



Data ownership revisited: clarifying data accountabilities in times of big data and analytics

Martin Fadler & Christine Legner

To cite this article: Martin Fadler & Christine Legner (2021): Data ownership revisited: clarifying data accountabilities in times of big data and analytics, Journal of Business Analytics, DOI: [10.1080/2573234X.2021.1945961](https://doi.org/10.1080/2573234X.2021.1945961)

To link to this article: <https://doi.org/10.1080/2573234X.2021.1945961>



© 2021 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 04 Aug 2021.



Submit your article to this journal [↗](#)



Article views: 1672



View related articles [↗](#)



View Crossmark data [↗](#)

Data ownership revisited: clarifying data accountabilities in times of big data and analytics

Martin Fadler and Christine Legner

Faculty of Business and Economics (HEC), University of Lausanne, Switzerland

ABSTRACT

Today, a myriad of data is generated via connected devices and digital applications. In order to benefit from these data, companies have to develop their capabilities related to big data and analytics (BDA). A critical factor that is often cited concerning the “soft” aspects of BDA is data ownership, i.e., clarifying the fundamental rights and responsibilities for data. IS research has investigated data ownership for operational systems and data warehouses, where the purpose of data processing is known. In the BDA context, defining accountabilities for data is more challenging because data are stored in data lakes and used for previously unknown purposes. Based on four case studies, we identify ownership principles and three distinct types: *data*, *data platform*, and *data product ownership*. Our research answers fundamental questions about how data management changes with BDA and lays the foundation for future research on data and analytics governance.

ARTICLE HISTORY

Received 29 November 2020
Accepted 14 June 2021

KEYWORDS

Data ownership; data governance; analytics governance; big data and analytics; data lake; data products

1. Introduction

There is no doubt that data are leading to a rising new economy (The Economist, 2017) and are fundamentally changing how business is conducted (Davenport et al., 2012; Wamba et al., 2015). With decreasing computing costs and the myriad of data generated via connected devices and digital applications, enterprises are seeking opportunities to improve existing processes and products as well as to develop new data-driven business models (Wixom & Ross, 2017). This goes along with improving their capabilities to manage big data and analytics (BDA) (Grover et al., 2018). A cornerstone of BDA is data lakes, which store large volumes of data in various formats and enable innovation through data exploration and experimentation (Farid et al., 2016; Madera & Laurent, 2016; Watson, 2017). Since data are nonrival, the business potential scales with data being used for multiple purposes at the same time without losing their value (Jones & Tonetti, 2019). However, this idiosyncrasy and the increasing number of data consumer–provider relationships leads to complexity in data ownership. While there is consensus that data ownership clarifies fundamental rights and responsibilities for data (Hart, 2002), the related debates in practice and research view the concept from different, often contrasting disciplinary perspectives. The legal perspective is reflected in the increasing number of data privacy regulations that governments issue to give individuals more rights and to control businesses’ uses of personal data (Labadie & Legner, 2019). Economists emphasise

that data ownership affects and potentially harms social welfare (Jones & Tonetti, 2019). In IS literature, data ownership is often cited as a critical factor concerning the “soft” aspects in the creation and use of enterprise data, specifically BDA. Data ownership is not only important to gain business value from big data (Alexander & Lyytinen, 2017; Comuzzi & Patel, 2016; Grover et al., 2018); it also clarifies fundamental rights and responsibilities that underpin data governance (Loshin, 2001; Winter & Meyer, 2001). Grover et al. (2018) emphasise: “*governance that delineates responsibility and accountability for data, [is a catalyst] for BDA value creation*” (p. 417).

Data ownership has been discussed since electronic data processing began (Maxwell, 1989; Spirig, 1987; Van Alstyne et al., 1995; Wang et al., 1995). The focus of the subsequent debates has been on data ownership for operational systems and data warehouses (Winter & Meyer, 2001), where the purpose of data processing is known. While we can assume that data ownership is still beneficial in today’s corporate environment, practitioners underline that data lakes require a different approach to data governance (Chessell et al., 2018). Defining accountabilities for data is more challenging for BDA because data are stored in data lakes and used for new, previously unknown purposes. When data are repurposed, data flow across organisational units and need to satisfy different data consumer’ requirements in terms of data format, granularity, and quality. Such cross-unit data flows require effective coordination, as emphasised by the concept of enterprise-wide

information logistics (Dinter, 2013; Winter, 2008). These developments raise the question how we need to reinterpret and apply data ownership concepts so as to cope with emerging challenges in BDA environments.

To address this gap, our objective is to understand how data ownership concepts change in the context of BDA. Thus, we ask:

RQ: *How do enterprises define and adapt data ownership in the big data and analytics context?*

To integrate academic and practitioner perspectives, we performed an extensive literature review and conducted explorative research based on multiple case studies to explore data ownership in the real-world context (Benbasat et al., 1987; Yin, 2003). From our analysis of the literature and of four companies with significant BDA experience, we identify data ownership principles and three data ownership types: *data*, *data platform*, and *data product*. We also demonstrate the implications of data repurposing on data ownership assignment and data dependencies. Our findings extend the prevailing data ownership concept from IS literature by integrating the *data platform* perspective, which serves as the required mediator between data supply (*data*) and data demand (*data product*) in BDA environments. Our insights into ownership contribute to the data and analytics governance literature generally. They particularly address structural aspects of data governance according to Tallon et al. (2013) and help clarify the decision rights in Tiwana et al. (2013)'s *IT Governance Cube*. Based on Grover et al.'s (2018) research framework, our study lays the foundation for *BDA governance to facilitate the value creation process*. Our findings also complement prior research on enterprise-wide information logistics (Dinter, 2013; Winter, 2008), by adding the perspective of data ownership to cross-unit information flows. The three data ownership types support the effective coordination of enterprise-wide information delivery in order to generate synergies and attain overarching goals.

The remainder of this paper is structured as follows: We start by reviewing the research field on data ownership from different disciplinary perspectives and outline the research gap. We then motivate our qualitative research approach and provide an overview of the research process. Third, we present each case in detail. Based on our cross-case analysis, we synthesise our findings into six propositions. We conclude with a summary and discussions of our contributions as well as an outlook on future research.

2. Background

Data ownership is grounded in the general concept of ownership, which is a fundamental mechanism in our society and can relate to different disciplinary lenses,

including legal, economics and management. Accordingly, different paradigms can be applied to determine who could or would be entitled to claim ownership of data. In the IS field, data ownership has been studied since the early days of electronic data processing, resulting in data ownership principles for operational systems and data warehouses. With BDA, an increasing variety of data sources are used for new, previously unknown purposes and are stored in data lakes so as to enable data exploration and experimentation. This requires us to revisit the data ownership concept for the BDA context.

2.1. Relevance of data ownership from different disciplinary perspectives

Ownership is a fundamental concept that is grounded in our everyday life and in fundamental mechanisms of society (Shleifer, 1998). It denotes the assignment of rights and responsibilities for a property to an individual or an organisation: "*Property rights [...] are the rights of ownership. In every case, to have a property right in a thing is to have a bundle of rights that defines a form of ownership (Becker 1980, pp. 189–190)*" (as cited in (Hummel et al., 2020), p. 3). These rights can apply to material and immaterial objects alike (ibid). Independent of the underlying object, the concept of ownership links various research disciplines among them law, economics, or management. In each of these disciplines, data ownership is discussed with varying objectives (see Table 1).

In law, data ownership is mostly associated with the privacy of individuals. With personal identifiable information being collected in an ever-increasing volume by large tech companies, this discipline aims at defining the actual owner of this data collection and the extent of control that remains with the data's subjects. This legal perspective is particularly important as companies must be held accountable when it comes to data leakages or alienation of use that can harm data's subjects as happened in the Cambridge Analytica scandal (Confessore, 2018). Although some governments are introducing privacy regulations to give individuals more rights and to control businesses' uses of personal data (Labadie & Legner, 2019; Tikkinen-Piri et al., 2018), the dominant legal view remains that data cannot be owned (Hummel et al., 2020). Nonetheless, contractual and intellectual property laws have to be respected for governing data in different situations (ibid). They put forwards that data property rights can be transferred through a licence agreement or that data property rights are obtained through mere creation (ibid).

