



# A taxonomy of data governance decision domains in data marketplaces

Rene Abraham<sup>1</sup> · Johannes Schneider<sup>1</sup> · Jan vom Brocke<sup>1</sup>

Received: 1 July 2022 / Accepted: 7 February 2023 / Published online: 22 May 2023  
© The Author(s) 2023

## Abstract

Commercializing data and data-related services has gained in importance in recent years. Driven by digitalization and the Internet-of-Things (IoT), companies and individuals continuously generate vast amounts of data. Data marketplaces have emerged to support these data providers in selling their data to different data consumers. However, data marketplaces face challenges in different data governance decision domains that inhibit their adoption. To get a better understanding of how data marketplaces counteract these challenges, this paper develops a taxonomy of data governance decision domains in data marketplaces. We used a taxonomy development method to inspect 13 data marketplaces from eight countries. The resulting taxonomy shows an overview of mechanisms concerning data quality, data security, data architecture, metadata, data lifecycle, data storage, and data pricing. We discuss common instantiation patterns, highlight gaps, and propose possible solutions. The taxonomy sets a foundation for further research and theory-building on data marketplaces. Practitioners can use the taxonomy to develop customized data governance strategies for data marketplaces.

**Keywords** Data governance · Data marketplace · Decision domain · Taxonomy

**JEL classification** M19

## Introduction

Digitalization profoundly transforms business today. For some time, digitalization focused on the automation of back-office and customer-oriented processes (Urbach, et al., 2019). Today, digitalization increasingly encompasses the collection and usage of data, adding to global data growth (IDC, 2018). Monitoring and collecting data from Internet-of-Things (IoT) sensors, mobile devices, and business processes not only facilitate companies in creating innovative products and improving customer experience. These activities also open

new ways of monetizing data as a product itself. According to Forrester Research (2018), 48 percent of global data and analytics decision-makers commercialize the data they own by either selling it or the derived insights. In doing so, data becomes part of a company's revenue model. Simultaneously, the demand for access to external data sources is growing (Forrester, 2018). Companies need diverse datasets to train machine learning models and enable improved decision-making. For example, sharing data between manufacturers can help to improve predictive maintenance algorithms and therefore increase machine performance (World Economic Forum, 2020). Additionally, making more data available for analysis supports tackling societal and environmental challenges (European Commission, 2020).

Data marketplaces have been emerging in recent years to match the supply and demand for data. Data marketplaces provide a digital platform enabling data providers to sell their data to data consumers. Concurrently, data consumers get access to otherwise inaccessible data sources. Though data marketplaces bring along benefits, they pose multiple challenges for data governance. These challenges manifest in a variety of issues within different data governance decision domains, including how to protect data, ensure data quality, and define and model

---

Responsible Editor: Martin Smits

✉ Jan vom Brocke  
jan.vom.brocke@uni.li

Rene Abraham  
rene.abraham@uni.li

Johannes Schneider  
johannes.schneider@uni.li

<sup>1</sup> Institute of Information Systems, University of Liechtenstein, Fürst-Franz-Josef-Strasse, 9490 Vaduz, Liechtenstein

data consistently. Regarding data security, companies fear a loss of control over their data, which could lead to a competitive disadvantage (Roman & Stefano, 2016; Spiekermann, 2019; van den Broek & van Veenstra, 2015). Concerning data quality, data consumers require insights into the quality of data products before they can use them for certain purposes (Jansen et al., 2012). If data products are further processed and used as the basis for producing other goods, a high level of data quality is required (Stahl et al., 2017). Issues with the quality of data products can hamper efforts to provide a marketplace service (Smith et al., 2016). Regarding data architecture, the lack of standards inhibits data sharing (European Commission, 2022). It also poses barriers to finding, analyzing, and processing published data (Smith et al., 2016). These challenges indicate that further research is required regarding data governance in the context of data marketplaces.

This paper adopts a comprehensive perspective by establishing a taxonomy of data governance decision domains in data marketplaces. A taxonomy structures and organizes the body of knowledge within a certain field (Glass & Vessey, 1995). It builds a foundation for future research by allowing researchers to determine relationships between the taxonomy's dimensions and other variables of interest. A taxonomy is also helpful in identifying divergence in previous research findings (Sabherwal & King, 1995). We aim to create a taxonomy that describes the different mechanisms available to instantiate data governance decision domains in the context of data marketplaces. Hence, our study answers the following research question: *how do data marketplaces instantiate data governance decision domains?* Our study advances the body of knowledge on data governance in data marketplaces, which has to date been little researched (Koutroumpis et al., 2020). Our findings also contribute to the body of knowledge on the wider topic of data governance for ecosystems of public and private organizations (Tiwana et al., 2014).

The remainder of this paper is structured as follows. First, we present the theoretical background regarding data marketplaces and data governance decision domains. Second, we describe the taxonomy development method applied for the study. Third, we present our findings concerning the dimensions, subdimensions, and characteristics of the taxonomy. Fourth, we discuss our findings in the context of scientific literature and present the final taxonomy. In doing so, we highlight current limitations in the instantiation of data governance decision domains. We then conclude with a summary, describe the limitations of our study, and suggest avenues for future research.

## Theoretical background

### Data marketplaces

Data marketplaces are digital platforms that offer data or data-related services as primary goods (Stahl et al., 2016).

Data goods encompass manually and automatically created personal and commercial data, such as age, gender, purchase history, and IoT sensor data. Data-related services comprise capabilities such as data aggregation, analysis, and visualization (Roman & Stefano, 2016; Spiekermann, 2019). In our study, we focus on data marketplaces that act as independent intermediaries connecting two or more market participants (Stahl et al., 2016). The main actors involved in data trades are data providers offering data, and data consumers buying data. Marketplace providers offer an infrastructure that allows these actors to upload, discover, buy, and sell data (Spiekermann, 2019; Stahl et al., 2016).

### Data governance decision domains in data marketplaces

Data governance is a framework that provides structure and formalization for the management of data (Morabito, 2015; Rifaie et al., 2009; Weber et al., 2009). As part of data governance, organizations typically need to specify what must be governed, i.e., the scope of data (Abraham et al., 2019), who governs the data, i.e., the roles and governance bodies (Khatri & Brown, 2010; Otto, 2011), and what decisions must be made in data-related areas, i.e., the data governance decision domains (Abraham et al., 2019; Khatri & Brown, 2010; Lee et al., 2019). This study deals with the last component. Based on Khatri and Brown (2010) and Abraham et al. (2019), we distinguish between the following six data governance decision domains: (a) data quality; (b) data security; (c) data architecture; (d) metadata; (e) data lifecycle; (f) data storage and infrastructure. Based on Schrieck et al. (2016) we add (g) data pricing as an additional data governance decision domain since data marketplaces contain the aspect of data trade and data valuation. In the remainder of this section, we describe the seven data governance decision domains that form the foundation of the taxonomy.

#### Data quality

Data quality refers to the ability of data to satisfy its usage requirements in a given context (de Abreu Faria et al., 2013; Khatri & Brown, 2010). Data quality is characterized by quality dimensions such as completeness, credibility, accuracy, timeliness, and consistency of data (DAMA International, 2009; Khatri & Brown, 2010). Scientific literature proposes both preventive and reactive measures to manage data quality (Otto et al., 2012). In the context of data marketplaces, preventive measures inhibit data providers from onboarding data products with insufficient quality. For example, data providers apply automated test scripts to examine the quality of their data products before making the products available on

the data marketplace (Smith et al., 2016). Reactive measures aim to support the identification and reporting of data quality issues after data products have been made available on the data marketplace. Examples include rating systems that allow data consumers to rate and provide feedback on data products (Zuiderwijk et al., 2014) or data providers (Mišura & Žagar, 2016; Ramachandran et al., 2018).

