What Constitutes a Data space? Conceptual Clarity beyond Technical Aspects

Completed Research Full Paper

Andreas Hutterer
Johannes Kepler University Linz, Linz, Austria
andreas.hutterer@jku.at

Barbara Krumay
Johannes Kepler University Linz, Linz, Austria
barbara.krumay@jku.at

Manuel Mühlburger
Johannes Kepler University Linz, Linz, Austria
manuel.muehlburger@jku.at

Abstract

In the data economy, data has become an essential strategic resource for gaining a competitive advantage. Data spaces represent a relatively new phenomenon aimed at encouraging businesses to fully leverage the potential of data. Despite various approaches for definitions there remains a lack of clarity surrounding the conceptualization of data space, its perceived value, and the factors that drive its adoption. The conceptual ambiguity and synonymous usage of the term in academic and business literature present significant obstacles to targeted conceptualization and use. This paper addresses these issues by proposing primary properties of data space and contributes to the field by applying a semantic decomposition. Through this approach, we identified data space as having the following conceptual aspects Nature, Element, Function, Utility and Governance. These primitives highlight the growing need for security and privacy when sharing interorganizational data. In addition, we offer an initial definition of data spaces.

Keywords

Data space, data ecosystem, semantic decomposition, concept, definition.

Introduction

Data space has become a widely used term to describe phenomena related to data sharing, data integration, and competitive advantage. In general, data spaces can facilitate data integration among different players whilst at the same time supporting organizations in their aim to create value. Thus, it is expected that data integration across different levels enables the leveraging of innovation potential (Gelhaar et al. 2021; Kuebler-Wachendorff et al. 2021). To enable a competitive advantage for participants, the development of dynamic digital ecosystems seems to be a possibility (Oliveira et al. 2019; Oliveira and Lóscio 2018). Moreover, by sharing data for the benefit of all in data spaces, organizations engage in data ecosystems (Otto et al. 2015). The development of the data space concept (Franklin et al. 2005) has traditionally been focused on extending its benefits beyond individual companies and positively impacting entire economies and societies (Capiello et al. 2020). Unfortunately, still organizations still face obstacles regarding data sharing (Fassnacht et al. 2023; Heinz et al. 2022; Kraemer et al. 2021) and refrain from participating in data spaces. This may be due to the fact that many platforms have faced failures in recent years (Özcan et al. 2022; Spiekermann 2019). Despite this reluctance in adopting data spaces, they have been identified as a solution for data sovereignty (Haße et al. 2020; Hutterer and Krumay 2022; Winter et al. 2022) and were identified as a promising approach for achieving an increase in global data availability (Schleimer et al. 2023). Especially in Europe, data spaces should exploit this potential and increase competitiveness

(European Commission 2020; Minghini et al. 2022). As a result, data spaces have gained more interest lately and have been identified as a research topic (Beverungen et al. 2022).

Current research suggests that data spaces can form the technical basis for data ecosystems (Hutterer and Krumay 2022), yet most studies predominantly addresses these technical aspects of data spaces (Curry et al. 2022). Information systems (IS) research, for example, addresses data spaces mainly as a way of integrating various data sources (Curry 2020; Curry et al. 2022). This focus on technical aspects seems to hinder the widespread adoption of data spaces (Otto and Jarke 2019). Furthermore, a clear definition has not evolved, since definitions of data spaces vary depending on the context – i.e, a technical, academic, or business point of view (Otto et al. 2022). This leads to a discontinuity between the existing academic body of knowledge about data spaces and their implementation in real-world scenarios (Otto et al. 2022). Thus, the authors have asserted that the idea of what a data space constitutes has emerged, but the term is rarely defined in a clear and holistic way (Strnadl and Schöning 2023). Previous endeavors in the field of IS research have produced various definitions that lack consistency (Otto et al. 2022). To sum up, the growing research on data spaces refers to various understandings and design options (Schleimer et al. 2023), whick lack a concise definition of the term. Thus, a common conceptual understanding of the structures, requirements, and implementations of data spaces is needed (Hirsch-Kreinsen et al. 2022). Therefore, this study aims at contributing to increasing clarity regarding what constitutes a data space in a more holistic way by answering the following research question: "What are the fundamental conceptual aspects of a data space?" To reach this goal, existing definitions have been used as the basis for a semantic decomposition process; the deducted semantic primitives were then used to develop a new definition. The present paper is structured as follows: Section two elucidates the current state-of-the-field in this context. In section three, we explicate our methodological approach, specifically the semantic decomposition. Section four offers a comprehensive analysis of the results. Next, we discuss these results and propose a conceptual definition for data spaces. Lastly, section five encompasses our conclusions, limitations, and future research.

