222. https://doi.org/10.1287/ISRE.2.3.192
Müller, J., Veile, J., & Voigt, K.-I. (2020). Prerequisites and incentives for digital information sharing in Industry 4.0—An international comparison across data types. *Computers & Industrial Engineering, 148*(1), 106733. https://doi.org/10.1016/j.cie.2020.106733
Muschalle, A., Stahl, F., Löser, A., & Vossen, G. (2012). Pricing approaches for data markets. *International Workshop on Business Intelligence for the Real-Time Enterprise 2012 Proceedings*.
Nickerson, R. C., Varshney, U., Muntermann, J. & Kundisch, D. (2024). Hierarchical taxonomies in information systems. *AMCIS 2024 Proceedings*
Nickerson, R. C., Varshney, U., & Muntermann, J. (2013). A method for taxonomy development and its application in information systems. *European Journal of Information Systems, 22*(3), 336–359. https://doi.org/10.1057/ejis.2012.26
Nokkala, T., Salmela, H., & Toivonen, J. (2019). Data governance in digital platforms. *AMCIS 2019 Proceedings*.
O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. *International Journal of Qualitative Methods, 19*. https://doi.org/10.1177/1609406919899220
Oliveira, M. I. S., Lima Glória de Fátima Barros, & Lóscio, B. F. (2019). Investigations into Data Ecosystems: a systematic mapping study. *Knowledge and Information Systems, 61*(2), 589–630. https://doi.org/10.1007/s10115-018-1323-6
Opriel, S., Skubowius, E., & Lamberjohann, M. (2021). How usage control fosters willingness to share sensitive data in interorganizational processes of supply chains. *International Scientific Symposium on Logistics 2021 Proceedings*.
Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: findings from the International Data Spaces case. *Electronic Markets, 29*(4), 561–580. https://doi.org/10.1007/S12525-019-00362-x
Palinkas, L. A., Horwitz, S. M., Green, C. A., Wisdom, J. P., Duan, N., & Hoagwood, K. (2015). Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Administration and Policy in Mental Health and Mental Health Services Research, 42*(5), 533–544. https://doi.org/10.1007/s10488-013-0528-y
Reiberg, A., Niebel, C., & Kraemer, P. (2022). *What is a data space?*. Retrieved March 21, 2024, from https://gaia-x-hub.de/wp-content/uploads/2022/10/White_Paper_Definition_Dataspace_EN.pdf
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*(1), 4–29. https://doi.org/10.1007/s40319-018-00777-7
Rousseeuw, P. J. (1986). A visual display for hierarchical classification. In E. Diday, Y. Escoufier, L. Lebart, J. Pages, & Y. Schektman (Eds.), *Data Analysis and Informatics*, *4*, 743–748 North-Holland.
Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7
Rukanova, B., Tan, Y. H., Huiden, R., Ravulakollu, A., Grainger, A., & Heijmann, F. (2020). A framework for voluntary business-government information sharing. *Government Information Quarterly, 37*(4), 101501. https://doi.org/10.1016/J.GIQ.2020.101501
Rupasinghe, T., Burstein, F., & Rudolph, C. (2019). Blockchain based dynamic patient consent: A privacy-preserving data acquisition architecture for clinical data analytics. *ICIS 2019 Proceedings*.
Saldaña, J. (2015). *The coding manual for qualitative researchers* (3rd.). SAGE Publications.
Sayogo, D. S., & Pardo, T. A. (2011). Understanding the capabilities and critical success factors in collaborative data sharing network: The case of dataONE. *ACM International Conference 2011 Proceedings*.
Sayogo, D. S., & Pardo, T. A. (2012). Exploring the motive for data publication in open data initiative: Linking intention to action. *HICSS 2012 Proceedings*.
Schäffer, T., & Stelzer, D. (2017). Towards a taxonomy for coordinating quality of master data in product information sharing. *International Conference on Information Quality 2017 Proceedings*.
Schäffer, T., & Stelzer, D. (2018). Barriers to adopting data pools for product information sharing—A literature review. *International Conference on Business Informatics 2018 Proceedings.*
Scheider, S., Lauf, F., Geller, S., Möller, F., & Otto, B. (2023). Exploring design elements of personal data markets. *Electronic Markets, 33*, 28. https://doi.org/10.1007/s12525-023-00646-3
Schüritz, R., Seebacher, S., Satzger, G. & Schwarz, L. (2017). Datatization as the next frontier of servitization—Understanding the challenges for transforming organizations. *ICIS 2017 Proceedings*.
Schüritz, R., Farrell, K., Wixom, B. & Satzger, G. (2019). Value co-creation in data-driven services: Towards a deeper understanding of the joint sphere. *ICIS 2019 Proceedings*.
Schweihoff, J., Jussen, I., Dahms, V., Möller, F., & Otto, B. (2023a). How to share data online (fast)—A taxonomy of data sharing business models. *HICSS 2023 Proceedings*.
Schweihoff, J., Jussen, I., & Möller, F. (2023b). Trust me, I'm an intermediary! Exploring data intermediation services. *International Conference on Business Informatics 2023 Proceedings.*
Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. *Intereconomics, 54*(4), 208–216. https://doi.org/10.1007/S10272-019-0826-Z
Spiekermann, S., Acquisti, A., Böhme, R., & Hui, K. L. (2015). The challenges of personal data markets and privacy. *Electronic Markets, 25*(2), 161–167. https://doi.org/10.1007/S12525-015-0191-0