### Data security

Data security refers to the preservation of security requirements concerning the accessibility, authenticity, availability, confidentiality, integrity, privacy, and reliability of data (Carretero et al., 2017; de Abreu Faria et al., 2013; Donaldson & Walker, 2004; ISACA, 2013). In the context of data marketplaces, requirements concern the control of when, to whom, and to what extent data is being sold (Mišura & Žagar, 2016; Tzianos et al., 2019) and how and where data is being used (Otto & Jarke, 2019; Roman & Stefano, 2016). To store data confidentially, data marketplaces use encryption techniques (Roman & Stefano, 2016; Shaabany et al., 2016; Tzianos et al., 2019). To protect sensitive data during data usage, data marketplaces apply methods that only provide access to parts of the data or even fully restrict access to raw data. Examples include the utilization of anonymization techniques to hide identity data (Fung et al., 2010; Ha et al., 2019) and the application of homomorphic encryption to enable mathematical operations on encrypted data (Roman & Stefano, 2016). To control data usage and protect data ownership rights, data marketplaces apply data usage terms that describe the appropriate uses of data (Otto & Jarke, 2019; Truong et al., 2012; Tzianos et al., 2019). Similarly, data contracts help to negotiate and assure the authorizations, obligations, and prohibitions on data covered by the contract (Allen, et al., 2014; Matteucci et al., 2012). They enable data providers to have a remedy against data consumers in case of contract infringements (Truong et al., 2012).

### Data architecture

Data architecture is a set of data specifications, which is used to define data requirements and guide data integration (DAMA International, 2009). Data architecture also contains comprehensive data models on a conceptual, logical, and physical level (DAMA International, 2009; Watson et al., 2004). In the context of data marketplaces, data standards are often mentioned as being crucial to supporting interoperability and data exchange between data providers and data consumers (Lis & Otto, 2020; Spiekermann, 2019). However, data marketplaces apply different approaches regarding the data format to facilitate data exchange, ranging from standardized through proprietary to hybrid. Within the standardized approach, data marketplaces

define standardized vocabularies and formats, to which all marketplace participants must adhere (Ito, 2016; Otto & Jarke, 2019). The proprietary approach allows data providers to offer their data products using their own proprietary data formats (Özyilmaz et al., 2018). The most convenient approach for both data providers and consumers is a hybrid approach, where data providers can offer data products in proprietary formats, which then are automatically normalized on the marketplace platform using a standardized data model (Nagorny et al., 2018).

### Metadata

Metadata represents data about data (DAMA International, 2009; Were & Moturi, 2017). It gives meaning and context to data by providing a structured description of the content, quality, and other characteristics of data (Hovenga & Grain, 2013; Khatri & Brown, 2010). Within data marketplaces, rich contextual metadata is important for supporting data consumers in finding data of interest (Tzianos et al., 2019), determining the usefulness of data products (Ramachandran et al., 2018), and correctly interpreting and processing data (Zuiderwijk et al., 2014). Scientific literature provides two approaches regarding the metadata vocabulary in the context of data marketplaces. The first approach contains a marketplace-specific metadata vocabulary, which is used by data providers for describing and publishing metadata, and by data consumers for looking up and retrieving metadata (Otto & Jarke, 2019). The second approach comprises the application of standardized metadata vocabularies such as CERIF and DCAT (W3C, 2020; Zuiderwijk et al., 2014).

### Data lifecycle

The data lifecycle represents the approach of defining, collecting, creating, using, maintaining, archiving, and deleting data (Khatri & Brown, 2010; Morabito, 2015). In the context of data marketplaces, the main data lifecycle phases are data onboarding, data discovery, data purchase, and data usage. During data onboarding, data providers capture, create, and store data, which is made available to the data marketplace (Otto & Jarke, 2019). Within the data discovery phase, data consumers try to find the right data for their purpose (Mišura & Žagar, 2016; Ramachandran et al., 2018). During the data purchase step, data consumers pay for data products, and data providers grant access to the purchased data (Musso et al., 2019; Tzianos et al., 2019). In the data usage phase, data consumers use the data, e.g., by enriching and aggregating it (Otto & Jarke, 2019).

### Data storage and infrastructure

Data storage and infrastructure focus on information technology (IT) artifacts that enable effective data management

(Dreibelbis et al., 2008; Tallon et al., 2014). In a data marketplace context, one major question concerns how the data should be stored. Spiekermann (2019) distinguishes between the centralized, decentralized, and hybrid storage approaches. With the centralized approach, data products are offered by data providers via a central location such as cloud infrastructure. With the decentralized approach, data products remain with data providers. The hybrid approach is a combination of both the centralized and the decentralized approaches.

### Data pricing

With the trade of data between independent parties, the question of how to price data products becomes relevant. Data marketplaces often apply pay-per-use or subscription-based pricing models in line with their business models. Within a pay-per-use pricing model, data marketplaces charge data consumers for each consumed data product (Spiekermann, 2019; Truong et al., 2012). Within a subscription-based pricing model, data consumers have access to data products for a certain period. In addition to pay-per-use and subscription-based pricing models, data products can be provided free of charge. This often includes data from public authorities and non-profit organizations (Spiekermann, 2019). We also identified hybrid pricing models, such as the freemium pricing model, where data providers offer basic data products free of charge while charging a premium for more detailed data products (Thomas & Leiponen, 2016). Furthermore, data pricing contains the determination of the right price for data products (Truong et al., 2012). Apart from fixed prices for data products, data marketplaces apply more dynamic pricing strategies such as bidding (Maruyama et al., 2013; Parra-Arnau, 2018), progressive pricing (Spiekermann, 2019), the “pay what you want” approach (Zuiderwijk et al., 2014), and packaged pricing (Spiekermann, 2019).

## Methodology

### Taxonomy development method

Taxonomies support systematically organizing and describing the body of knowledge in a certain field (Glass & Vessey, 1995). As we aim to create a structured overview of mechanisms to instantiate data governance decision domains, a taxonomy is particularly well suited. The development of a taxonomy is a multistep process (Fiedler et al., 1996). We use the taxonomy development method by Nickerson et al. (2013) which comprises four main steps. The first step is the identification of the meta-characteristic. This is based on the purpose of the taxonomy and guides the choice of the remaining characteristics within the taxonomy. The second

step comprises the selection of objective and subjective ending conditions. The third step initiates an iterative approach for the development of the taxonomy. Within each development cycle, the researcher can choose between an empirical-to-conceptual or a conceptual-to-empirical approach. The fourth step contains the validation of the taxonomy against the objective and subjective ending conditions. The taxonomy development ends if all objective and subjective ending conditions are fulfilled. Otherwise, the taxonomy development continues with the third step.

### Data collection

The empirical-to-conceptual part of the taxonomy development method requires the researcher to select a sample of objects from which to derive characteristics (Nickerson et al., 2013). We considered objects for inclusion if they met the following two criteria: (a) the object represents a data marketplace or a data marketplace protocol enabling trading between data providers and data consumers; (b) the main products traded on the data marketplace are data products or data services. We selected multiple objects since the evidence is considered more robust and generalizable than from a single object (Herriott & Firestone, 1983).

We retrieved an initial list of 177 data marketplaces via the datarade.ai website. In a first round, we reviewed all 177 data marketplaces against the two inclusion criteria and reduced the list to 63 data marketplaces. In a second round, we reviewed the remaining 63 data marketplaces and excluded instances where the marketplace provider did not act as an independent intermediary but rather as an aggregator of different data providers. We also excluded instances which did not provide sufficient official and openly accessible information to be analyzed. In doing so, we reduced the list to 13 data marketplaces to be included in our study. The selected data marketplaces enabled trading personal data, corporate data, and IoT sensor data. Personal data represents data about natural persons such as gender, date of birth, place of residence, and personal interests. Corporate data comprises data about companies, such as company descriptions and financial market data. IoT sensor data encompasses data specifically collected from such IoT devices as sensors installed in cars, smart homes, and smart factories. Table 1 provides an overview of the examined data marketplaces.