State of the field

Challenged by the constantly increasing heterogeneity of data, the development of effective data management concepts has become inevitable (Kraemer et al. 2021). Whilst the growing significance of data as a strategic factor has further pushed the research on the concept of data spaces (Otto et al. 2015), no unified understanding of the concept has emerged. Data spaces in their most common understanding have been characterized as a set of related resources and interrelationships (Liu et al. 2010) or, more specifically, as digital infrastructures that allow users to access and analyze data from remote locations (Grossman and Mazzucco 2002). The objective of a data space has broadly been described as offering basic functionality for handling data sources regardless of their integration (Franklin et al. 2005), as managing data irrespective of its format or structure (Sun et al. 2010), and as managing data sources regardless of their structure or location in a virtual space (Tchuissang et al. 2010). Concurrently existing data space systems aim at processing data from diverse formats and systems (Wang et al. 2016), managing diverse data without assuming semantic integration, and coordinating their management in a cohesive manner (Singh 2013). Data spaces thus enable the concurrent use and management of multiple data sources (Halevy et al. 2006). Additionally, data spaces offer traditional data integration benefits with lower initial costs and incremental improvements, enabling a pay-as-you-go approach and allowing cost-effective, agile data integration (Hedeler et al. 2010). Data spaces have lately found more attention as a result of achieving data sovereignty in data-rich environments (Hutterer and Krumay 2022; Otto et al. 2022). However, due to the various architecture design options (Schleimer et al. 2023), there are different gradations, e.g., between the more central International Data Spaces Association (IDSA) framework (IDSA 2022) and the more decentralized Ocean Protocol (Ocean Protocol Foundation 2022), which is based on distributed ledger technology. Currently, only a few businesses have implemented data space applications, such as Catena-X (2023), Mobility Data Space (2023) or EuProGigant (2023). Further initiatives are aiming for a federation of data spaces, e.g., for mobility data (PrepDSpace4Mobility 2023).

As mentioned above, data spaces have mainly been investigated regarding their technical and structural aspects in addition to their implementation from a technical standpoint. From very early on, the term *data space* was used to describe a three-dimensional space that integrates digital connectivity (Imieliński and Goel 1999). This was even before the data space concept had been published (Franklin et al. 2005). Other attempts to describe and define a data space focus on data spaces from a structural point of view. They conceptualize data spaces as a collection of multimedia and diverse data sources (Tchuissang et al. 2012). Sometimes, the focus is more on the way data is retrieved. For example, a very technical approach is to describe and identify the efficiency of the query processing (Tchuissang et al. 2012). When it comes to specific implementations of data spaces, technical and structural aspects are even more frequently the focus. For example, the description of an industrial data space focuses on the implementation of virtual knowledge graphs (VKG) to represent heterogeneous data from multiple sources (Jiang et al. 2019). This specific representation of a data space is described as a broker for running cyber-physical-social production systems (CPSPS) in the context of manufacturing (Jiang et al. 2019). Another conceptualization of data spaces focuses on their ability to operate in real-time. They address and investigate the data transfer (via linked data and knowledge graphs), the real-time provision of data, and pricing models (Curry 2020; Curry et al. 2019). Other publications focus on the technical barriers to adopting data spaces (Sarma et al. 2009). Again, they adopt a technical approach to overcome these barriers by proposing a platform—data space support platforms (DSSP)—minimize the efforts (e.g., data integration) for participants by establishing mechanisms (e.g., self-starting) and services without human intervention (Sarma et al. 2009). They further address the evolution of the data space related to input from the data space (e.g., user feedback, additional data sources) to constitute a continuous integration process (Sarma et al. 2009). In addition, the existing literature also addresses differences between traditional data integration and a data space. For example, control over data and gradual integration is possible via data spaces (Sarma et al. 2009). Beyond technical aspects, some authors describe the term *data space* as the handling of knowledge schema and complex data in data management (Li et al. 2020), while others suggest that a data space refers to a group of participants and their relationships (Singh and Jain 2011). Decentralization, which has lately found more ground in the discussion (Beverungen et al. 2022), is often used to describe a data space as a decentralized infrastructure for trustworthy data sharing and exchange in data ecosystems based on commonly agreed principles (Nagel and Lycklama 2021). Again, this is a singular definition, which does not integrate other aspects. Even more, definitions of data spaces in a broader context, e.g., those related to data ecosystems, remain unclear (Nagel and Lycklama 2021). In sum, conceptual clarity of the term *data space* is missing. Although there seems to exist a common understanding about data spaces, to the best of our knowledge there is no structured approach to holistically describe what constitutes a data space beyond technical aspects. Even more, a clear and widely accepted definition is missing (Otto et al. 2022).