In economics, property rights for data are defined as the ability to control the amount of data collected and to monetise it (Dosis & Sand-Zantman, 2019, pp. 3–4). With the recent explosion of data,

economists are seeking answers on “*how different property rights for data determine its use in the economy, and thus affect output, privacy, and consumer welfare*” (Jones & Tonetti, 2019, p. 2819). Inherent to the economic perspective are data’s unique characteristics as nonrival goods. In contrast to most other goods, data thereby are infinitely usable and are the source of increasing returns for companies (Jones & Tonetti, 2019). This characteristic can have negative economic consequences in cases where property rights for data are wrongly distributed. First, firms may not adequately respect the privacy of consumers (ibid). Second, firms may hoard data and limit potential gains of data being broadly used (ibid). Finding the optimal allocation of property rights for data therefore remains an open quest. Interestingly, a recent study by Dosis and Sand-Zantman (2019, p. 32) finds that the optimal allocation of rights crucially depends on the value of the data, or equivalently on the relative weight between the market in which the data are generated and the market in which they are used. Notably, there are already initiatives that drive open access to data (e.g., open data) (Link et al., 2017) and to machine learning models (e.g., Open AI) (Open AI, 2020) which directly stimulates reuse and thus generates value.

In management, ownership rights are an important element of corporate governance that guarantee the mere survival of organisations. Recent studies argue that property rights of a company should be assigned in a way that increases a company’s overall market value (Schulze & Zellweger, 2020). Here, a company is the owner of data that it collects or creates, while the companies’ property rights holders are undertaking the inherent risk of this venture. Linked to this perspective are also the separation and delegation of different decision rights to manage an organisation’s inherent complexity and achieve a desirable outcome (Fama & Jensen, 1983; Winkler & Wessel, 2018). In their seminal study, Fama and Jensen (1983) view a company “*as a nexus of contracts (written and unwritten)*” between different agents (p. 321). As implication, an effective system for decision control implies, almost by definition, that the control (ratification and monitoring) of decisions is to some extent separate from the management (initiation and implementation) of decisions (ibid, p. 304). Besides the general differentiation of decision rights and their separation, it remains important to understand for what object (material or immaterial) a certain decision is made. This question is further studied in the corresponding sub-disciplines of management research, for instance, in the IS discipline.

In IS, early studies investigate how the allocation of data ownership affects system success (Maxwell, 1989; Spirig, 1987; Van Alstyne et al., 1995; Wang et al., 1995). Although the authors use the term “*data*

ownership”, they do not interpret “*ownership*” in the same way as the other disciplines mentioned earlier. Data ownership in the context of IS governance is decision control rights rather than property rights (as in the economic or management perspectives). For instance, in their seminal paper, Van Alstyne et al. (1995) distinguishes between *ownership* as the residual right of control (i.e., the right to determine access privileges for others), and *usage rights* as the ability to access, create, standardise, and modify data as well as all intervening privileges (p. 8). Allocating decision control rights on data has a direct effect on system implementations. Several studies confirm that data ownership should always stay with its origin (i.e., where the data are created) to ensure system success (Maxwell, 1989; Spirig, 1987; Van Alstyne et al., 1995; Wang et al., 1995). While this logic sounds intuitive, its practical implementation remains complex, especially in analytical information systems where data flow across organisational units (Dinter, 2013).

2.2. Data ownership paradigms – how to assign the data owner?

In the enterprise context, data ownership provides the underpinning principles for data governance to define roles, responsibilities, and processes (Loshin, 2001; Winter & Meyer, 2001). Grover et al. (2018, p. 417) argued that “*without appropriate organizational structures and governance frameworks in place, it is impossible to collect and analyze data across an enterprise and deliver insights to where they are most needed*”. The assignment of certain ownership rights to roles has proven to be beneficial: most importantly, people feel responsible, act in their self-interest, and take care of data. Thus, data ownership has been found to positively impact on data quality and system success (Loshin, 2001; Van Alstyne et al., 1995). While the assignment of ownership rights and responsibilities has clear advantages, it can also lead to conflict concerning data sharing (Hart, 2002).

Generally, the allocation of data ownership is a “*control issue – control of the flow of [data], the cost of [data], and the value of [data]*” (Loshin, 2001, p. 28). Since responsibilities can depend on its context of use, Loshin (2001) explored different data ownership paradigms. Although Loshin (2001) followed a fairly pragmatic approach, the suggested paradigms can be linked to different general philosophical ownership approaches outlined by Hart (2002). These approaches can help us to understand the underlying rationale for assigning ownership as well as to structure the research field (see Table 2). We classify the paradigms according to the socio-organisational context into three categories: individual, organisational, and shared ownership (everyone). We will now present each category.

2.2.1. Individuals as data owner (data ownership outside of the organisation)

Data ownership is increasingly being claimed by individuals as the subjects of data (*subject as owner*). This paradigm reflects *libertarian theory* by Robert Nozick and John Rawls, where ownership must be allocated in ways that do not limit the freedom of others to act autonomously (Hart, 2002). With the Internet, personal data are being collected, used, and even sold in non-transparent ways. Thus, the private ownership paradigm often emerges as a reaction once the data collection has been unveiled, and *individual data ownership* rights are increasingly enforced with data protection policies such as the European Union's General Data Protection Regulation (GDPR). With the emergence of the Internet of Things (IoT), the debate about individual data ownership has gained a new facet because it remains unclear who owns personal data produced by machines (Janeček, 2018). For instance, the data collected by smart metres enable electricity providers to optimise their network and service offerings, but also unveil highly sensitive data about private households, which can easily be misused (McKenna et al., 2012).

2.2.2. Organisations as data owner (data ownership inside of the organisation)

In the context of organisations (*enterprise as owner*), the data ownership concept is getting more complex as a result of distributed data creation and processing in organisations (Van Alstyne et al., 1995). Here, three reasons for claiming ownership can be distinguished. First, organisations claim ownership owing to monetary factors of funding (*funding organisation as owner*) or purchasing/licencing data (*purchaser/licensor as owner*). These paradigms build on *labour theory* by John Locke and assign ownership according to the extent of value added through labour (Hart, 2002). They always involve two parties: the organisation that funds the party who creates data, and the organisation that purchases or licences data owned by another party. While in the first case data ownership is transferred to the funding organisation without any restrictions, in the second case, data ownership is transferred to the purchasing/licencing party under certain restrictions. Second, an organisation may claim ownership by using data. This approach reflects the view of first *occupancy theory* by Immanuel Kant, which assigns ownership to the first who possesses a property or object (Hart, 2002). This is typically the case for consuming parties (*consumer as owner*) that require high confidence in the data and therefore take over accountability. It may also apply to parties who read data from different sources (*reader as owner*) to create or add these to their knowledge base. Third, organisations create business value through data processing and therefore claim ownership. In line with

personality theory by Georg Wilhelm Friedrich Hegel, it determines ownership by a person's will to invest in an object, which makes him this object's owner (Hart, 2002). Four paradigms can be distinguished depending on the processing type: creating data (*creator as owner*) or formatting data (*packager as owner*) for a certain purpose, compiling information from various data sources (*compiler as owner*), and decoding data (*decoder as owner*).

2.2.3. Everyone as data owner

Data ownership often implies that an individual or organisation has sole ownership rights. The opposite is the case in the paradigm *everyone as owner*, which is applied when data are intended to be shared with a broad user group. In this case, data ownership is not assigned to any individual or organisational party; instead, everyone can become an owner of certain data, and with the same access rights. This paradigm builds on *utility theory* by Jeremy Bentham and John Stuart Mill, where ownership maximises the benefits for all involved parties (Hart, 2002). It is often emphasised in discussions related to open data, which is "data that anyone can access and use" (Link et al., 2017). Especially when the data are created in a crowdsourced way – as is the case with OpenStreetMap (OpenStreetMap, 2019), for instance, – the community is the data owner and everyone shares the same rights to access and use the data, under certain restrictions. Still, open data repositories require data governance, which is often hard to establish when responsibilities are distributed, and accountabilities cannot be assigned to an individual or organisational entity. This is especially the case with public health data, but also with data collected in smart cities, for instance. Thus, while open data hold the potential for great innovation, issues develop around privacy, confidentiality, and control of data (Kostkova et al., 2016).