We consider the sample set of data marketplaces as appropriate for studying the instantiation of data governance decision domains since they support various types of data products and are located in diverse countries. Also, some data marketplaces enable the creation and sale of data analysis results, which require the management of additional data stakeholders and complex data infrastructures. The primary source of evidence for our study was official publications

**Table 1** Overview of data marketplaces as of 24 June 2022

<i>Name</i>	<i>Country</i>	<i>Traded data product</i>	<i>Status</i>
Airbloc	South Korea	Personal data	Active
Databroker	Belgium	IoT sensor data	Active
Datapace	France	IoT sensor data	Inactive
Datum	Switzerland	Personal and corporate data (including IoT sensor data)	Active
Enigma	USA	Personal and corporate data (including IoT sensor data)	Active
IOTA	Germany	IoT sensor data	Active
MADANA	Germany	Personal and corporate data (including IoT sensor data)	Inactive
oneTRANSPORT	UK	Corporate data (including IoT sensor data)	Active
Opiria	Germany	Personal data	Active
Streamr	Switzerland	IoT sensor data	Active
VETRI	Switzerland	Personal data	Active
Wibson	USA	Personal data	Inactive
Zenome	Russia	Personal health data	Active

from the sample set of data marketplaces. We collected data provided in whitepapers, reports, and on data marketplace websites.

## Taxonomy development

Following the taxonomy development method by Nickerson et al. (2013), we started with the definition of the meta-characteristic. Since our purpose was to investigate how data marketplaces instantiated data governance decision domains, we chose data governance instantiation in data marketplaces as our meta-characteristic. We then determined the ending conditions for the validation of our taxonomy. We omitted the objective ending condition that at least one object was classified under every characteristic. We perceived characteristics solely derived from scientific literature as valid because they provided for differentiation and thus robustness of the taxonomy. Tables 2 and 3 provide an overview of the selected objective and subjective ending conditions. After each taxonomy development cycle, we checked to see whether the resulting taxonomy met all selected objective and subjective ending conditions. If the test result was negative, we initiated a new development cycle.

In total, we conducted four development cycles to reach the final version of the taxonomy. In the first taxonomy development cycle, we used a conceptual-to-empirical approach and conceptualized dimensions, subdimensions, and characteristics taken from scientific literature. Based on Khatri and Brown (2010) and Abraham et al. (2019), we chose the following data governance decision domains as the initial dimensions: data quality, data security, data architecture, metadata, data lifecycle, and data storage and infrastructure. The resulting version of the taxonomy gave us a first impression of the taxonomy structure. During the second development cycle, we applied an empirical-to-conceptual approach. We used the open coding technique to turn collected raw data from the sample of objects into dimensions, subdimensions, and characteristics (Corbin & Strauss, 2015). We first assigned labels to key areas of text using a software tool for content analysis (Tallon et al., 2014). In doing so, we created 167 codes in total. Then, we reviewed the codes and filtered for those with a focus on the meta-characteristic, which reduced the number of codes to 80. We searched for common patterns among the codes and underlying raw data to derive generic characteristics. Afterward, we grouped related characteristics under dimensions and subdimensions.

**Table 2** Overview of objective ending conditions derived from Nickerson et al. (2013)

### *Objective ending conditions*

The taxonomy consists of dimensions and subdimensions each with mutually exclusive and collectively exhaustive characteristics

All objects or a representative sample of objects have been examined

No object was merged with a similar object or split into multiple objects in the last development cycle

No new dimensions, subdimensions, or characteristics were added in the last development cycle

No dimensions, subdimensions, or characteristics were merged or split in the last development cycle

Every dimension and subdimension is unique and not repeated (i.e., there is no dimension or subdimension duplication)

Every characteristic is unique within its dimension (i.e., there is no characteristic duplication within a dimension)

Each cell (combination of characteristics) is unique and is not repeated (i.e., there is no cell duplication)

**Table 3** Overview of subjective ending conditions derived from Nickerson et al. (2013)*Subjective ending conditions*

Concise: the number of dimensions allow the taxonomy to be meaningful without being unwieldy or overwhelming
Robust: the dimensions, subdimensions, and characteristics provide for differentiation among objects sufficient to be of interest
Comprehensive: all objects or a (random) sample of objects within the domain of interest can be classified
Extendible: a new dimension, subdimension, or a new characteristic of an existing subdimension can be easily added
Explanatory: the dimensions, subdimensions, and characteristics provide useful explanations of the nature of the objects

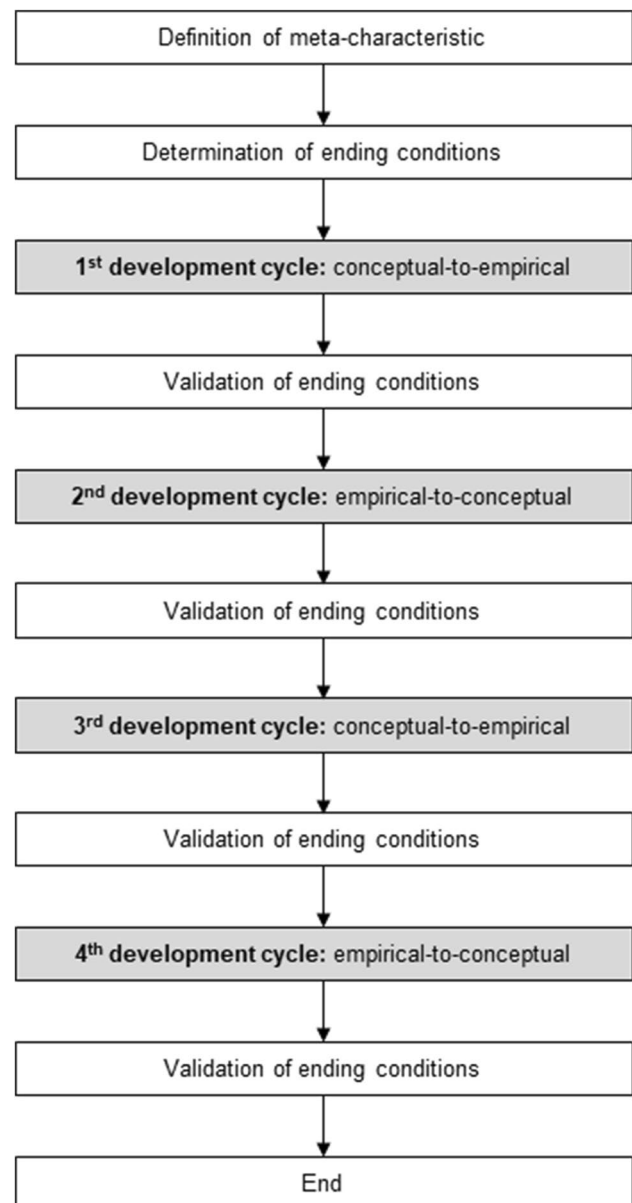
During this analysis step, we added data pricing as a new dimension, among others. We conducted a third development cycle because we had previously added the data pricing dimension. Also, the validation via objective ending conditions showed that the characteristics within the data lifecycle dimension were not mutually exclusive. We applied a conceptual-to-empirical approach to derive additional relevant subdimensions and characteristics from scientific literature for the data pricing dimension. Furthermore, we restructured characteristics within the data lifecycle dimension to make the characteristics mutually exclusive. Since we restructured characteristics, we did not meet all objective ending conditions. Therefore, we conducted a fourth empirical-to-conceptual development cycle and classified again all 13 data marketplaces using the final version of the taxonomy. As we met all objective and subjective ending conditions this time, we concluded the taxonomy development process. Figure 1 shows an overview of the taxonomy development process.