Methodological Approach

To effectively study the data space phenomenon (Sekaran and Bougie 2013) and contribute to more clarity regarding its conceptual nature, we utilized an inductive qualitative approach. We applied semantic decomposition to methodically deconstruct existing definitions into a sequence of constituent primitives (Akmajian et al. 2017) representing the central conceptual aspects of data spaces. Our research process involved a three-stage procedure: (1) deriving primitives based on definitions in existing literature, (2) evaluating literature-based primitives for their conceptual coverage of data space definitions (through non-academic sources) and perceptions (through interviews), and (3) establishing clarity and proposing a definition for data spaces. In stage one we aimed at identifying a broad set of data space conceptualizations for our semantic decomposition approach by applying a two-step process. First, we added the synopsis of 11 definitions presented by Curry et al. (2022) to our definition set. Second we analyzed each article in the results set of a recent literature review on data spaces (Hutterer and Krumay 2022) to derive additional data space definitions and conceptualizations. From the 59 papers analyzed in the second stage, seven presented an original conceptualization and definition, leading to a total set of 18 definitions available for semantic decompositioning. Selected parts with semantic meaning from the definitions were assessed based on their semantic content, resulting in the derivation of a first set of primitives. Subsequently, we compared these primitives across different definitions to identify the fundamental expressions related to data spaces. In the second stage, we conducted semi-structured interviews with 28 experts (E01–E28). The experts were selected based on a non-probability sampling strategy (Sekaran and Bougie 2013) often used when the phenomenon is rather new and the number of experts available is low (Etikan et al. 2016). The experts came from the German-speaking D-A-CH region. They were selected on the basis of their project experience with data spaces. The interviews were conducted in an online format from October to December 2022 and lasted between 28 and 113 minutes (average 47 minutes); all interviews were recorded and transcribed. All interview sections containing information regarding the conceptual understanding of data spaces were coded utilizing the literature-based set of conceptual primitives to evaluate and expand our literature-based understanding. Furthermore, we included six non-academic sources, which are among the most well-known sources regarding data spaces (Bitkom 2022; DIO 2021; European Commission 2023; GAIA-X Hub Austria 2023; GAIA-X Hub Germany 2022; IDSA 2022) in this process.

Results

Based on the semantic decomposition applied to the selected IS literature, we identified four primitives that comprise the primary properties of established definitions of data spaces: (1) *Nature*, (2) *Element*, (3) *Function*, and (4) *Utility*. The fundamental characterization of a data space is based on its physical or logical appearance, which is defined in the primitive *Nature*. The primitive *Element* defines the structure that constitutes the data space, along with the characteristics of its elements. The primitive *Function* explains the system-related aspect of the data space, including the functional relationships between the elements and their respective services. Finally, the primitive *Utility* pertains to the practical application and value of the data space implementation. Based on the expert interviews, we identified an additional primitive (5), *Governance*, that outlines the rules and regulations that oversee the management of a data space. The data were used to further specify the primitives in terms of expressions (see Table 1).

Primitive	Expression	Occurrence
Nature (Grossman and Mazzucco 2002; Halevy et al. 2006; Imieliński and Goel 1999; Jiang et al. 2019; Li et al. 2020; Nagel and Lycklama 2021; Sarma et al. 2009; Scerri et al. 2020; Singh 2013; Tchuissang et al. 2010, 2012; Wang et al. 2016)	Physical	E01, E03, E05, E07, E09, E10, E12, E15, E18, E19, E20, E21, E24, Bitkom (2022), DIO (2021), European Commission (2023), GAIA-X Hub Germany (2022)
	Logical	E04, E06, E09, E14, E18, E19, E22, E23, E28, GAIA-X Hub Austria (2023), IDSA (2022)
Element (Grossman and Mazzucco 2002; Imieliński and Goel 1999; Jiang et al. 2019; Liu et al. 2010; Nagel and Lycklama 2021; Scerri et al. 2020; Singh 2013; Singh and Jain 2011; Tchuissang et al. 2012; Wang et al. 2016)	Structure	E01, E02, E03, E06, E08, E10, E12, E15, E16, E18, E19, E25, E28, DIO (2021), European Commission (2023), GAIA-X Hub Austria (2023)
	Characteristic	E01, E02, E09, E10, E13, E17, E19, E21, E24, E25, E28, GAIA-X Hub Austria (2023), GAIA-X Hub Germany (2022)
Function (Curry 2020; Curry et al. 2019; Imieliński and Goel 1999; Li et al. 2020; Liu et al. 2010; Nagel and Lycklama 2021; Sarma et al. 2009; Scerri et al. 2020; Singh 2013; Tchuissang et al. 2010, 2012; Wang et al. 2016)	System	E01, E02, E03, E07, E09, E10, E11, E13, E16, E18, E19, E25, DIO (2021)
	Service	E02, E03, E09, E10, E13, E28, European Commission (2023)
Utility (Curry 2020; Curry et al. 2019; Grossman and Mazzucco 2002; Halevy et al. 2006; Hedeler et al. 2010, 2011; Imieliński and Goel 1999; Jiang et al. 2019; Nagel and Lycklama 2021; Sarma et al. 2009; Singh 2013; Sun et al. 2010; Tchuissang et al. 2010; Wang et al. 2016)	Form	E01, E04, E05, E06, E09, E12, E14, E15, E17, E18, E19, E20, E25, E26, E28, Bitkom (2022), DIO (2021), European Commission (2023), GAIA-X Hub Austria (2023), IDSA (2022)
	Context	E01, E04, E05, E06, E09, E12, E14, E15, E16, E17, E19, E20, E25, E26, E28, DIO (2021), GAIA-X Hub Austria (2023), Hub Germany (2022), IDSA (2022)
Governance	Principle	E01, E02, E03, E05, E09, E10, E11, E13, E14, E15, E16, E17, E19, E20, E22, E26, Bitkom (2022), Hub Germany (2022), IDSA (2022)
	Operation	E01, E02, E03, E04, E06, E09, E15, E19, E21