2.3. Approaches to data ownership for operational and analytical systems

Data ownership has been specifically investigated for operational systems (Maxwell, 1989; Spirig, 1987; Wang et al., 1995) and data warehouses (Winter & Meyer, 2001). Operational systems seek to enable business processes with quality data, defined as data that fit its purpose (Wang & Strong, 1996). Enterprises have sought to centralise operational systems to ease maintenance and control for IT departments. This has resulted in a misconception that IT departments are the data owner and must be responsible for data quality (Van Alstyne et al., 1995). Business users create the data while executing business processes, but also need high confidence (quality) in the data they use. Thus, in operational systems, it is recommended that data

ownership holds to its original aim of ensuring high data quality (Maxwell, 1989; Spirig, 1987). This implies that the data ownership paradigms *creator as owner* and *consumer as owner* fall together.

While data ownership in operational systems follows the logic of business processes, data warehouses and particularly data marts (in the means of analytics systems) integrate data from multiple business processes (Watson & Wixom, 2007). Data warehouses bring together data from operational systems (*push*). To fulfill a certain information demand (e.g., management report), data are integrated for this particular use in data marts (*pull*). Thus, data ownership in data warehouses and data marts must be data-centric and depends on the number of data integration layers. In the case of one data warehouse and one data mart layer, two ownership types can be distinguished (Winter & Meyer, 2001). Since data are typically not changed when it is brought into a data warehouse, data ownership on the data warehouse layer stays the same as in operational systems (*data supply*). On the data mart layer, data are typically changed to fulfill a certain information need. Thus, data ownership on this layer is assigned to the party who requests particular information (*data demand*), which is often also the sponsor of such activities.

In the context of analytical information systems, data are used in different organisational units than from which they originate (Dinter, 2013; Winter, 2008). The resulting data supply issues have been discussed from the perspective of information logistics, i.e., “*the planning, control, and implementation of the entirety of cross-unit data flows as well as the storage and provisioning of such data*” (Winter, 2008, p. 41). Hereby, data ownership, in the form of governance structures (Dinter, 2013) enables efficient and effective information delivery.

2.4. The research gap

Debates about data ownership have multiple facets and, with increasing privacy concerns, they go well beyond the boundaries in which data are created. In the enterprise context, data ownership remains more complex compared to other assets. Still, data ownership is needed to clarify rights and responsibilities to ensure business value with effective data governance (Grover et al., 2018; Otto, 2011; Tallon et al., 2013). The research distinguishes two approaches to data ownership: In operational systems, data ownership is business process-centric, i.e., the creator and the consumer of operational data are often the same. This perspective stands in contrast to analytical systems (e.g., data warehouses), where data ownership is data-centric: the consumer is not the creator because a data mart integrates data from multiple business processes. IS

research on data ownership has focused mostly on operational systems although even more managerial challenges emerge in the context of analytical systems (Dinter, 2013; Winter, 2008). To the best of our knowledge, only one early study elaborates specifically on data ownership in data warehouses (Winter & Meyer, 2001). A few studies investigate related topics, such as data governance in the context of data warehousing (e.g., Watson et al., 2004) or governance mechanisms for data analytics (Baijens et al., 2020), data quality management (e.g., Weber et al., 2009), and data lifecycle management (Tallon et al., 2013).

BDA as an emerging analytical paradigm differs from traditional business intelligence and data warehouse infrastructures, where the structure is predefined, and data are cleaned upfront to deliver high-quality reports and insights (Watson, 2009). BDA introduces larger volumes and a higher variety of data that are stored in data lakes, without a predefined structure and in raw format, to enable data exploration and innovation (Farid et al., 2016; Madera & Laurent, 2016; Watson, 2017). With this paradigm shift, new challenges emerge for enterprises (Grover et al., 2018; Sivarajah et al., 2017). On the one hand, with data repurposing, they need to manage an increasing number of data provider–consumer relationships. Providing data for multiple purposes (Chen et al., 2012) imposes higher requirements on data quality, data integration, and data security (Grover et al., 2018). In fact, data quality remains one of the key challenges to enable business value from BDA (Abbasi et al., 2016; Grover et al., 2018; Wamba et al., 2015). On the other hand, the development and operation of analytics go beyond the mere aggregation and visualisation of data. With artificial intelligence (AI) (Watson, 2017), it is harder to keep track of how data are processed. Further, the high dependency of machine learning applications on data may lead to the risk of high technical debt (Sculley et al., 2015). At the same time, the increasing use of AI is fuelling debates about ethical questions. For instance, deep learning techniques operate as “*black box*” algorithms whose working mechanisms are somehow hard to understand (Castelvecchi, 2016). This is why analytics can lead to “*discriminatory effects and privacy infringements*” (Custers, 2013, p. 3) and why debates have emerged about accountabilities for algorithmic decision-making (Diakopoulos, 2016).

These developments are resulting in new issues and questions relating to data ownership, while showing the relevance of defining accountabilities for data. Besides the consideration of these contemporary requirements in research on accountabilities, a holistic view on data governance, which comprises operational and analytical systems, is currently missing.

3. Methodology

We seek to understand *how enterprises define and adapt data ownership in the BDA context* – a complex phenomenon that requires that one analyse rich information related to the adoption of BDA and the definition of data-related roles in enterprises. This is why we opted for an explorative case study research design, which is well suited for answering *how* questions (Yin, 2003) and studying such contemporary phenomena in their particular context (Benbasat et al., 1987; Yin, 2003). Specifically, we studied multiple-case studies so as to ensure our theory's robustness and to draw generalisable conclusions (Benbasat et al., 1987; Yin, 2003).

3.1. Case selection

We integrated our research activities into a research programme on data management that included close interactions with 11 data management experts from seven high-profile European companies over 12 months. In early 2019, we initiated an expert group to investigate data management challenges in the context of BDA and met 14 times between January and November 2019. The participants were data experts responsible for establishing organisational and technological structures to manage BDA. They represent large corporations from different industries with some maturity in leveraging BDA.

The discussions in the expert group allowed us to develop an understanding of the current situation and to select four (out of seven) companies for further investigation (see Table 3-7). Three companies were discarded because their data lake initiative was only in the pilot phase, and they had limited practical experience with data ownership in BDA environments. The selected four companies had already established an enterprise data lake and had practical experience with introducing data and analytics roles, including the data ownership concept. As each case company has a high BDA maturity and belongs to a different industry, the case selection process followed literal replication logic, leading to similar rather than contrasting results (Benbasat et al., 1987; Yin, 2003).

3.2. Data collection

Our data collection approach aimed at gathering information from multiple sources, including expert interviews and internal documents, to allow for triangulation and ensure construct validity (Yin, 2003). For the expert interviews, we selected key informants that have strategic and operational responsibility to manage BDA and who are aware

of the relevance of and issues relating to data ownership. For identifying the experts, we used a snowball sampling approach (Naderifar et al., 2017): We were already in contact with at least one key informant for the data lake initiative in the respective company through the expert group that we formed (see above). We requested them to identify further key informants in case our requirements were not met. Thereby, we interviewed at least two experts per company, which were knowledgeable about BDA platforms, roles and accountabilities. At least one expert was working in the company for more than five years to ensure a solid understanding of the company's strategic initiatives and challenges. As a starting point, we conducted one initial semi-structured interview of 1–1.5 h with the key informants to understand each's technological and organisational structures to manage BDA. For instance, we asked the open-ended questions “*What is the architectural structure of your data lake?*”, “*What are your key accountabilities for managing data on the data lake?*” and “*How do you assign those accountabilities?*”. These interviews gave us the opportunity to understand the challenges and approaches concerning assigning accountabilities for data in greater depth. In parallel, we collected primary data through internal documents provided by the firms (e.g., BDA platform designs, role models, and organisational structures). These documents informed us not only about their approach to data ownership but also about the context and related topics, such as technical infrastructure as well as established roles or processes.

3.3. Within- and cross-case analysis

We performed the case analysis in two steps. First, we conducted a within-case analysis (Yin, 2003) to understand the different data ownership types in each enterprise. Here, we used an analysis framework and documented the company-specific data ownership types, their descriptions, and the organisational assignment in each type based on the interview transcripts and the additional company documents provided. In a subsequent expert group meeting, we discussed and compared each company's data ownership approach. The discussion helped us to understand the similarities and peculiarities of each case. Second, we performed a cross-case analysis (Yin, 2003), comparing the findings of the within-case analysis with one another so as to identify common data ownership types and their responsibilities. Further, we linked each identified type to the corresponding data ownership paradigms suggested by Loshin (2001), which helped us to understand its mechanism in

a simplified way. Based on our analysis, we outlined four propositions for data ownership in the BDA context. We discussed our findings in another expert group meeting, which gave us a better understanding of whether the enterprises agreed with our conclusions or if we had missed aspects we had not reflected on. To verify specific aspects with the case companies and to ensure robust findings, we conducted an additional interview with one key informant from each company. At the end, we held another expert group meeting to discuss common challenges resulting from data repurposing and derived two further propositions.