## Findings

The following chapter presents the findings from the analysis of 13 data marketplaces. We describe the results for each data governance decision domain. We also present selected citations from our sources to substantiate our findings.

## Data quality

Regarding data quality, we identified preventive and reactive measures among data marketplaces. On the preventive side, data marketplaces applied automated data validation methods during data onboarding. In doing so, they prevented flawed data products from being further processed and offered on the data marketplace. On the reactive side, we identified one data marketplace that offered a rating system based on consumer feedback to rank data providers. If data consumers detected and reported fake or incorrect data, data marketplaces reduced the rating of the data provider. Several data marketplaces planned to establish similar rating systems to rate data providers or data products.

**Fig. 1** Overview of the taxonomy development process

“The data model is used within the data normalization process and plays a key role. It defines how values should be stored in the local data store and is used to

identify rule violations, thus establishing a consistent level of quality and consistency. It enforces specific units, length and a structure on the stored data, making it possible to analyze the data. Only if the data is accurate, reliable, and formatted consistently, further processing will be possible.” (MADANA, 2018, p. 27)

– Preventive measure

“In order to guarantee the validity of the data, Datapace employs several mechanisms - like seller reputation rating (...).” (Datapace, 2017, p. 4) – Reactive measure

## Data security

Data marketplaces applied several mechanisms to ensure data security. Concerning the confidential storage of data, most data marketplaces applied encryption techniques such as public-key cryptography. Furthermore, we identified data fragmentation as a method whereby either the data payload was split into different fragments or the data providers’ identity information and the data payload were split. The fragments were stored in different physical storage locations. The data marketplaces we analyzed implemented data fragmentation as an additional mechanism to data encryption and therefore applied a hybrid approach.

“VETRI users will store their most sensitive data locally on their device by using state-of-the-art encryption techniques (...).” (VETRI, n.d., p. 9) – Data encryption

“Full Privacy: In this case data are stored in encrypted form, (...). Standard Privacy: Genetic data are stored as fragments making it impossible to identify the user.” (Zenome, 2017, p. 29) – Hybrid storage protection

“The data is fragmented to a number of unknown physical locations, and it is protected by strong encryption while in transit and in storage.” (Streamr, 2017, p. 19)

– Hybrid storage protection

In terms of data access control, data marketplaces provided the options to grant access instantly or based on consent. When applying instant access, data consumers received access to a data product immediately after purchase and without any explicit consent from data providers. Consent-based access enabled data providers to decide which data they wanted to sell to selected data consumers. We also identified two data marketplaces that applied a hybrid approach by allowing both authorization options.

“The user clients shows [sic] currently running projects requesting data access and users can control whether to give access or not based on their decision.” (Datum, 2017, p. 10) – Consent-based access

“Subscribing to streams can be restricted to certain users only, or be free to the public.” (Streamr, 2017, p. 13) – Hybrid access control measure

Regarding the confidential usage of data, we identified a few data marketplaces that only provided access to parts of the data by anonymizing the data. Two data marketplaces did not provide access to the raw data at all. Instead, they restricted data processing to the marketplace platform without providing data consumers with direct access to raw data.

“Data can be offered anonymously, so privacy is not violated.” (Streamr, 2017, p. 10) – Access to data parts

“Data will only be processed in secured environments and afterward deleted to minimize the risk of unwanted data breaches.” (MADANA, 2018, p. 17) – No access to data

We also found evidence that data marketplaces supported the application of data contracts. These allowed data providers to determine the conditions under which their data products should be used and enabled data consumers to describe how they planned to use the data.

“Your rights to use the data are governed by a licence that has been drafted by the data provider. When you purchase data, you need to confirm that you accept the terms of the licence.” (Databroker, 2021) – Contract-based data usage control

## Data architecture

Concerning data architecture, we identified standardized, proprietary, and hybrid approaches regarding the data format. Using the standardized approach, some data marketplaces required data providers to format data products according to a unified data model before publishing these products on the data marketplace. Applying the proprietary approach, data marketplaces allowed data providers to publish data products using proprietary data formats. Since the latter could have inhibited data consumers in automatically interpreting data, some data marketplaces required the submission of the data payload format as part of the metadata. We also identified a data marketplace that applied a hybrid approach where data products in proprietary formats were automatically pre-processed and standardized based on standardized data models before storage.

“For each new sensor, we ask you to provide the following information: (...) Data Fields: The most essential part of the sensor configuration. Please provide information for every parameter that will be captured by the sensor and stored on the Tangle. (...) Parameter information consist [sic] of 3 fields: Field ID (...). Field Name (...). Field Unit (...).” (IOTA, 2020) – Standardized data format

“It is responsibility of data seller to provide a valid data source URL and give detailed description of the data stream and it’s [sic] format (it’s [sic] JSON schema) – so it can be easily consumed by data buyer.” (Datapace, 2017, p. 4) – Proprietary data format

“The normalization process builds on the interpretation of the data before the data is put into the local data store. The standardization process then reformats the data and creates a consistent data representation with fixed and discrete columns based on the data model. The advantage of standardization is that the conformity of the data guarantees simpler and more secure processing of the data.” (MADANA, 2018, p. 27) – Hybrid data format

## Metadata

The data marketplaces used metadata to support data providers in organizing their data products and facilitate data consumers in discovering relevant datasets. Most analyzed data marketplaces provided a marketplace-specific set of metadata fields to capture metadata when onboarding new data products. Common metadata fields comprised the description of the offered data product, the data owner, the price, access permissions, and the terms and conditions of data use.

“For each new sensor, we ask you to provide the following information: Device ID (...). Device Type (...). Company (...). Location (city/country) (...). GPS Coordinates (latitude/longitude) (...). Price of the data stream (...).” (IOTA, 2020) – Marketplace-specific vocabulary

## Data lifecycle

Regarding the data lifecycle, we identified two types of data marketplaces. The first type of marketplace focused on data trading, encompassing the phases of data onboarding, data discovery, data purchase, and data offboarding. Most analyzed data marketplaces fell under this category. The second marketplace type contained an additional data usage phase. The data usage phase supported the processing of data within the marketplace platform and the provisioning of analysis results to data consumers.

“Via their gateway operator, the sensor owners place the data generated by their sensors up for sale (...), and buyers can discover and purchase access to the data using that same DTX token. (...) Data generated by the sensors of their clients is sent (...) to their dAPI which check who has purchased access and send the data directly on to the location specified by the buyer on purchasing.” (Databroker, n.d., p. 6) – Data trade-focused marketplace

“In case of a mobile survey all answers from all consumers worldwide are aggregated and visually presented on selectable charts and in table form. Since GPS position of each consumer is tracked, Opiria can display the location of each answer on a world map.” (Opiria, 2017, p. 13) – Data usage-focused marketplace

## Data storage and infrastructure

Regarding data storage, we identified the centralized and decentralized storage approach within the data marketplaces. A few data marketplaces applied the centralized storage approach and used a central database or a central, cloud-based storage solution to store the data. However, most analyzed data marketplaces applied a decentralized storage approach, of which we identified three forms. The first form encompassed data being stored on the data provider’s device such as a mobile phone. The second form comprised a decentralized storage node architecture provided by the data marketplace. Independent storage nodes were paid for providing computing power and storage capacity to replicate and store the data in a distributed network. The third decentralized storage form entailed data being stored at a storage vendor of the data provider’s choice.