Table 1: Primitives, Expressions, and their Sources

Nature

The primitive Nature is what a data space is considered as. Widely, it has been described as a system (Sarma et al. 2009; Singh 2013; Wang et al. 2016). The primitive Nature was derived from all data sources (i.e., academic literature samples, interviews, and non-academic literature samples). It is expressed in the data in the sense of a logical and/or physical system. The physical expression of the nature of a data space refers to the tangible components within the data space and has been described for example as a "3-dimensional physical space" (Imieliński and Goel 1999). Experts emphasized the importance of ensuring a physical infrastructure. One expert referred to a "data space as…a technical instrument and vehicle" (E04). Another expert described a data space "as a connection of connectors" (E22). Two experts (E09, E19) as well as the non-academic literature (DIO 2021; GAIA-X Hub Germany 2022) focused on a federated, decentralized data infrastructure. On the other hand, the logical expression of the primitive Nature of a data space deals with the semantics of the components. Examples in the academic literature are handling (Li et al. 2020) or approach (Halevy et al. 2006). Experts have also emphasized the need to manage the elements within a data space effectively and comprehend its logical nature. In line with this, one expert proposed that "a data space is a federated environment" (E01), and others expressed the same opinion (E06, E15, E28). Still others saw it as an opportunity (E14, E18), construct (E05, E21), platform (E03), or marketplace (E07).

Element

The primitive Element describes what a data space consists of. It was derived since most of the definitions outline a certain set of structural elements, while some of them further describe their characteristics. The primitive Element is reflected for instance by defining a data space as a "set of participants and a set of relationships among them" (Singh and Jain 2011). Furthermore, it was reflected in definitions as "data models, datasets, ontologies, [and] data sharing contracts" (Scerri et al. 2020), data sources (Singh and Jain 2011), resources (Liu et al. 2010), objects (Imieliński and Goel 1999), or virtual knowledge graphs (Jiang et al. 2019), to name just some. Experts (E01, E02, E10, E19, E24, E28) mentioned mainly the different actors, data sets (E25), and the data shared (E03, E12, E15). The primitive Element is expressed in the data in the sense of a structure of elements or their characteristics. Structure is represented in the data in terms of structural elements such as identities (E09), servers (E13), central bodies (E28), and services in general (E02). Characteristics are expressed in connection to certain elements-e.g., decentralized (E01, E09, E19, E24, Nagel and Lycklama 2021), distributed (E02, Singh 2013), heterogenous (E24, Tchuissang et al. 2012), or mobile (Imieliński and Goel 1999)–or to a sum of structural elements–e.g., decentralized infrastructure (E09, Nagel and Lycklama 2021). Characteristics in the non-academic literature are represented in terms of certification (IDSA 2022), federation (GAIA-X Hub Germany 2022), or openness (GAIA-X Hub Germany 2022). While characteristics and structure are expressed in the definitions, they are not homogeneous. This is also reflected by the experts, who referred to a characteristic in terms of a closed system where data is made available in a confined space (E28). In general, the experts described characteristics in terms of decentralization, federation (E01, E09, E13, E18, E19, E21, E24), and distribution (E13).

Function

The primitive Function describes the range of capabilities a data space may perform. It was derived from the definitions since most of them describe certain functions like the data management function (Li et al. 2020; Tchuissang et al. 2010), semantic integration function (Liu et al. 2010; Singh 2013; Wang et al. 2016) or processing capabilities (Curry 2020; Curry et al. 2019). The primitive Function is expressed in the data in the sense of system or service functions. In the academic literature, the expression system of the primitive Function is represented in terms of data management (Li et al. 2020; Tchuissang et al. 2010), principles (Nagel and Lycklama 2021), semantic integration (Liu et al. 2010; Singh 2013; Wang et al. 2016), the pay-as-you-go approach (Sarma et al. 2009) and soft competencies (Scerri et al. 2020). The experts mentioned identity and access management (E02, E03, E10, E16, E19), data catalogues (E02, E13), and search possibilities (E02). In the non-academic literature, basic functionalities (Bitkom 2022), sovereignty (DIO 2021), and interoperability (European Commission 2023) represent the expression system of the primitive Function. On the other hand, the service expression of the primitive Function of data spaces refers to specific services that can be performed within the data space, represented in terms of user feedback (Sarma et al. 2009), processing capabilities (Curry 2020; Curry et al. 2019), or process queries (Tchuissang et al. 2012). Experts remain rather unclear, although they mentioned advanced services in general (E02, E09, E10). In the same way, the non-academic literature remains rather unclear in this regard.

Utility

The primitive Utility describes how data spaces can be employed by actors. It was derived from the definitions as most of them address the utility of the data space, such as its handling of resources in various use-cases (Grossman and Mazzucco 2002; Halevy et al. 2006; Hedeler et al. 2010, 2011; Nagel and Lycklama 2021; Singh 2013; Sun et al. 2010; Wang et al. 2016). The primitive Utility is expressed in the data in the sense of form and context. The expression form of the primitive Utility includes concepts like userless system management (Sarma et al. 2009), transparency of data (Jiang et al. 2019), and enhanced connectivity (Imieliński and Goel 1999). Experts mentioned data exchange (E01, E02, E04, E14, E17, E22, E27), restricted access (E01), transparency (E23) and guaranteed sovereignty (E26) as forms of utility. The latter depends on context, i.e., conditions of the data space. Context is represented in terms of ways to collect data (Curry 2020; Curry et al. 2019; Jiang et al. 2019; Singh 2013; Sun et al. 2010; Wang et al. 2016), which are constituted as data ecosystems (Nagel and Lycklama 2021), or the existence of unintegrated data sources (Halevy et al. 2006; Tchuissang et al. 2010). Experts also addressed the specific context in terms of the purpose of the data collection (E13), technical requirements (E18), and the federation among trustworthy partners (E20, E28). In the non-academic literature, context has been addressed in a rather nonspecific way (DIO 2021; GAIA-X Hub Austria 2023; GAIA-X Hub Germany 2022; IDSA 2022).