4. Data ownership in the four case companies

To provide insights into the case setting, we start by presenting the general context, i.e., BDA’s role in each enterprise and each’s approach to data ownership.

4.1. Company A

Company A is undergoing a digital transformation and is introducing innovative digital products (in addition to its traditional product portfolio), which shifts its core business model from business-to-business to business-to-consumer. Through this change, the company faces an increasing amount of data created via sensors embedded in the digital product and in new customer touchpoints (e.g., points of sale or web applications). These data are enabling company A to improve the way it understands and interacts with its customers; but, to lever these data, the company had to enhance its data and analytics capabilities. In a first step, it formed a central group that is responsible for enterprise data and analytics. It also established a data lake as a central big data platform (commercialised Hadoop stack from Cloudera, on-premise and partially in the cloud), which enables data scientists to conduct analytics across the traditional business functions based on internal and external datasets. This platform is primarily used for exploration and experimentation, but also for industrialisation of analytics use cases. It has three major components: the data repository for storing and

staging data from internal and external sources, data science labs for exploration and experimentation, and data products for industrialisation of analytics use cases.

Company A distinguishes three data ownership types: data source owner, platform owner, and data product owner. The data source owner is “*primary decision maker about the data entities under his responsibility and accountable for the overall integrity, data lifecycle and data quality of data created in his ownership*”. This role is typically assigned at a director level or even above, to the head of a business function that creates but also consumes data of this domain. In the data platform context, the data source owner “*provides approval for data usage in data product*”. Thus, company A ensures compliant access to sensitive data (e.g., identifiable personal information). When data are then used in a data product, the company arranges a service-level agreement with the corresponding owner of the data sources so as to ensure quality on both sides. Thus, the data source owner must “*fulfill service-level agreements for data products*”. The platform owner is accountable for the platform infrastructure (technology stack) and is assigned to the head of the digital analytics team. Concerning data, he “*maintains data sanity and business context while data are going through the technology stack*”. This includes that he “*oversees and controls work in data labs*”. Further, he “*is accountable for the availability of data pipelines*”. In this sense, he must ensure that business requirements for data products are being fulfilled. The data product owner, as a head of a business function, represents the data use side and “*addresses business need for data driven by analytics use cases*”. This makes him “*accountable for output of the technology stack*”. Once a data product is developed and ready to use, he “*ensures the business value of a data product over its lifetime*”.

4.2. Company B

Company B is an infrastructure provider. It is undergoing a digital transformation following a corporation-wide program with three main goals: improve interactions with customers, increase internal efficiency, and enhance capacity management. Thus, the company has invested in

Table 1. Disciplinary perspectives on ownership and data ownership.

Discipline	Ownership concept	Objectives of data ownership
Law	Enablement and protection of rights with respect to one’s property (external) and identity (internal).	Ensure data privacy, while holding firms accountable for fraudulent data use.
Economics	Allocation of ownership rights for economic goods and their effect on market equilibriums as well as welfare.	Distribute property rights for data in the way that increases output and consumer welfare, while protecting individual privacy.
Management	Allocation of property rights to maintain an organisation’s survival and increase its value (Firm is owner of the data it collects and creates)	Define a firm’s accountability and assess its risk undertaking through data collection and monetisation.
Management information systems	Allocation of decision rights for IT artefacts to achieve a desired outcome.	Assign decision control rights for data among different organisational entities to increase value generation through data.

Table 2. Data ownership paradigms and discourses.

The socio-organisational context	The data ownership paradigm (Loshin, 2001)	Example	The related philosophical perspective on ownership (Hart, 2002)
Individual	Subject as owner	A private person accuses a company of selling his or her personal data to a third party	Libertarian theory: Ownership does not limit the freedom of others
Organization	Consumer as owner	A sales team uses customer phone numbers that are essential for its daily operation	First occupancy theory: Ownership by being the first to possess an object
	Reader as owner	A consultancy collects information on industry trends to extend its knowledge base	
	Enterprise as owner	An enterprise creates, processes (adds value), and distributes data about its products	
	Funding organisation as owner	A company pays a research company to collect panel data	
	Purchaser/Licensor as owner	A company buys an address list of potential customers	
	Creator/Generator as owner	A research firm invests in collecting qualitative data for a market study	Personality theory: Ownership through personal will to invest in an object
	Compiler as owner	A business intelligence department builds a central data warehouse	
	Packager as owner	A web agency designs and formats a web page for a customer	
	Decoder as owner	A company synthesises information from DNA data	
Everyone	Everyone as owner	A crowdsourced collection of geo-information in a public database	Utility theory: Ownership maximises the benefits for all involved parties

Table 3. Selected cases.

Case name	Industry	Size	Key informants (years in the company)	Big data and analytics context
Company A	Fast-moving consumer goods	Revenue: 50 USDB to 100 USDB Employees: ~80,000	Manager: data governance (>10y), Enterprise data architect (1–5y)	<u>Organisation</u> : central data and analytics management organisation <u>Infrastructure</u> : central big data platform for innovation and industrialisation of analytics use cases
Company B	Public transportation and mobility infrastructure	Revenue: 1 USDB to 50 USDB Employees: ~35,000	Leader: Business information management (>10y), Data governance manager (6–10y), Big data platform architect (1–5y)	<u>Organisation</u> : central data management organisation and central/decentralised data science team <u>Infrastructure</u> : corporate data lake for data exploration/experimentation and the operation of analytics use case
Company C	Manufacturing	Revenue: 1 USDB to 50 USDB Employees: ~90,000	Director: Data architecture and engineering (6–10y), Project manager: Data platform (3y)	<u>Organisation</u> : corporate data management organisation and central platform team <u>Infrastructure</u> : central data platform to enable digital innovations and scale the operation of data products
Company D	Healthcare and life science	Revenue: 1 USDB to 50 USDB Employees: ~50,000	Leaders: Head of Data Products and Solutions (>10y), Global Enterprise Data Strategy Lead (1–5y)	<u>Organisation</u> : federated organisation with data and analytics centre of excellence and staff in line of business <u>Infrastructure</u> : Multiple data platforms serving specific analytics needs and an enterprise-wide data platform

Table 4. Data ownership in company A.

Data owner type	Description	Organizational assignment
Data source owner	<i>“Primary decision-maker about the data entities under his responsibility and accountable for the overall integrity, data lifecycle and data quality of data created in his ownership”. “Provides approval for data usage in data product”. “Fulfils service-level agreements for data products”.</i>	Head of a business function: director level or above
Platform owner	<i>“Maintains the data sanity and business context while data are going through the technology stack”. “Oversees and controls work in data labs”. “He is accountable for the availability of data pipelines”.</i>	Head of the digital analytics team
Data product owner	<i>“Addresses the business need for data driven by analytics use cases”. “Accountable for the output of the technology stack”. “He ensures business value of data product over its lifetime”.</i>	Head of a business function: director level or above

new digital applications and sensor technologies to collect data from its assets. Further, it provides noncritical data to third parties through open access so as to stimulate innovation from the outside. Advanced and big data analytics are key drivers of company B's digitalisation initiative and are strategically relevant to the company. Thus, it established a central big data platform (commercialised Hadoop stack from Cloudera, on-premise) to provide access to data from diverse sources simultaneously for innovation and production. To ensure the reusability of data on the platform, it was decided that data must be actively managed through corresponding organisational roles and structures. A central data management organisation was established to ensure data governance. On the analytics side, a central data science team coordinates the activities, while data scientists form part of each business unit. The platform has four major components: data lake, data labs, data apps, and user homes. The data lake serves as an underlying data storage and processing entity that operates along a staging, an integration, and a business transformation layer. Data labs operate on the data lake and serve the data scientists' need to explore and experiment with data, for instance, a group of data scientists is accessing machine state data in a data lab to develop a predictive maintenance algorithm. The data app represents an operationalised application that uses data from the data lake, for instance, the predictive maintenance application signals service workers in case of required maintenance activity. A user home comprises specific data from the data lake that is private to the user, for instance, a business analyst conducts ad hoc analyses of daily customers.