“oneTRANSPORT provides a cloud-based platform (...).” (oneTRANSPORT, 2017, p. 6) – Centralized storage

“VETRI users will store their most sensitive data locally on their device (...).” (VETRI, n.d., p. 9) – Decentralized storage

“A Storage Node receives the data and stores the data. The data is replicated to many other storage nodes.” (Datum, 2017, p. 14) – Decentralized storage

“Built in the dAPI, there are connectors to integrate with the leading IoT and bigdata [sic] storage vendors, leaving the choice on where their data needs to be sent.” (Databroker, n.d., p. 24) – Decentralized storage

## Data pricing

Regarding data pricing, most analyzed data marketplaces applied a pay-per-use or subscription-based pricing model. Some data marketplaces applied a hybrid pricing model where data providers could decide if they offered data products at a certain price per use, based on a subscription, or free of charge.

“Consumers that consent to provide their data would trigger a smart contract between the consumer and the company. On this basis the consumer is paid with PDATA tokens and the company receives the requested personal data.” (Opiria, 2017, p. 3) – Pay-per-use

“(…) Enigma creates a decentralized data marketplace that allows people, companies and organizations to contribute data (...), which users of the system can then subscribe to and consume.” (Enigma, 2020) – Subscription-based

“Data can be purchased as one-off, or on an on-going subscription basis.” (Datum, 2017) – Hybrid pricing model



“The Marketplace is filled with both paid and free products, offering data producers an opportunity to either monetise their data or make it freely available to everyone.” (Streamr, 2020) – Hybrid pricing model

For their pricing strategy, most analyzed data marketplaces applied a fixed price approach. Some data marketplaces applied the bidding process where data consumers offered a price to data providers, who then accepted or declined the offer or made a counteroffer.

“Price of the data stream: Here you can define the cost of the sensor data.” (IOTA, 2020) – Fixed price

“A Data Consumer declares interest to purchase the piece of data. (...) The User receives a data purchase request with the details such as purchaser and price offered. He can agree to the purchase request or counter offer with a modified proposal.” (Datum, 2017, p. 14) – Dynamic pricing

## Discussion

Our findings show that data marketplaces have multiple options to instantiate data governance decision domains. We observed certain tendencies, but also limitations of specific mechanisms. In the following, we discuss our main findings in the context of scientific literature and present our final taxonomy.

Regarding data security, our findings would suggest that data marketplaces offer limited protection of sensitive data. Though data marketplaces apply anonymization techniques, Narayanan and Shmatikov (2008) have demonstrated that the de-anonymization of datasets is possible with little auxiliary information. Secured execution environments, which restrict direct access to raw data, promise a higher level of protection. Nevertheless, in cases where raw data is used to train machine learning algorithms, adversaries could identify the raw data by using model inversion attacks (Fredrikson et al., 2015). It becomes essential, therefore, for data providers to undertake a thorough check to assess whether their data products comprise sensitive data. Furthermore, in most analyzed data marketplaces, data providers transfer data products to data consumers. Thus, the main mechanism to control data usage and protect data ownership rights is the application of data contracts between data providers and consumers. However, the application of data contracts cannot fully prevent the illegal and malevolent copying and reselling of data products. In cases of unauthorized reselling, being able to prove data ownership is essential. Technology-based data usage control mechanisms such as watermarking could help to prove data ownership rights. By applying watermarking, data providers could embed watermark data such as data ownership information into data products (Agarwal et al., 2019; Vlachos et al., 2015).

Furthermore, our findings would suggest that data quality management within data marketplaces is still at an early stage. Almost half of the data marketplaces did not actively approach the topic of ensuring data quality. Where we identified measures, these were mainly rating systems consistent with those proposed by Mišura and Žagar (2016), Ramachandran et al. (2018), and Zuiderwijk et al. (2014). Using these reactive approaches, the responsibility for identifying and reporting data quality issues often lies with data consumers. A hybrid approach comprising both preventive and reactive measures can help to overcome this drawback. For example, guided approaches such as LANG can help data consumers reactively identify data quality issues in datasets for which they have minimal control or knowledge of underlying rules (Zhang et al., 2019). Simultaneously, data marketplace providers can use LANG to prevent flawed data products from being onboarded on the marketplace. Another solution comprises the provisioning of warranties for data products. If data providers do not deliver data products at the expected level of quality, data consumers have the right to cancel the purchase and demand a refund. The terms and conditions could either be stipulated by law or by data marketplace providers similar to guarantees provided by marketplaces such as Amazon (Amazon, 2022) and payment providers such as PayPal (PayPal, 2022).

In terms of data architecture, our findings do not reveal a clear tendency towards a specific data formatting approach. Instead, our findings would suggest that the approaches described by Tzianos et al. (2019), Özyilmaz et al. (2018), and Nagorny et al. (2018) are valid in different contexts. Data marketplaces, which aim to support data consumers in the automatic processing of data products, are likely to provide a standardized data format for the data payload. This applies in particular to data products that are published regularly or in real-time such as data streams. Data marketplaces that want to keep adoption barriers low for data providers might enable the use of proprietary data formats. This might especially be the case for data marketplaces new on the market. Those data marketplaces that focus on convenience for both data providers and data consumers might apply a hybrid approach.

Regarding data storage and infrastructure, our findings show a tendency towards decentralized storage solutions. Preserving data within data providers' storage systems offers data providers an increased level of control over their data assets (Spiekermann, 2019). Also, storing data using distributed storage nodes enables the scalability of storage and facilitates the availability of data and fault tolerance through data replication. These findings contrast with data governance in traditional companies where the inclusion of external IT infrastructure is negatively related to data governance maturity (Borgman et al., 2016). Also, the storage of data in several disparate databases often inhibits the accessibility and consistency of

data (Cheong & Chang, 2007; Tallon et al., 2014). The high level of standardization and integration within the marketplace platform architectures could be reasons why a decentralized storage approach is successfully applied.

Concerning data pricing, our results confirm earlier findings showing that the pay-per-use model and the subscription-based model are the more common pricing models applied in data marketplaces (Truong et al., 2012). Most data marketplaces within our sample applied one of these two pricing models. However, we also identified data marketplaces that allowed data providers to offer data products using a pricing model of their choice. This approach might support marketplace adoption as it might attract more data marketplace participants. Within the set of dynamic pricing strategies, we only identified bidding as an applied pricing strategy, omitting other dynamic pricing strategies such as progressive pricing, “pay what you want”, or packaged pricing. We suspect there are different reasons behind this result. Progressive pricing is applied where the dissemination of data products is to be restricted (Spiekermann, 2019). Given that data marketplaces are fairly novel, data providers have likely no incentive to restrict the dissemination of their data products. Furthermore, the commercial interest of data providers and data marketplaces pre-empts the application of a “pay what you want” pricing approach. However, increased adoption of data marketplaces could bring along the application of additional pricing strategies.

Figure 2 shows our resulting taxonomy of data governance decision domains in data marketplaces. Per characteristic, the number in the bottom right corner illustrates how often we found that characteristic within the analyzed data

marketplaces. A hybrid characteristic indicates that a data marketplace applies a combination of the characteristics within the respective subdimension. The taxonomy meets all selected objective ending conditions. Each dimension of the taxonomy consists of mutually exclusive and collectively exhaustive characteristics. All objects were examined. During the last development cycle, we did not merge or split any objects. We also did not add, merge, or split any dimensions, subdimensions, or characteristics during the last development cycle. Every dimension, subdimension, characteristic, and combination of characteristics is unique within the taxonomy. In addition, the taxonomy meets all subjective ending conditions. The taxonomy is concise since the number of dimensions is in the proposed range of seven plus or minus two. The taxonomy is robust since the added characteristics provide for differentiation among objects. The taxonomy is comprehensive since we were able to classify all 13 data marketplaces. By adding dimensions and characteristics during development cycles, we were able to demonstrate that the taxonomy is extendible. Furthermore, the taxonomy is explanatory because the dimensions, subdimensions, and characteristics provide explanations of the nature of the objects.