Sterk, F., Peukert, C., Hunke, F., & Weinhardt, C. (2022). Understanding car data monetization: A taxonomy of data-driven business models in the connected car domain. *International Conference on Business Informatics 2022 Proceedings.*

Sterk, F., Stocker, A., Heinz, D., & Weinhardt, C. (2024). Unlocking the value from car data: A taxonomy and archetypes of connected car business models. *Electronic Markets, 34*, 13. https://doi.org/10.1007/s12525-024-00692-5

Steudner, T., Widjaja, T., & Schumann, J. (2019). An exploratory study of risk perception for data disclosure to a network of firms. *International Conference on Business Informatics 2019 Proceedings.*

Strauss, T., & von Maltitz, M. J. (2017). Generalising Ward's method for use with Manhattan distances. *PLoS ONE, 12*(1), e0168288. https://doi.org/10.1371/journal.pone.0168288

Susha, I., Janssen, M., & Verhulst, S. G. (2017). Data collaboratives as a new frontier of cross-sector partnerships in the age of open data: Taxonomy development. *HICSS 2017 Proceedings.*

Susha, I., & Gil-Garcia, J. R. (2019). A collaborative governance approach to partnerships addressing public problems with private data. *HICSS 2019 Proceedings*.

Szopinski, D., Schoormann, T., & Kundisch, D. (2019). Because Your Taxonomy is Worth It: Towards a Framework for Taxonomy Evaluation. *ECIS 2019 Proceedings*.

Szopinski, D., Schoormann, T., & Kundisch, D. (2020). Criteria as a prelude for guiding taxonomy evaluation. *HICSS 2020 Proceedings.*

Treiblmaier, H., & Pollach, I. (2007). Users' perceptions of benefits and costs of personalization. *ICIS 2007 Proceedings.*

van den Broek, T., van Veenstra, A. F., (2015). Modes of governance in inter-organizational data Collaborations. *ECIS 2015 Proceedings.*

van de Ven, M., Abbas, A., Kwee, Z., & Reuver, M. (2021). Creating a taxonomy of business models for data marketplaces. *BLED eConference 2021 Proceedings.*

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

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association, 58*(301), 236–244. https://doi.org/10.1080/01621459.1963.10500845

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://www.misq.org/misreview/announce.html

Wilms, K. L., Brenger, B., & Rehwald, S. (2018). Open data in higher education—What prevents researchers from sharing research data? *ICIS 2018 Proceedings*.

Xiao, J., Xie, K., & Hu, Q. (2013). Inter-firm IT governance in power-imbalanced buyer-supplier dyads: Exploring how it works and why it lasts. *European Journal of Information Systems, 22*(5), 512–528. https://doi.org/10.1057/EJIS.2012.40

Zaheer, N., & Trkman, P. (2017). An information sharing theory perspective on willingness to share information in supply chains. *International Journal of Logistics Management, 28*(2), 417–443. https://doi.org/10.1108/IJLM-09-2015-0158

Zuiderwijk, A., Janssen, M., & Davis, C. (2014). Innovation with open data: Essential elements of open data ecosystems. *Information Polity, 19*(1–2), 17–33. https://doi.org/10.3233/IP-140329