Company B distinguishes three of data ownership types on the big data platform, according to its components: *data owner*, *owner of the data lab/data app/user home*, and *owner of the data lake*. The *data owner* is responsible for a data feed in the context of the big data platform and is typically assigned to a business role. Thus, this role is "*responsible for data quality, definition, classification, security, compliance and data lifecycle of a data attribute, set of attributes, or dataset*". The data definition (e.g., documentation in data catalogue) and classification must be done when data are brought to the big data

platform. This implies that the data owner "*controls reading access to his data through data feed on big data platform and ensures compliant use through the provision of no-join policies under the respect of interests of existing and future data user*". These policies must be revisited as new data are continuously brought to the platform. Since not every data feed has a *data owner* assigned when it is brought to the big data platform, the data user is required to find the *data owner*. If the *data owner* cannot be identified, the user must fill this gap and becomes the owner of the requested data. The *owner of the data lake* is "*accountable for the standardisation of the overall big data solution architecture*". This includes that he "*proves the compliance of analytics solutions*". Thus, this role is assigned to the role of the *big data solution architect*, who is also responsible for platform development and provides "*information on planned extensions of the data lake*". This role's responsibilities go beyond the architecture of the big data platform, since he "*ensures that new and valuable data are onboarded to the data lake according to the business need and potential. For this, he searches proactively new data sources, values their business potential, and initiates the onboarding process*". In this regard, the *owner of the data lake* serves as a mediator between the *data owner* and the *owner of the data lab/data app/user home*. The latter holds the rights to use data either through a data app that is typically assigned to a business role or through a data lab or user home that is typically assigned to technical roles, for instance, a data scientist. This owner also "*manages access to data lab, app, or user home and is accountable for any activity (operational activity or data privacy) on it over its lifetime*". He is also obliged to inform the *platform owner* about whether the environment still generates value or can be removed. A data scientist, as a user of the *owner of the data lab*, "*needs to comply with a conduct of ethics when working with data in a data lab*".

4.3. Company C

Company C has a long tradition in the automotive industry. It has invested heavily in R&D to embed

Table 5. Data ownership in company B.

Data owner type	Description	Organizational assignment
<i>Data owner</i>	"Responsible for data quality, definition, classification, security, compliance, and data lifecycle of data attribute, set of attributes, or dataset". "Controls reading access to his data through data feed on big data platform and ensures compliant use through the provision of no-join policies under the respect of interests of existing and future data users".	Business role
<i>Owner of the data lake</i>	"Accountable for the standardisation of the overall big data solution architecture. Proves compliance of analytics solutions". "Gives information on planned extensions of the data lake". "Ensures that new and valuable data are onboarded to the data lake according to the business need and potential. For this, he proactively searches for new data sources, values their business potential, and initiates the onboarding process".	Big data solution architect
<i>Owner of the data lab/data app/user home</i>	"Manages access to the data lab, app, or user home, and is accountable for any activity (operational activity or data privacy) on it over its lifetime". "Data scientists must comply with conduct of ethics when working with data in a data lab".	Business role for the data app Technical role for the data lab/user home

software in its products to collect and process data. With this data, the company is seeking to monitor its products' conditions and to provide value adding services to its customers. Thus, it strongly relies on data as an essential component of its future business. For traditional data domains, it has established a corporate organisation for master data management. Owing to new requirements to manage sensor data and to develop analytics, company A has extended this function's scope and has set up new organisational units. A central platform team has been built up and manages a platform with a virtualised and physical data lake (Microsoft Azure Cloud) to enable digital innovations and to scale the operation of data products. Company C has also flattened its organisational hierarchies so as to become more agile. Its data platform has two major components: a data hub and data solutions. The data hub connects to the data sources and encompasses a physical and a virtual storage for various types and formats of data. The data solution accesses and processes data to develop/deliver a data application for/to a data consumer.

In the context of the data platform, company C distinguishes between three ownership types: *data domain manager*, *infrastructure owner*, and *business logic ownership*. The *data domain manager* "controls and monitors the data management for his domain". Each data domain comprises a homogenous set of data attributes describing a business object, for instance, a customer or an asset. This domain approach to structuring data ownership is a typical approach in organisations with mature data management practices. Company C's *data domain manager* "receives requests for data processing and provides data for data usage" and is accountable for data content and responsible for maintaining data according to business requirements. This role is assigned to a business role in lower management to ensure the efficient handling of requests, which corresponds to company C's agile management approach. Company C does not yet distinguish between the input and output data of a data application. Thus, the data domain manager is the owner of input data to the platform and output data of data applications as long as they belong to his

domain of responsibility. This includes reporting errors and suggesting improvements. The infrastructure owner is accountable for the data platform's development and operation. Thus, he "oversees the implementation and availability of data pipelines to onboard data to the data hub and provision data to data solutions". At company C, this role is assigned to the head of the data platform team, which is part of the corporate IT function. The *business logic owner* is "accountable for data applications over its lifetime, which includes compliant implementation, the maintenance of data application, and support of users". This role can either be assigned to a business or/and an IT role (central/decentral) depending on a data application's importance and complexity.

4.4. Company D

Company D is a long-lasting player in the healthcare and life-science industries and exist on the market since more than a century. As science and technology are at the core of this company, data and analytics have become major enablers for the company's ability to develop innovative products. Thus, an enterprise-wide data platform has been established that aims at data democratisation by capturing, curating, exposing, and understanding data to answer innovative business questions. This enterprise-wide platform comprises a wide array of capabilities among them are advanced analytics, text analytics, data lake, and a data catalogue. Data can be onboarded from internal operational (e.g., CRM) and analytical systems (e.g., data warehouse) as well as external data sources. The data catalogue is a central element of this platform that helps in coordinating data onboarding workflows (data supply) and simultaneously in finding relevant data (data demand).

Company D distinguishes four data ownership types in the context of the enterprise-wide data platform: *data owner*, *data product owner*, *business owner*, and *platform owner*. It makes a clear distinction between so-called *left-hand operations* and *right-hand operations* on the enterprise-wide platform. The "*left-hand operations are basically how you fill your data*

Table 6. Data ownership in company C.

Data owner type	Description	Organizational assignment
<i>Data domain manager</i>	"Controls and monitors the data management for his domain". "Receives requests for data processing and provides data for data usage". "Reports errors and suggests improvements".	Business role: lower management
<i>Infrastructure owner</i>	"Develops and operates the data platform". "Oversees the implementation and availability of data pipelines to onboard data to the data hub and to provision data to data solutions".	Corporate IT role: Head of the data platform team
<i>Business logic owner</i>	"Accountable for a data application over its lifetime, which includes compliant implementation, the maintenance of the data application, and support of users".	Business or/and IT role: lower management

Table 7. Data ownership in company D.

Data owner type	Description	Organizational assignment
Data owner	<i>"Data owner has a strong contributory role in governing the data in the means of their purposeful, compliant, ethical use as well as their quality". "Data owner is accountable for the careful delegation of responsibilities for supporting processes, systems and uses".</i>	Organizational unit that has primary use of a dataset
Data product owner	<i>"Data product owners [...] oversee a portfolio or bundle of use cases which might result or can be bundled to a product".</i>	Data organisation
Business owner	<i>"The data product owner is complemented with a business owner, someone who has skin in the game and makes sure that their staff are really using the data product, e.g., in digital sales".</i>	Line of business: team leader
Data platform owner	<i>"Data platform owner prioritises all the requirements coming from all areas and own the platform, and basically give the direction how the platform will develop further".</i>	Data organisation

catalogue and how you curate your data and organise it", the "right-hand operations are actually how you use that data for a certain purpose and that purpose is what we call product".

On the left hand, data owners are assigned to organisational entities that are the primary users of a specific dataset. The data owner "has a strong contributory role in governing the data in the means of their purposeful, compliant, ethical use as well as their quality". This role is accountable for delegating corresponding data responsibilities to data stewards and data custodians. Data owners are nominated for a dataset's *context of use* which defines the *Who*, *What*, *Why*, *Where*, *How*, and *When* respective the *terms and conditions* of a dataset's use. This context of use can be adapted when a dataset is used for a different use case, for instance. While small deviations of a dataset's context of use (e.g., a dataset is used as it is for creating a report) have no effect of its responsibilities, greater deviations (e.g., a dataset's attributes must be extended) may lead to defining a new context of use and nominating a dedicated data owner. The central data organisation acts as an intermediary and is responsible for assigning data owners and negotiating these contractual agreements.