Governance

The primitive Governance constitutes the rules and regulations for the management of a data space, expressed in terms of principles and operation. This primitive was only weakly represented in the definitions from the academic literature, yet it was evident in the expert interviews. The primitive Governance is reflected in terms of compliance and general rules (E02, E05, E09, E17, E20) as well as ”transparent and fair mechanisms” (E19). The expression principle of the primitive Governance describes general governing rule sets. The experts mentioned sovereignty preserving regulations (E04, E19, E21), authentication and identification rules (E01, E09), confidentiality and integrity measures (E03, E10, E14, E16), and regulations to ensure a trusted environment” (E15). The operation expression of the Governance primitive refers to specific rules and regulations related to a specific operation (e.g., data sharing) within the data space. Experts mentioned terms and conditions (E02, E03, E06, E15) of the data space but also ethical issues (E01). In addition, experts mentioned the possibility for authorized actors to act freely within governing boundaries (E01, E09).

Upon analyzing the perspectives of data spaces, it became apparent that there are differences in the definitions of data spaces in terms of their scope. As a result of the semantic decomposition, we can provide conceptual clarity of the term *data space*. Our semantic decomposition indicates that these primitives capture the conceptual understanding of data spaces in modern contexts. Based on our results, we propose a conceptual definition for the concept of data space, considering the derived primitives as follows: ***A data space is a system of physical and/or logical nature consisting of elements and functions for providing a certain utility while at the same time relying on appropriate governance.***

Discussion

The goal of this study was to increase clarity regarding what constitutes a data space in a more holistic way that goes beyond technical aspects. We applied semantic decomposition on various data sources (IS literature, interviews, non-academic literature) and were able to identify five primitives (Nature, Element, Function, Utility, Governance) and two expressions per primitive. The primitives, alongside their expressions, reflect a conceptual structure that encapsulates existing aspects of data space definitions in a systematic manner. The primitive Governance evolved mainly from the expert interviews and the non-academic literature and was scarcely reflected in the academic literature in our sample. Although the term *data space* seems to be used widely in academia and business, the understanding of what constitutes a data space differs greatly. Thus, with our study, we contribute to the current body of knowledge by describing these primitives and their expressions. In addition, the study may contribute to a more unified understanding between academia and business. In the discussion, we mainly focus on the relationships among the primitives, but also propose a more holistic conceptual definition of the concept of data space.

The primitives derived (i.e., Nature, Element, Function, Utility, Governance) show certain dependencies and relationships. The underlying Nature of a data space is that of a system or environment with specific boundaries and interfaces. As described above, Nature has two expressions: physical and logical. These two expressions are not mutually exclusive but exist at the same time. This is consistent with the literature from different fields, which identifies the physical and logical aspects of a system, e.g., a network (Hou et al. 2009). Looking at the physical Nature means focusing on the tangible components such as servers, physical networks, and network components. Looking at the logical Nature of the data space does not neglect the physical Nature but focuses on the semantics of the components. Therefore, the term *federated* may refer to both–the physical infrastructure and the logical semantics behind this. Interestingly, the academic literature often adopted a rather narrow point of view, whereas the experts described the primitive Nature and its expressions more holistically in terms of addressing both–physical and logical Nature. Clearly, the primitive Nature determines the primitive Element. For example, the logical Nature of a marketplace determines Elements such as sellers and buyers. However, all Elements reflect structure and characteristics. The primitive Function is expressed in terms of system and service. It can also be seen as system-inherent capabilities compared to additional services. Again, the primitive Function is mainly influenced by the primitive Nature and the system property of a data space. The physical Nature, for example, influences the system-inherent Function (e.g., processing capabilities). The primitive Utility is the utilization of resources, e.g., resources to guarantee sovereignty (expression form) in accordance with the way to collect data (expression context). Finally, the primitive Governance reflects rules and regulations necessary to govern a data space. Governance is essential to ensuring data sovereignty (Jarke et al. 2019). It is necessary to establish trust in and reliability of the data space (Nagel and Lycklama 2021). The primitive Governance also reflects current discussions regarding legal and even ethical issues (E01), but also privacy and security aspects (Hernandez et al. 2022).

Notably, there is a misfit between the understanding of data spaces in academia and business. Based on the primitives and their expressions described above, we hope to answer the question of what constitutes a data space. We assume that the conceptual definition will also help to improve clarity among those who want to use or provide a data space–although on a more abstract level–. For businesses or individuals who want to benefit from participating in a data space, the primitives can be used to assess the properties of a specific data space. For businesses or organizations that want to provide data spaces, the primitives may function as a starting point for the development of the data spaces. And finally, for policymakers who are already engaged in the idea of a data space for many, it might be helpful to describe the Nature, Element, Function, Utility, and Governance primitives in the respective data space in an understandable way.