### Conclusion

Data is at the center of digital transformation. Facilitating the exchange of data between independent market participants has the potential to generate significant economic,

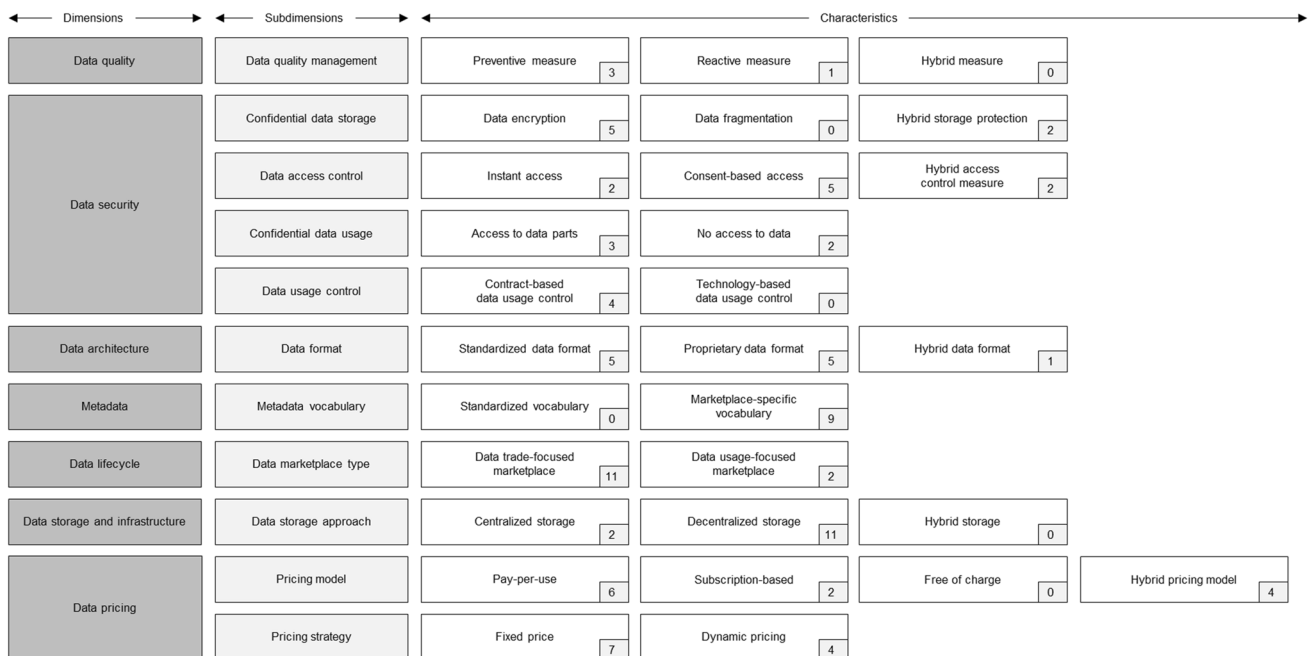


Fig. 2 Taxonomy of data governance decision domains in data marketplaces

societal, and environmental benefits. However, the fear of unintentionally releasing sensitive data, the lack of control over data usage, alongside accessibility and interoperability issues create trust-related and technical barriers to data sharing (World Economic Forum, 2020). To overcome these barriers, the trade and exchange of data should be accompanied by robust data governance practices, increasing the level of certainty and producing new opportunities. Hence, the research reported in this paper analyzed the emerging topic of data marketplaces from a data governance perspective. The following research question framed our study: how do data marketplaces instantiate data governance decision domains? We answered this question by developing a taxonomy comprising the subdimensions and characteristics of data governance decision domains in the context of data marketplaces.

Our study has the following limitations. As the primary source of evidence, we reviewed official documentation such as whitepapers, reports, and information provided on data marketplace websites. Future research should conduct case studies collecting data from interviews and observations to enhance our findings. In addition, we used a sample size of 13 data marketplaces. Future research should examine a larger sample size to validate the robustness and comprehensiveness of our taxonomy. Furthermore, our data did not allow for rigorous testing of how different types of data products influence the instantiation of data governance decision domains. Future research should analyze the possible configurations of data governance decision domains based on different types of data products. Also, future research should investigate which roles and governance bodies have the decision-making authority within each decision domain and which marketplace actor takes on which role.

The results of our study advance scientific literature. To the best of our knowledge, our study is the first to investigate data marketplaces through a data governance lens and thus combine these two research strands. The taxonomy of data governance decision domains offers a common terminology that can be used by researchers to share their findings with other members of the information systems community. Moreover, the taxonomy represents a foundation for further scientific investigation and theory-building. For example, the taxonomy can be used to study relationships between the taxonomy concepts and other variables of interest (Glass & Vessey, 1995). The taxonomy also enables researchers to understand divergence in previous research findings regarding data marketplaces (Sabherwal & King, 1995). Additionally, the taxonomy can be used to define data governance archetypes for data marketplaces. From a practitioner's perspective, the taxonomy highlights relevant data governance decision domains and instantiation options in the context of data marketplaces. When trading and exchanging data with third parties, neglecting aspects such as data security,

privacy, and data quality can have unforeseen consequences. Marketplace providers can use our findings to develop a data governance strategy in a structured and thoughtful manner. Our results can also be used by traditional companies aiming to implement an internal data marketplace.

**Funding** Open access funding provided by University of Liechtenstein

## Declarations

**Conflict of interest** The authors declare no competing interests.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, 49, 424–438. <https://doi.org/10.1016/j.ijinfomgt.2019.07.008>
- Agarwal, N., Singh, A. K., & Singh, P. K. (2019). Survey of robust and imperceptible watermarking. *Multimedia Tools and Applications*, 78, 8603–8633. <https://doi.org/10.1007/s11042-018-7128-5>
- Allen, C., Des Jardins, T. R., Heider, A., Lyman, K. A., McWilliams, L., Rein, A. L., & Turske, S. A. (2014). Data governance and data sharing agreements for community-wide health information exchange: Lessons from the beacon communities. *eGEMs*, 2(1), 1–9. <https://doi.org/10.13063/2327-9214.1057>
- Amazon. (2022). <https://www.amazon.com/>. Retrieved February 26, 2022, from <https://www.amazon.com/gp/help/customer/display.html?nodeId=GQ37ZCNECJKTIFYQV>
- Borgman, H., Heier, H., Bahli, B., & Boekamp, T. (2016). Dotting the I and Crossing (out) the T in IT Governance: New Challenges for Information Governance. In *49th Hawaii International Conference on System Sciences* (pp. 4901–4909). <https://doi.org/10.1109/HICSS.2016.608>
- Carretero, A. G., Gualo, F., Caballero, I., & Piattini, M. (2017). MAMD 2.0: Environment for data quality processes implantation based on ISO 8000-6X and ISO/IEC 33000. *Computer Standards & Interfaces*, 54(3), 139–151. <https://doi.org/10.1016/j.csi.2016.11.008>
- Cheong, L. K., & Chang, V. (2007). The need for data governance: Acase study. In *18th Australasian Conference on Information System* (pp. 999–1008).
- Corbin, J., & Strauss, A. (2015). *Basics of qualitative research* (4 ed.). SAGE Publications, Inc.
- DAMA International. (2009). *The DAMA guide to the data management body of knowledge*. Technics Publications, LLC.
- Databroker. (2021). <https://www.databroker.global/>. Retrieved November 25, 2021, from <https://www.databroker.global/help/buying-data/topic/using-data>