On the right hand, "data product owners look after certain domains like sales, supply chain or marketing and they oversee a portfolio or bundle of use cases which might result or can be bundled to a product". Hence, the data product owner manages a portfolio of use cases and collaborates with data and analytics experts to bring these use cases on the platform. First, data need to be onboarded to the platform. This data acquisition "is driven by the data product/use case, first data stewards or data detectors find the data, then data engineers acquire the data and lead engineers organise the data". The data product owner is "complemented with a business owner, someone who has skin in the game and makes sure that their staff are really using the data product, e.g., in digital sales". So, while data product owners seek to transform use cases into products in the central data organisation,

business owners ensure that these products are actually used to generate value in the lines of business.

Besides the accountabilities for managing data supply and demand, case company D is at the moment establishing the data platform owner role who "prioritises all the requirements coming from all areas and own the platform, and basically give the direction how the platform will develop further". This role is also part of the data organisation.

5. Data ownership types and principles in the context of BDA

Through a cross-case analysis, our study unveils that BDA leads to significant changes and extensions to data ownership. In the following, we formulate propositions related to the three data ownership types and on the specific implications of data repurposing on data ownership.

5.1. Data ownership types

Proposition 1: In the context of BDA, companies define data ownership at three levels: data source or dataset (data supply), data product (data demand), and data platform.

Our cross-case analysis reveals that three different data ownership types were present in all four enterprises. These ownership types characterise relevant organisational data accountabilities and responsibilities in the context of BDA. They can be linked to the corresponding data ownership paradigm suggested by Loshin (2001) and the related philosophical assumptions (see Table 8).

Proposition 2: The data owner ensures compliant access to and use of data, not only in the source system, but also on the platform and in data products. This addition extends beyond the traditional responsibility of ensuring data quality and requires one to manage more data dependencies.

Table 8. Data ownership types in the context of big data and analytics.

Data owner type	Responsibilities	Support in cases	Exemplary statement
Data owner	Accountable for quality and lifecycle of data in his domain of responsibility.	A, B, C, D	"[...] accountable for the overall integrity, data lifecycle, and data quality of data created in his ownership". (A)
Data platform owner	Fulfills quality requirements for data in his domain of responsibility for data products. Ensures compliant access and use of data in his domain of responsibility by handling requests, providing access, and approving usage. Ensures data quality on the platform by managing data pipelines to onboard and provision data.	A, D A, B, C, D A, C	"Fulfills service-level agreements for data products". (A) "Controls reading access [...] ensures compliant use through the provision of no-join policies [...]". (B) "Oversees the implementation and availability of data pipelines to onboard data to the data hub and to provision data to data solutions". (C) "Ensures that new and valuable data are onboarded to the data lake according to the business need and potential". (B)
Data product owner	Accountable for onboarding of valuable data according to a business need and potential. Responsible for the development and operation of the data platform. Approves compliance of data products according to data platform standards. Ensures that a data product addresses a business need and generates business value over its lifetime. Accountable for a data product over its lifetime, including use case portfolio management, development, maintenance, and user support. Ensures compliant access and use of data product.	B, C, D A, D A, C, D B	"Data platform owner prioritises all the requirements coming from all areas and own the platform, and basically give the direction how the platform will develop further". (D) "He ensures business value of a data product over its lifetime". (A) "Accountable for a data application over its lifetime, which includes compliant implementation, maintenance of the data application, and support of users". (C) "Manages access to data lab, app, or user home and is accountable for any activity [...] on it over its lifetime". (B)

The *data owner* is first the creator but can also be user of data (sources) in his or her domain of responsibility. This implies the accountability for the definition, the quality and the lifecycle of data and can be associated with the paradigms of *creator as owner* and *consumer as owner*. The *data owner* is a pure business role in all four case companies, but with varying organisational assignment levels. While in company A, this role is assigned on a director level, in company C, it is assigned to a lower management function so as to ensure efficiency in handling data requests. In company D, this role is assigned to any organisational unit which is primary user of a dataset.

The *data owner* is accountable for making data fit its purpose, as outlined by seminal papers (Wang & Strong, 1996). However, data owners also play a key role in advancing the digital transformation by increasing the availability of quality data captured by digital technologies (Vial, 2019). Interestingly, we find that BDA also extends the responsibilities of *data owners* to also provide the input data for new data products. First, the *data owner* is expected to address the particular requirements of data products according to service-level agreements – as in company A and D. For instance, in company D such contractual agreements are handled through a dataset's context of use and a central organisation helps in moderating and defining them. Second, the *data owner* ensures compliant access and use of the data on the platform, i.e., manages data requests, approves usage, and provides access. For instance, the *data owner* in company B must continually revisit the no-join policies so as to ensure compliant use, also when the number of data available on the platform increases. This responsibility requires both additional effort and knowledge of potential implications when data are combined with data from other domains. In this regard, the *data owner* controls the decentralised access, which is one of the key data security issues to be solved in BDA environments (Grover et al., 2018), and may even be needed at an intra-organisational level (Günther et al., 2017).

Proposition 3: The data product owner ensures business value of a data product over its lifetime, including use case portfolio management, development, maintenance, and user support. Depending on the data product's complexity, this role may require technical expertise; thus, this may be a shared role between business and IT.

The *data product owner* is accountable for the data product. Notably, the companies differentiated between data products that are yet in their development (typically, a sandbox environment ("data lab") used by an analytics development team to explore and experiment with a dataset) and data products that are already developed and used downstream in productive

systems (e.g., a customer churn prediction model used by sales teams). In companies A and C, the *data product owner* is accountable for the data product over its lifetime, including development, maintenance, and user support. In company D, the *data product owner* manages data products for a portfolio of use cases of varying maturity. For use cases with low maturity (i.e., hypothesis that yet need to be validated), the *data product owner* collaborates with data and analytics experts to acquire all necessary data and turn these use cases into value generating products. Here, the paradigms *decoder as owner* (e.g., a data scientist who decodes a pattern in the data) or *compiler as owner* (e.g., data analysts who aggregate multiple data sources) are more suitable as the *data product owner* involved in the creation of the data product that is then consumed by a user. In company A, this role mainly ensures that the data product generates a business value over its lifetime. In company D, the *data product owner* is complemented with the role of a business owner who makes sure that data products are actually used. In this sense, the *data product owner* can also be associated with the *consumer as owner* paradigm.

Proposition 4: In BDA environments, the data platform owner role facilitates data supply (data owners) and data demand (data product owners). This activity ensures the availability of data on the platform for data exploration and experimentation, but also the operation of data products.

Companies manage BDA with data platforms, storing data from multiple sources and delivering data products for data exploration/experimentation and for direct use. This observation underpins the disruptive nature of BDA to amalgamate technologies to derive knowledge from big data into platforms (Abbasi et al., 2016). All enterprises have the role of a *data platform owner*, which serves as a mediator and facilitates data supply (*data owner*) and data demand (*data product owner*). While there are many *data owners* and *data product owners*, there is usually only one *data platform owner* assigned to an IT role in an enterprise. Thus, we can link this ownership type to the paradigms *compiler as owner* since this role brings data from various sources to the platform, and *packager as owner* since they reformat data for particular uses in data products. In company B, this role has the important (even strategic) function to “proactively” search for and bring valuable data (according to a business potential and need) to the platform. This role is also accountable for the development and operation of the platform – as is also the case in company C and D. This also includes controlling whether data products comply with data platform standards. In

sum, the *data platform owner* is responsible for the availability of data on the platform since she or he manages the data pipelines to bring data to the platform and to provide data to data products. Our findings thereby also support Wamba et al.’s (2015, p. 242) study that “emphasizes not only the support but also the active involvement of senior management for successful implementation of the shared platform to leverage ‘big data’ capabilities”.

5.2. Implications of data repurposing

With BDA, the analytical paradigm changes from using data in known ways towards finding innovative ways of using data in unknown ways (data repurposing). From the challenges that enterprises encounter when repurposing data, we derive further propositions related to the assignment of data ownership and changes in responsibilities.

Proposition 5: With data repurposing, data’s context of use deviates more often from its origin. Thus, new data owners may be assigned if the data creators are not able to cope with the additional data requirements.