Conclusion

Although widely used in academia and business, the concept of data space lacks a clear definition. In our study, we applied semantic decomposition on current interpretations of the concept of data space from various sources. This resulted in five primitives alongside with ten expressions to describe what constitutes a data space. By explaining their relationships and proposing a more holistic definition beyond technical aspects, we lay the ground for further research, but also contribute to a better understanding for researchers and businesses alike. In this regard, we provide a fundament for further research not only on the concept of data space, but also data spaces adoption and participation. Since our semantic decomposition research design was based on only a small sample, a more in-depth analysis on the expressions and their representations was not possible. As a next step, it would make sense to analyze these underlying representations to better understand the primitives in full. First, our results can be a starting point for the development of a data space taxonomy. Future research could develop specific data space definitions for concrete implementation strategies. Second, although there are no existing theories on why organizations choose not to adopt data spaces, the lack of governance mechanisms seems to have an impact on such decisions. For this reason, our conceptual definition can also serve as a framework for research focusing on data spaces in a more application-oriented way. Clearly, this study also faces some limitations due to the small sample size and the fact that the assessments we conducted are solely based on qualitative data. In order to assess the adequacy of the presented conceptual definition, we invite researchers to extend our investigation to obtain a more comprehensive understanding of expectations concerning data spaces. However, the fact that our study focused solely on a specific region limits the generalizability of our findings. Furthermore, future research should address the realization of data spaces and the challenges involved in designing them by identifying potential pathways. It is important to note that our paper is focused on the conceptual aspects that characterize dataspaces and does not detail regulatory aspects or economic considerations that might influence their adoption.

REFERENCES

Akmajian, A., Farmer, A. K., Demers, R. A., and Harnish, R. M. (eds.). 2017. Linguistics: An Introduction to Language and Communication, (7th ed.), Cambridge, Mass: MIT Press.

Beverungen, D., Hess, T., Köster, A., and Lehrer, C. 2022. "From Private Digital Platforms to Public Data Spaces: Implications for the Digital Transformation," Electronic Markets (32:2), Springer, pp. 493–501.

Bitkom. 2022. Bitkom Position Paper EU Data Act Proposal, Bitkom e.V. (https://www.bitkom.org/sites/main/files/2022-07/Bitkom_Data_Act_Position_Paper_July22_Cloud_13.07.22.pdf).

Capiello, C., Gal, A., Jarke, M., and Rehof, J. 2020. Data Ecosystems: Sovereign Data Exchange among Organizations (Dagstuhl Seminar 19391), Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik GmbH, Wadern/Saarbruecken, Germany, 69 pages. (https://doi.org/10.4230/DAGREP.9.9.66).

Catena-X. 2023. "Automotive Network." (https://catena-x.net/en/, accessed April 19, 2023).

Curry, E. 2020. "Dataspaces: Fundamentals, Principles, and Techniques," in Real-Time Linked Dataspaces, Cham: Springer International Publishing, pp. 45–62. (https://doi.org/10.1007/978-3-030-29665-0_3).

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

Curry, E., Scerri, S., and Tuikka, T. 2022. Data Spaces: Design, Deployment and Future Directions, Springer Nature.

DIO. 2021. Datenräume und Datenkreise - Im Fokus der Data Intelligence, Data Intelligence Offensive. (https://www.dataintelligence.at/wp-content/uploads/2021/10/DIO_Whitepaper_Datenraeume-und-Datenkreise_2021.pdf).

Etikan, I., Musa, S. A., Alkassim, R. S., and others. 2016. "Comparison of Convenience Sampling and Purposive Sampling," American Journal of Theoretical and Applied Statistics (5:1), New York.

EuProGigant. 2023. "The Lighthouse Project for Gaia-X in Manufacturing." (https://euprogigant.com/en/, accessed April 19, 2023).

European Commission. 2020. A European Strategy for Data. (https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0066&from=EN).

European Commission. 2023. REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on Harmonised Rules on Fair Access to and Use of Data (Data Act), European Commission. (https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52022SC0034).

Fassnacht, M., Benz, C., Heinz, D., Leimstoll, J., and Satzger, G. 2023. "Barriers to Data Sharing among Private Sector Organizations," in Hawaii International Conference on Systems Sciences (HICCS).

Franklin, M., Halevy, A., and Maier, D. 2005. "From Databases to Dataspaces: A New Abstraction for Information Management," ACM SIGMOD Record (34:4), pp. 27–33. (https://doi.org/10.1145/1107499.1107502).

GAIA-X Hub Austria. 2023. "Begrifflichkeiten." (https://www.gaia-x.at/begrifflichkeiten/, accessed April 19, 2023).

GAIA-X Hub Germany. 2022. "Was Ist Ein Datenraum?," Gaia-X Hub Germany c/o Acatech – Deutsche Akademie Der Technikwissenschaften e.V. (https://gaia-x-hub.de/wp-content/uploads/2022/10/20220914_White_Paper_22.1_Definition_Datenraum_final.pdf).

Gelhaar, J., Gürpinar, T., Henke, M., and Otto, B. 2021. "Towards a Taxonomy of Incentive Mechanisms for Data Sharing in Data Ecosystems," Twenty-Fifth Pacific Asia Conference on Information Systems (PACIS).