- Databroker. (n.d.). databroker dao – Global market for local data. Retrieved October 20, 2019, from <https://www.allcryptowhitepapers.com/wp-content/uploads/2018/11/Databroker-DAO.pdf>
- Datapace. (2017). Datapace – Decentralized data marketplace based on blockchain. Retrieved February 15, 2020, from [https://www.datapace.io/datapace\\_whitepaper.pdf](https://www.datapace.io/datapace_whitepaper.pdf)
- Datum. (2017). Datum Network – The decentralized data marketplace. Retrieved September 15, 2019, from <https://datum.org/assets/Datum-WhitePaper.pdf>
- de Abreu Faria, F., Maçada, A. C., & Kumar, K. (2013). Information governance in the banking industry. In *Proceedings of the 46th Hawaii International Conference on System Sciences* (pp. 4436–4445). Wailea. <https://doi.org/10.1109/HICSS.2013.270>
- Donaldson, A., & Walker, P. (2004). Information governance — a view from the NHS. *International Journal of Medical Informatics*, 73, 281–284. <https://doi.org/10.1016/j.ijmedinf.2003.11.009>
- Dreibelbis, A., Hechler, E., Milman, I., Oberhofer, M., & van Run, P. (2008). *Enterprise master data management: An SOA approach to managing core information*. IBM Press.
- Enigma. (2020, May 1). Retrieved from <https://blog.enigma.co/towards-a-decentralized-data-marketplace-part-2-1362c8e11094>
- European Commission. (2020). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions: A European strategy for data. Retrieved January 17, 2021, from <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066>
- European Commission. (2022). Proposal for a regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act). Retrieved March 21, 2022, from <https://ec.europa.eu/newsroom/dae/redirection/document/83521>
- Fiedler, K. D., Grover, V., & Teng, J. T. (1996). An empirically derived taxonomy of information technology structure and its relationship to organizational structure. *Journal of Management Information Systems*, 13(1), 9–34. <https://doi.org/10.1080/07421222.1996.11518110>
- Forrester. (2018). Design data governance for the data economy.
- Fredrikson, M., Jha, S., & Ristenpart, T. (2015). Model inversion attacks that exploit confidence information and basic countermeasures. In *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* (pp. 1322–1333). Denver. <https://doi.org/10.1145/2810103.2813677>
- Fung, B. C., Wang, K., Chen, R., & Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys*, 42(4), 1–53. <https://doi.org/10.1145/1749603.1749605>
- Glass, R. L., & Vessey, I. (1995). Contemporary application-domain taxonomies. *IEEE Software*, 12(4), 63–76. <https://doi.org/10.1109/52.391837>
- Ha, M., Kwon, S., Lee, Y. J., Shim, Y., & Kim, J. (2019). Where WTS meets WTB: A blockchain-based marketplace for digital me to trade users' private data. *Pervasive and Mobile Computing*, 59, 1–15. <https://doi.org/10.1016/j.pmcj.2019.101078>
- Herriott, R. E., & Firestone, W. A. (1983). Multisite qualitative policy research: optimizing description and generalizability. *Educational Researcher*, 12(2), 14–19. <https://doi.org/10.3102/0013189X012002014>
- Hovenga, E. J., & Grain, H. (2013). Health data and data governance. In *Health information governance in a digital environment* (pp. 67–92).
- IDC. (2018). Data age 2025: The digitization of the world. Retrieved August 12, 2019, from <https://www.seagate.com/de/de/our-story/data-age-2025/>
- IOTA. (2020, May 9). <https://data.iota.org/>. Retrieved from <https://blog.iota.org/iota-data-marketplace-tech-intro-d54b29774f1a-d54b29774f1a/>
- ISACA. (2013). COBIT 5: Enabling information. .
- Ito, R. (2016). ID-Link, an enabler for medical data marketplace. In *2016 IEEE 16th International Conference on Data Mining Workshops (ICDMW)* (pp. 792–797). <https://doi.org/10.1109/ICDMW.2016.0117>
- 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>
- Khatri, V., & Brown, C. V. (2010). Designing data governance. *Communications of the ACM*, 53(1), 148–152. <https://doi.org/10.1145/1629175.1629210>
- Koutroumpis, P., Leiponen, A., & Thomas, L. D. (2020). Markets for data. *Industrial and Corporate Change*, 29(3), 645–660. <https://doi.org/10.1093/icc/dtaa002>
- Lee, S. U., Zhu, L., & Jeffery, R. (2019). Data governance decisions for platform ecosystems. In *52nd Hawaii International Conference on System Sciences* (pp. 6377–6386).
- Lis, D., & Otto, B. (2020). Data governance in data ecosystems – Insights from organizations. *AMCIS 2020 Proceedings*, 12.
- MADANA. (2018). MADANA white paper. Retrieved March 23, 2020, from <https://www.madana.io/download/MADANA-White-Paper.pdf>
- Maruyama, H., Okanohara, D., & Hido, S. (2013). Data marketplace for efficient data placement. In *2013 IEEE 13th International Conference on Data Mining Workshops* (pp. 702–705). <https://doi.org/10.1109/ICDMW.2013.146>
- Matteucci, I., Petrocchi, M., Sbodio, M. L., & Wiegand, L. (2012). A design phase for data sharing agreements: data privacy management and autonomous spontaneous security. In *DPM 2011, SETOP 2011. Lecture Notes in Computer Science* (Vol. 7122, pp. 25–41). [https://doi.org/10.1007/978-3-642-28879-1\\_3](https://doi.org/10.1007/978-3-642-28879-1_3)
- Mišura, K., & Žagar, M. (2016). Data marketplace for internet of things. In *2016 International Conference on Smart Systems and Technologies (SST)* (pp. 255–260). <https://doi.org/10.1109/SST.2016.7765669>
- Morabito, V. (2015). *Big data and analytics: Strategic and organizational impacts*. Springer International Publishing.
- Musso, S., Perboli, G., Rosano, M., & Manfredi, A. (2019). A decentralized marketplace for M2M economy for Smart cities. In *2019 IEEE 28th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE)* (pp. 27–30). <https://doi.org/10.1109/WETICE.2019.00014>
- Nagorny, K., Scholze, S., Ruhl, M., & Colombo, A. W. (2018). Semantical support for a CPS data marketplace to prepare Big Data analytics in smart manufacturing environments. In *2018 IEEE Industrial Cyber-Physical Systems (ICPS)* (pp. 206–211). <https://doi.org/10.1109/ICPHYS.2018.8387660>
- Narayanan, A., & Shmatikov, V. (2008). Robust de-anonymization of large sparse datasets. In *2008 IEEE Symposium on Security and Privacy (SP 2008)* (pp. 111–125). <https://doi.org/10.1109/SP.2008.33>
- 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, 336–359. <https://doi.org/10.1057/ejis.2012.26>
- oneTRANSPORT. (2017). oneM2M and oneTRANSPORT – a global approach to Intelligent Mobility. Retrieved February 16, 2020, from <https://service.onetransport.io/>
- Opiria. (2017). PDATA TOKEN Whitepaper. Retrieved April 27, 2020, from <https://www.opiria.com/wp-content/uploads/2017/09/PDATA-White-Paper-20171119.pdf>