The role of the data owner becomes an elementary role in the context of BDA. As data repurposing results in changes of a dataset’s context of use, it often results in new data requirements, e.g., a specific data attribute must be collected at a data source to be used in a data product. Thus, in order to manage these deviations, responsibilities are required at the source level for maintaining data requirements, while ensuring compliant access and use. The identification and assignment of data owners must follow a governed process to align data supply and demand effectively. In company D, the context of use comprises six dimensions which define a dataset’s functional bounds a data owner looks after: *Who*, *What*, *Why*, *Where*, *How*, and *When*. *Who* defines the qualifications and skills of dataset user, *What* defines the dataset itself and its sensitivity level, *Why* describes its purpose of use, *Where* the location of use and how data are flowing to and from that location, *How* governs the maintenance and use of data, and *When* specifies a dataset’s time of use and retention restrictions. When a dataset’s context of use changes, it is either extended or a new context of use is defined and assigned to a new data owner. The latter case will only happen when one or more dimensions need to be adapted in a way that goes beyond the original data owner’s area of expertise, for instance.

Propositions 6: With data repurposing, the number of dependencies between datasets and data products are increasing. The data platform owner assumes additional responsibilities for maintaining transparency

and contractual agreements between data owners and data product owners.

Data repurposing immediately results in an increasing number of dependencies between datasets and data products. On the one side, these dependencies need to be managed on the source level where data requirements are maintained. On the other side, these dependencies also need to be managed on the platform level where data products consume data. For instance, engineers at Google warn about data dependencies in machine learning applications that can lead to high technical debt (Sculley et al., 2015). Transparency on these data dependencies is needed to ensure traceability of data quality impacts, for instance. Hence, the *data platform owner* acts as an intermediary role with additional responsibilities regarding transparency and contractual agreements between data owners and data product owners. In line with the concept of information logistics, the *data platform owner* plays an important role in coordinating enterprise-wide information flows and managing the increasing number of data consumer-provider relationships.

6. Summary and outlook

6.1. Contribution

Our study contributes to the emerging field of research on data governance, which is considered a critical success factor for BDA (Grover et al., 2018) and for digital transformation in general (Vial, 2019). More specifically, we link data ownership to the underlying philosophical assumptions (Hart, 2002) and identify data ownership types that help assigning the decision rights for governing the content of IT artefacts according to Tiwana et al. (2013)'s IT Governance Cube. Our findings confirm that data ownership remains a key concept to clarify rights and responsibilities but should be revisited in the BDA context. While BDA environments come with specific challenges, due to the nature of advanced analytics products and the more frequent repurposing of data, some of the established principles for operational systems and data warehouses still hold true; most importantly, the clear distinction between the owner on the data supply side (*data owner*) and the owner on the data demand side (*data product owner*). Despite these similarities, BDA environments require also a change in responsibilities and additional role of the *data platform owner* to mediate data supply (*data owner*) and data demand (*data products*). We conclude that building BDA environments leads to even more complex data provider-consumer relationships and requires effective

coordination of enterprise-wide information flows. Our propositions and the suggested ownership types represent a first step towards studying *BDA governance to facilitate the value creation process*, which is a key theme of Grover et al.'s (2018) research framework.

6.2. Limitations

This study comes not without limitations. Since the four case companies represent large organisations, the findings may not be transferrable to smaller enterprises. Also, case studies only allow for analytical generalisation, and we suggest quantitative empirical studies to further validate our findings.

6.3. Implications for research

While prior research has mostly looked at either data or analytics governance, our findings illustrate how these two worlds are interconnected and inform future research on these topics. Eventually, the three types of data ownership may guide the definition of governance mechanisms for BDA and should be considered as the basis for more comprehensive data governance roles and frameworks. We show how data governance designs must be extended to include analytics-related accountabilities for data products and data platforms. Moreover, the three data ownership types are highly interdependent and will need to interact frequently. These interdependencies underline not only the importance of relational governance mechanisms, but also the collaboration between data and analytics teams with business and IT departments. Data and data product ownership are accountabilities ideally assigned to business stakeholders who understand best how to create business value. However, their domain expertise must be complemented with knowledge about data and analytics. This augmentation requires the collaboration with data and analytics experts. Platform ownership lies with the data and analytics teams, which onboard the data and deliver data products, and the IT teams, which operate and develop the infrastructure.

From the perspective of enterprise-wide information logistics, the assignment of data ownership can be interpreted as coordination mechanism in analytical information systems. By setting clear data ownership frameworks, organisations foster “*the planning, control and implementation of cross-unit data flows in order to realize enterprise-wide (or even inter-organizational) synergies*” (Winter, 2008, p. 47). In correspondence to Winter (2008), the *data owner* represents the unit in which data are generated, the *data product owner* the unit in which data are analytically processed, and the *data platform*

owner manages the platform infrastructure which is essential for information logistics success. We envision that organisations will be highly data-driven in the future. As data demands increase, the organisation inevitably evolves into a complex network of data producers and data consumers. The assignment of data ownership therefore plays a significant role to coordinate these raising data provider–consumer relationships and requires further research to understand the involved processes in greater depth.

6.4. Implications for practice

Practitioners may use our findings to define their approach to ownership as well as the related roles and responsibilities. Our findings can help them to increase consistency in role definitions and establish an understanding of data supply and demand in their data governance initiatives. For instance, the three data ownership types can be used to derive further roles, such as data engineers who typically work alongside data platform owners to implement data pipelines and data scientists who collaborate with data product owners to build advanced analytics models. Moreover, the ownership types and governance structures need to be complemented by new data quality management practices as data repurposing more frequently changes the data use contexts. Ideally, companies establish scalable and agile approaches for onboarding data in the right quality to create immediate business value through data exploration and experimentation.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Competence Center Corporate Data Quality (CC CDQ). The authors would like to thank all CC CDQ partner companies for their financial support, and their active contributions to this research.

References

Abbasi, A., Sarker, S., & Chiang, R. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, 17(2), 1–XXXII. <https://doi.org/10.17705/1jais.00423>

Alexander, D., & Lyytinen, K. (2017). Organizing successfully for big data to transform organizations. *AMCIS 2017 proceedings*. <http://aisel.aisnet.org/amcis2017/DataScience/Presentations/30>

Bajens, J., Helms, R. W., & Velstra, T. (2020, June 15). Towards a framework for data analytics governance mechanisms. *Proceedings of the 28th European*

Conference on Information Systems (ECIS), an online AIS conference.

Becker, L. C. (1980). The moral basis of property rights. *Nomos XXII: Property*, 22, 187–220

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The case research strategy in studies of information systems. *MIS Quarterly*, 11(3), 369–386. <https://doi.org/10.2307/248684>

Castelvecchi, D. (2016). Can we open the black box of AI? *Nature News*, 538(7623), 20. <https://doi.org/10.1038/538020a>

Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188. <https://doi.org/10.2307/41703503>

Chessell, M., Scheepers, F., Strelchuk, M., van der Starre, R., Dobrin, S., & Hernandez, D. (2018). *The journey continues: From data lake to data-driven organization*. Redbooks.

Comuzzi, M., & Patel, A. (2016). How organisations leverage Big Data: A maturity model. *Industrial Management & Data Systems*, 116(8), 1468–1492. <https://doi.org/10.1108/IMDS-12-2015-0495>

Confessore, N. (2018, April 4). Cambridge analytica and Facebook: The scandal and the fallout so far. *The New York Times*. <https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html>

Custers, B. (2013). Data dilemmas in the information society: Introduction and overview. In B. Custers, T. Calders, B. Schermer, & T. Zarsky (Eds.), *Discrimination and privacy in the information society: Data mining and profiling in large databases* (pp. 3–26). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30487-3_1

Davenport, T. H., Barth, P., & Bean, R. (2012). How ‘Big data’ is different. *MIT Sloan Management Review*, 54(1), 5. <https://sloanreview.mit.edu/article/how-big-data-is-different/>

Diakopoulos, N. (2016). Accountability in algorithmic decision making. *Communications of the ACM*, 59(2), 56–62. <https://doi.org/10.1145/2844110>

Dinter, B. (2013). Success factors for information logistics strategy—An empirical investigation. *Decision Support Systems*, 54(3), 1207–1218. <https://doi.org/10.1016/j.dss.2012.09.001>