Grossman, R., and Mazzucco, M. 2002. "DataSpace: A Data Web for the Exploratory Analysis and Mining of Data," Computing in Science & Engineering (4:4), IEEE, pp. 44–51.

Halevy, A., Franklin, M., and Maier, D. 2006. "Principles of Dataspace Systems," in Proceedings of the Twenty-Fifth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems.

Haße, H., van der Valk, H., Weißenberg, N., and Otto, B. 2020. "Shared Digital Twins: Data Sovereignty in Logistics Networks," Proceedings of the Hamburg International Conference of Logistics (HICL) (29), epubli. (https://doi.org/10.15480/882.3119).

Hedeler, C., Belhajjame, K., Paton, N. W., Campi, A., Fernandes, A. A., and Embury, S. M. 2010. "Dataspaces," in Search Computing, Springer, pp. 114–134.

Hedeler, C., Belhajjame, K., Paton, N. W., Fernandes, A. A., Embury, S. M., Mao, L., and Guo, C. 2011. "Pay-as-You-Go Mapping Selection in Dataspaces," in Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, pp. 1279–1282.

Heinz, D., Benz, C., Fassnacht, M., and Satzger, G. 2022. "Past, Present and Future of Data Ecosystems Research: A Systematic Literature Review," Pacific Asia Conference on Information Systems (PACIS).

Hernandez, J., McKenna, L., and Brennan, R. 2022. "TIKD: A Trusted Integrated Knowledge Dataspace for Sensitive Data Sharing and Collaboration," in Data Spaces: Design, Deployment and Future Directions, Springer International Publishing Cham, pp. 265–291.

Hirsch-Kreinsen, H., Kubach, U., Stark, R., von Wichert, G., Litsche, S., Sedlmeir, J., and Steglich, S. 2022. "Themenfelder Industrie 4.0 - Forschungs- und Entwicklungsbedarfe für die erfolgreiche Umsetzung von Industrie 4.0: Aktualisierte Fassung," Application/pdf, Forschungsbeirat der Plattform Industrie 4.0 / acatech – Deutsche Akademie der Technikwissenschaften, 42 pages. (https://doi.org/10.48669/FB40_2022-04).

Hou, Y., Zafer, M., Lee, K., Verma, D., and Leung, K. K. 2009. "On the Mapping between Logical and Physical Topologies," in 2009 First International Communication Systems and Networks and Workshops, IEEE, pp. 1–10.

Hutterer, A., and Krumay, B. 2022. "Integrating Heterogeneous Data in Dataspaces - A Systematic Mapping Study," in PACIS 2022 Proceedings. (https://aisel.aisnet.org/pacis2022/222).

IDSA. 2022. "Reference Architecture Model, Version 4.0." (https://docs.internationaldataspaces.org/ids-ram-4/, accessed April 19, 2023).

Imieliński, T., and Goel, S. 1999. "DataSpace—Querying and Monitoring Deeply Networked Collections in Physical Space," in Proceedings of the 1st ACM International Workshop on Data Engineering for Wireless and Mobile Access, Seattle Washington USA: ACM, August, pp. 44–51. (https://doi.org/10.1145/313300.313341).

Jarke, M., Otto, B., and Ram, S. 2019. "Data Sovereignty and Data Space Ecosystems," Business & Information Systems Engineering, Springer, pp. 549–550.

Jiang, P., Liu, C., Li, P., and Shi, H. 2019. "Industrial Dataspace: A Broker to Run Cyber-Physical-Social Production System in Level of Machining Workshops," in 2019 IEEE 15th International Conference on Automation Science and Engineering (CASE), IEEE, pp. 1402–1407.

Kraemer, J., Stüdlein, N., and Zierke, O. 2021. "Data as a Public Good: Experimental Insights on the Optimal Design of B2b Data Sharing Platforms," Available at SSRN 3970821.

Kuebler-Wachendorff, S., Luzsa, R., Kranz, J., Mager, S., Syrmoudis, E., Mayr, S., and Grossklags, J. 2021. "The Right to Data Portability: Conception, Status Quo, and Future Directions," Informatik Spektrum (44:4), pp. 264–272. (https://doi.org/10.1007/s00287-021-01372-w).

Li, P., Cheng, K., Jiang, P., and Katchasuwanmanee, K. 2020. "Investigation on Industrial Dataspace for Advanced Machining Workshops: Enabling Machining Operations Control with Domain Knowledge and Application Case Studies," Journal of Intelligent Manufacturing, Springer.

Liu, D., Yang, D., Nie, T., Kou, Y., and Shen, D. 2010. "Document Clustering in Personal Dataspace," in 2010 Seventh Web Information Systems and Applications Conference, Huhehot, China: IEEE, August, pp. 9–12. (https://doi.org/10.1109/WISA.2010.16).

Minghini, M., Kotsev, A., and Granell, C. 2022. "A European Approach to the Establishment of Data Spaces," Data (7:8), p. 118. (https://doi.org/10.3390/data7080118).

Mobility Data Space. 2023. "The Data Sharing Community." (https://mobility-dataspace.eu/, accessed April 19, 2023).