- Otto, B. (2011). Organizing data governance: findings from the telecommunications industry and consequences for large service providers. *Communications of the Association for Information Systems*, 29(1), 45–66. <https://doi.org/10.17705/1CAIS.02903>
- Otto, B., & Jarke, M. (2019). Designing a multi-sided data platform: Findings from the International Data Spaces case. *Electronic Markets*, 29, 561–580. <https://doi.org/10.1007/s12525-019-00362-x>
- Otto, B., Hüner, K. M., & Österle, H. (2012). Toward a functional reference model for master data quality management. *Information Systems and e-Business Management*, 10, 395–425. <https://doi.org/10.1007/s10257-011-0178-0>
- Özyilmaz, K. R., Doğan, M., & Yurdakul, A. (2018). IDMoB: IoT data marketplace on blockchain. In *Crypto Valley Conference on Blockchain Technology (CVCBT 2018)* (pp. 11–19). <https://doi.ieeeecomputersociety.org/10.1109/CVCBT.2018.00007>
- Parra-Arnau, J. (2018). Optimized, direct sale of privacy in personal data marketplaces. *Information Sciences*, 424, 354–384. <https://doi.org/10.1016/j.ins.2017.10.009>
- PayPal. (2022). <https://www.paypal.com/>. Retrieved February 26, 2022, from <https://www.paypal.com/us/webapps/mpp/paypal-safety-and-security>
- Ramachandran, G. S., Radhakrishnan, R., & Krishnamachari, B. (2018). Towards a decentralized data marketplace for Smart Cities. In *2018 IEEE International Smart Cities Conference (ISC2)* (pp. 1–8). <https://doi.org/10.1109/ISC2.2018.8656952>
- Rifaie, M., Alhaji, R., & Ridley, M. (2009). Data governance strategy: A key issue in building enterprise data warehouse. In *Proceedings of the 11th International Conference on Information Integration and Web based Applications & Services* (pp. 587–591). Kuala Lumpur. <https://doi.org/10.1145/1806338.1806449>
- Roman, D., & Stefano, G. (2016). Towards a reference architecture for trusted data marketplaces: The credit scoring perspective. In *2nd International Conference on Open and Big Data (OBD)* (pp. 95–101). <https://doi.org/10.1109/OBD.2016.21>
- Sabherwal, R., & King, W. R. (1995). An empirical taxonomy of the decision-making processes concerning strategic applications of information systems. *Journal of Management Information Systems*, 11(4), 177–214. <https://doi.org/10.1080/07421222.1995.11518064>
- Schreieck, M., Wiesche, M., & Krcmar, H. (2016). Design and governance of platform ecosystems – key concepts and issues for future research. In *Twenty-Fourth European Conference on Information Systems (ECIS)*. Istanbul.
- Shaabany, G., Grimm, M., & Anderl, R. (2016). Secure information model for data marketplaces enabling global distributed manufacturing. In *26th CIRP Design Conference* (Vol. 50, pp. 360–365). <https://doi.org/10.1016/j.procir.2016.05.003>
- Smith, G., Ofc, H. A., & Sandberg, J. (2016). Digital service innovation from open data: exploring the value proposition of an open data marketplace. In *49th Hawaii International Conference on System Sciences* (pp. 1277–1286). <https://doi.org/10.1109/HICSS.2016.162>
- Spiekermann, M. (2019). Data marketplaces: Trends and monetisation of data goods. *Intereconomics: Review of European Economic Policy*, 54(4), 208–216. <https://doi.org/10.1007/s10272-019-0826-z>
- Stahl, F., Schomm, F., Vossen, G., & Vomfell, L. (2016). A classification framework for data marketplaces. *Vietnam Journal of Computer Science*, 3, 137–143. <https://doi.org/10.1007/s40595-016-0064-2>
- Stahl, F., Schomm, F., Vomfell, L., & Vossen, G. (2017). Marketplaces for digital data: Quo Vadis? *Computer and Information Science*, 10(4), 22–37. <https://doi.org/10.5539/cis.v10n4p22>
- Streamr. (2017). Unstoppable data for unstoppable apps: DATAcoin by Streamr. Retrieved April 7, 2020, from [https://s3.amazonaws.com/streamr-public/streamr-datacoin-whitepaper-2017-07-25-v1\\_1.pdf](https://s3.amazonaws.com/streamr-public/streamr-datacoin-whitepaper-2017-07-25-v1_1.pdf)
- Streamr. (2020, May 2). <https://streamr.network>. Retrieved from <https://streamr.network/docs/marketplace/introduction-marketplace>
- Tallon, P. P., Ramirez, R. V., & Short, J. E. (2014). The information artifact in IT governance: Toward a theory of information governance. *Journal of Management Information Systems*, 30(3), 141–177. <https://doi.org/10.2753/MIS0742-1222300306>
- Thomas, L. D., & Leponen, A. (2016). Big Data Commercialization. *IEEE Engineering Management Review*, 44(2), 74–90. <https://doi.org/10.1109/EMR.2016.2568798>
- Tiwana, A., Konsynski, B., & Venkatraman, N. (2014). Special Issue: Information Technology and Organizational Governance: The IT Governance Cube. *Journal of Management Information Systems*, 30(3), 7–12. <https://doi.org/10.2753/MIS0742-1222300301>
- Truong, H.-L., Comerio, M., De Paoli, F., Gangadharan, G., & Dustdar, S. (2012). Data contracts for cloud-based data marketplaces. *International Journal of Computational Science and Engineering*, 7(4), 280–295. <https://doi.org/10.1504/IJCSSE.2012.049749>
- Tzianos, P., Pipelidis, G., & Tsiamitros, N. (2019). Hermes: An open and transparent marketplace for IoT Sensor data over distributed ledgers. In *2019 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)* (pp. 167–170). Seoul. <https://doi.org/10.1109/BLOC.2019.8751331>
- Urbach, N., Ahlemann, F., Böhm, T., Drews, P., Brenner, W., Schaudel, F., & Schütte, R. (2019). The impact of digitalization on the IT department. *Business & Information Systems Engineering*, 61, 123–131. <https://doi.org/10.1007/s12599-018-0570-0>
- van den Broek, T., & van Veenstra, A. F. (2015). Modes of governance in inter-organizational data collaborations. In *Twenty-Third European Conference on Information Systems (ECIS)* (pp. 1–12).
- VETRI. (n.d.). VETRI – VALUE YOUR DATA. Retrieved April 7, 2020, from <https://vetri.global/wp-content/themes/vetri/documents/whitepaper.pdf>
- Vlachos, M., Schneider, J., & Vassiliadis, V. G. (2015). On data publishing with clustering preservation. *ACM Transactions on Knowledge Discovery from Data*, 9(3), 1–30. <https://doi.org/10.1145/2700403>
- W3C. (2020). Data Catalog Vocabulary (DCAT) - Version 2. Retrieved December 31, 2020, from <https://www.w3.org/TR/vocab-dcat/>
- Watson, H. J., Fuller, C., & Ariyachandra, T. (2004). Data warehouse governance: Best practices at Blue Cross and Blue Shield of North Carolina. *Decision Support Systems*, 38, 435–450. <https://doi.org/10.1016/j.dss.2003.06.001>
- Weber, K., Otto, B., & Österle, H. (2009). One size does not fit all---a contingency approach to data governance. *Journal of Data and Information Quality*, 1(1), 1–27. <https://doi.org/10.1145/1515693.93.1515696>
- Were, V., & Moturi, C. (2017). Toward a data governance model for the Kenya health professional regulatory authorities. *The TQM Journal*, 29(4), 579–589. <https://doi.org/10.1108/TQM-10-2016-0092>
- World Economic Forum. (2020). Share to gain: Unlocking data value in manufacturing.
- Zenome. (2017). The Zenome Project: Whitepaper – blockchain-based genomic ecosystem. Retrieved September 16, 2019, from <https://zenome.io/download/whitepaper.pdf>
- Zhang, R., Indulska, M., & Sadiq, S. (2019). Discovering data quality problems: The case of repurposed data. *Business & Information Systems Engineering*, 61, 575–593. <https://doi.org/10.1007/s12599-019-00608-0>
- Zuiderwijk, A., Loukis, E., Alexopoulos, C., Janssen, M., & Jeffery, K. (2014). Elements for the Development of an Open Data Marketplace. In *Conference for E-Democracy and Open Government (CeDEM14)* (pp. 309–322). Krems an der Donau.

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.