Dosis, A., & Sand-Zantman, W. (2019). The ownership of data. *SSRN Electronic Journal*. 1–49. <https://doi.org/10.2139/ssrn.3420680>

Fama, E. F., & Jensen, M. C. (1983). Separation of ownership and control. *The Journal of Law and Economics*, 26(2), 301–325. <https://doi.org/10.1086/467037>

Farid, M., Roatis, A., Ilyas, I. F., Hoffmann, H.-F., & Chu, X. (2016). CLAMS: Bringing quality to data lakes. SIGMOD '16: Proceedings of the 2016 International Conference on Management of Data. 2089–2092. <https://doi.org/10.1145/2882903.2899391>

Grover, V., Chiang, R. H. L., Liang, T.-P., & Zhang, D. (2018). Creating strategic business value from big data analytics: A research framework. *Journal of Management Information Systems*, 35(2), 388–423. <https://doi.org/10.1080/07421222.2018.1451951>

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, 26(3), 191–209. <https://doi.org/10.1016/j.jsis.2017.07.003>

- Hart, D. (2002). Ownership as an issue in data and information sharing: A philosophically based review. *Australasian Journal of Information Systems*, 10(1). <https://doi.org/10.3127/ajis.v10i1.440>
- Hummel, P., Braun, M., & Dabrock, P. (2020). Own data? Ethical reflections on data ownership. *Philosophy & Technology*, 1–28. <https://doi.org/10.1007/s13347-020-00404-9>
- Janeček, V. (2018). Ownership of personal data in the Internet of Things. *Computer Law & Security Review*, 34(5), 1039–1052. <https://doi.org/10.1016/j.clsr.2018.04.007>
- Jones, C., & Tonetti, C. (2019). *Nonrivalry and the economics of data* (No. w26260). National Bureau of Economic Research. <https://doi.org/10.3386/w26260>
- Kostkova, P., Brewer, H., De Lusignan, S., Fottrell, E., Goldacre, B., Hart, G., Koczan, P., Knight, P., Marsolier, C., McKendry, R. A., & Ross, E. (2016). Who Owns the Data? Open Data for Healthcare. *Frontiers in public health*, 4, 1–7. <https://doi.org/10.3389/fpubh.2016.00007>
- Labadie, C., & Legner, C. (2019). Understanding data protection regulations from a data management perspective: A capability-based approach to EU-GDPR. Proceedings of the 14th international conference on Wirtschaftsinformatik (WI) (pp. 1292–1306). <https://aisel.aisnet.org/wi2019/track11/papers/3>
- Link, G., Lombard, K., Germonprez, M., Conboy, K., & Feller, J. (2017). Contemporary issues of open data in information systems research: Considerations and recommendations. *Communications of the Association for Information Systems*, 41(1), 587–610. <https://doi.org/10.17705/1CAIS.04125>
- Loshin, D. (2001). *Enterprise knowledge management: The data quality approach*. Morgan Kaufmann.
- Madera, C., & Laurent, A. (2016). The next information architecture evolution: The data lake wave. *Proceedings of the 8th international conference on management of digital ecosystems* (pp. 174–180). <http://doi.acm.org/10.1145/3012071.3012077>
- Maxwell, B. (1989). Beyond “Data validity”: Improving the quality of HRIS data. *Personnel*, 66(4), 48–58
- McKenna, E., Richardson, I., & Thomson, M. (2012). Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy*, 41, 807–814. <https://doi.org/10.1016/j.enpol.2011.11.049>
- Naderifar, M., Goli, H., & Ghaljaie, F. (2017). Snowball sampling: A purposeful method of sampling in qualitative research. *Strides in Development of Medical Education*, 14(3), 1–4. <https://doi.org/10.5812/sdme.67670>
- Open AI. (2020). *Open AI*. Open AI. <https://openai.com/>
- OpenStreetMap. (2019). *OpenStreetMap*. OpenStreetMap. <https://www.openstreetmap.org/copyright>
- Otto, B. (2011). Data governance. *Business & Information Systems Engineering*, 3(4), 241–244. <https://doi.org/10.1007/s12599-011-0162-8>
- Schulze, W. S., & Zellweger, T. M. (2020). Property rights, owner-management, and value creation. *Academy of Management Review*. <https://doi.org/10.5465/amr.2018.0377>
- Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J. F., Dennison, D. (2015). Hidden technical debt in machine learning systems. NIPS’15: Proceedings of the 28th International Conference on Neural Information Processing Systems, 2, 2503–2511. <https://doi.org/10.5555/2969442.2969519>
- Shleifer, A. (1998). State versus private ownership. *Journal of Economic Perspectives*, 12(4), 133–150. <https://doi.org/10.1257/jep.12.4.133>
- Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of big data challenges and analytical methods. *Journal of Business Research*, 70, 263–286. <https://doi.org/10.1016/j.jbusres.2016.08.001>
- Spirig, J. (1987). Compensation: The up-front issues of payroll and HRIS interface. *Personnel*, 66(100), 124–129.
- Tallon, P. P., Ramirez, R. V., & Short, J. E. (2013). The information artifact in IT governance: Toward a theory of information governance. *Journal of Management Information Systems*, 30(3), 141–178. <https://doi.org/10.2753/MIS0742-1222300306>
- The Economist. (2017). *Data is giving rise to a new economy*. The Economist Group Limited. <https://www.economist.com/briefing/2017/05/06/data-is-giving-rise-to-a-new-economy>
- Tikkinen-Piri, C., Rohunen, A., & Markkula, J. (2018). EU general data protection regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review*, 34(1), 134–153. <https://doi.org/10.1016/j.clsr.2017.05.015>
- Tiwana, A., Konsynski, B., & Venkatraman, N. (2013). 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>
- Van Alstyne, M., Brynjolfsson, E., & Madnick, S. (1995). Why not one big database? Principles for data ownership. *Decision Support Systems*, 15(4), 267–284. [https://doi.org/10.1016/0167-9236\(94\)00042-4](https://doi.org/10.1016/0167-9236(94)00042-4)
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *The Journal of Strategic Information Systems*, 28(2), 118–144. <https://doi.org/10.1016/j.jsis.2019.01.003>
- Wamba, S. F., Akter, S., Edwards, A., Chopin, G., & Gnanzou, D. (2015). How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study. *International Journal of Production Economics*, 165, 234–246. <https://doi.org/10.1016/j.ijpe.2014.12.031>
- Wang, R., & Strong, D. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33. <https://doi.org/10.1080/07421222.1996.11518099>
- Wang, R. Y., Storey, V. C., & Firth, C. P. (1995). A framework for analysis of data quality research. *IEEE Transactions on Knowledge and Data Engineering*, 7(4), 623–640. <https://doi.org/10.1109/69.404034>
- Watson, H. (2009). Business intelligence: Past, present and future. *AMCIS 2009 Proceedings*, 153. <https://aisel.aisnet.org/amcis2009/153/>
- Watson, H. J. (2017). Preparing for the cognitive generation of decision support. *MIS Quarterly*, 16(3), 153–169. <https://aisel.aisnet.org/misqe/vol16/iss3/3>
- 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(3), 435–450. <https://doi.org/10.1016/j.dss.2003.06.001>
- Watson, H. J., & Wixom, B. H. (2007). The current state of business intelligence. *Computer*, 40(9), 96–99. <https://doi.org/10.1109/MC.2007.331>
- 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/10.1145/1515693.1515696>

- Winkler, T. J., & Wessel, M. (2018). A primer on decision rights in information systems: Review and recommendations. Proceedings of the 39th International Conference on Information Systems (ICIS). <https://aisel.aisnet.org/icis2018/general/Presentations/5/>
- Winter, R. (2008). Enterprise-wide information logistics: Conceptual foundations, technology enablers, and management challenges. *Proceedings of ITI 2008* (pp. 41–50). <https://doi.org/10.1109/ITI.2008.4588382>
- Winter, R., & Meyer, M. (2001). Organization of data warehousing in large service companies—A matrix approach based on data ownership and competence centers. *Journal of Data Warehousing*, 6(4), 23–29. <https://www.alexandria.unisg.ch/publications/66638>
- Wixom, B., & Ross, J. (2017). How to monetize your data. *MIT Sloan Management Review*, 58(3), 10–13. <https://sloanreview.mit.edu/article/how-to-monetize-your-data/>
- Yin, R. (2003). *Case study research: Design and methods, third edition, applied social research methods series* (Vol. 5). Sage Publications, Inc.