Nagel, L., and Lycklama, D. 2021. "Design Principles for Data Spaces," Position Paper, Berlin: International Data Spaces Association, July 15, p. 111. (https://doi.org/10.5281/ZENODO.5105744).

Ocean Protocol Foundation. 2022. Ocean Protocol: Tools for the Web3 Data Economy, Ocean Protocol Foundation. (https://oceanprotocol.com/tech-whitepaper.pdf).

Oliveira, M. I. S., Barros Lima, G. de F., and Farias Lóscio, B. 2019. "Investigations into Data Ecosystems: A Systematic Mapping Study," Knowledge and Information Systems (61:2), pp. 589–630. (https://doi.org/10.1007/s10115-018-1323-6).

Oliveira, M. I. S., and Lóscio, B. F. 2018. "What Is a Data Ecosystem?," in Proceedings of the 19th Annual International Conference on Digital Government Research: Governance in the Data Age, Delft The Netherlands: Association for Computing Machinery, May 30, pp. 1–9. (https://doi.org/10.1145/3209281.3209335).

Otto, B., Bärenfänger, R., and Steinbuß, S. 2015. "Digital Business Engineering: Methodological Foundations and First Experiences from the Field," in BLED 2015 Proceedings. 13., Bled. (https://aisel.aisnet.org/bled2015/13).

Otto, B., ten Hompel, M., and Wrobel, S. 2022. Designing Data Spaces: The Ecosystem Approach to Competitive Advantage, Springer Nature.

Otto, B., and Jarke, M. 2019. "Designing a Multi-Sided Data Platform: Findings from the International Data Spaces Case," Electronic Markets (29:4), Springer, pp. 561–580.

Özcan, L., Koldewey, C., Duparc, E., van der Valk, H., Otto, B., and Dumitrescu, R. 2022. "Why Do Digital Platforms Succeed or Fail? - A Literature Review on Success and Failure Factors," in AMCIS 2022 Proceedings. 15. (https://aisel.aisnet.org/amcis2022/sig_dite/sig_dite/15).

PrepDSpace4Mobility. 2023. "Laying the Foundation for a Common European Mobility Data Space." (https://mobilitydataspace-csa.eu/, accessed April 19, 2023).

Sarma, A. D., Dong, X. L., and Halevy, A. Y. 2009. "Data Modeling in Dataspace Support Platforms," in Conceptual Modeling: Foundations and Applications, Springer, pp. 122–138.

Scerri, S., Tuikka, T., and Lopez de Vallejoan, I. 2020. Towards a European-Governed Data Sharing Space, Big Data Value Association. (https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf).

Schleimer, A. M., Fraunhofer, I., Jahnke, N., and Otto, B. 2023. "Architecture Design Options for Federated Data Spaces," in Proceedings of the 56th Hawaii International Conference on System Sciences. (https://hdl.handle.net/10125/103078).

Sekaran, U., and Bougie, R. 2013. Research Methods for Business: A Skill-Building Approach, (6th ed.), Chichester, West Sussex: Wiley & Sons.

Singh, M. 2013. "A Framework for Data Modeling and Querying Dataspace Systems," in Seventh International Conference on Data Mining and Warehousing (ICDMW).

Singh, M., and Jain, S. K. 2011. "A Survey on Dataspace," in Advances in Network Security and Applications (Vol. 196), Communications in Computer and Information Science, D. C. Wyld, M. Wozniak, N. Chaki, N. Meghanathan, and D. Nagamalai (eds.), Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 608–621. (https://doi.org/10.1007/978-3-642-22540-6_59).

Spiekermann, M. 2019. "Data Marketplaces: Trends and Monetisation of Data Goods," Intereconomics (54:4), Springer, pp. 208–216.

Strnadl, C., and Schöning, H. 2023. "Data Platforms, Data Spaces, and (Data-)Ecosystems [Manuscript in Preparation]," in Data Governance, Weber (Ed.); Springer.

Sun, D., Ma, A., Zhang, B., Gao, K., and Zhang, Y. 2010. "Metadata Matching Based on Bayesian Network in DataSpace," in 2010 International Conference On Computer Design and Applications (Vol. 5), IEEE, pp. V5-358.

Tchuissang, G. N. F., Wang, N., Kuicheu, N. C., Siewe, F., and Xu, D. 2012. "2PROM: A Two-Phase Image Retrieval Optimization on Dataspace Using Predictive Modeling," in 2012 IEEE 11th International Conference on Signal Processing (Vol. 2), IEEE, pp. 1176–1179.

Tchuissang, G. N. F., Wang, N., Kuicheu, N. C., Siewe, F., Xu, D., and Liu, S. 2010. "Predicting Dataspace Retrieval Using Probabilistic Hidden Information," IEICE TRANSACTIONS on Information and Systems (93:7), The Institute of Electronics, Information and Communication Engineers.

Wang, Y., Song, S., and Chen, L. 2016. "A Survey on Accessing Dataspaces," ACM SIGMOD Record (45:2), ACM New York, NY, USA, pp. 33–44.

Winter, J., Frey, A., and Biehler, J. 2022. "Towards the Next Decade of Industrie 4.0 – Current State in Research and Adoption and Promising Development Paths from a German Perspective," Sci (4:3), p. 31. (https://doi.org/10.3390/sci4030031).