Year: 2024

SCALING DATA PRACTICES IN MULTINATIONAL FIRMS: ESSAYS ON DATA GOVERNANCE AND DATA DEMOCRATIZATION

LEFEBVRE Hippolyte Antoine Marie

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DES SYSTÈMES D'INFORMATION

SCALING DATA PRACTICES IN MULTINATIONAL FIRMS: ESSAYS ON DATA GOVERNANCE AND DATA DEMOCRATIZATION

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de

Doctorat en systèmes d'information

par

Hippolyte Antoine Marie LEFEBVRE

Directrice de thèse
Prof. Christine Legner

Jury
Prof. Valérie Chavez-Demoulin, présidente
Prof. Stéphanie Missonier, experte interne
Prof. Yash Raj Shrestha, expert interne
Prof. Elena Parmiggiani, experte externe
Prof. Gregory Vial, expert externe

LAUSANNE
2024

UNIL | Université de Lausanne

FACULTÉ DES HAUTES ÉTUDES COMMERCIALES

DÉPARTEMENT DES SYSTÈMES D'INFORMATION

SCALING DATA PRACTICES IN MULTINATIONAL FIRMS: ESSAYS ON DATA GOVERNANCE AND DATA DEMOCRATIZATION

THÈSE DE DOCTORAT

présentée à la

Faculté des Hautes Études Commerciales
de l'Université de Lausanne

pour l'obtention du grade de

Doctorat en systèmes d'information

par

Hippolyte Antoine Marie LEFEBVRE

Directrice de thèse
Prof. Christine Legner

Jury
Prof. Valérie Chavez-Demoulin, présidente
Prof. Stéphanie Missonier, experte interne
Prof. Yash Raj Shrestha, expert interne
Prof. Elena Parmiggiani, experte externe
Prof. Gregory Vial, expert externe

LAUSANNE
2024

UNIL | Université de Lausanne
Faculté des hautes études commerciales
Internef - 261
CH-1015 Lausanne

IMPRIMATUR

La Faculté des hautes études commerciales de l'Université de Lausanne autorise l'impression de la thèse de doctorat rédigée par

Hippolyte Antoine Marie LEFEBVRE

intitulée

Scaling Data Practices in Multinational Firms: Essays on Data Governance and Data Democratization

sans se prononcer sur les opinions exprimées dans cette thèse.

Lausanne, le 23.08.2024

Professeure Marianne Schmid Mast, Doyenne

Members of the thesis committee

Prof. Christine LEGNER

Professor, Faculty of Business and Economics (HEC),

University of Lausanne, Switzerland.

Thesis supervisor

Prof. Stéphanie MISSONIER

Professor, Faculty of Business and Economics (HEC),

University of Lausanne, Switzerland.

Internal member of the thesis committee

Prof. Yash Raj SHRESTHA

Assistant Professor, Faculty of Business and Economics (HEC),

University of Lausanne, Switzerland.

Internal member of the thesis committee

Prof. Elena PARMIGGIANI

Associate Professor, Department of Computer Science,

Norwegian University of Science and Technology (NTNU), Norway.

External member of the thesis committee

Prof. Gregory VIAL

Associate Professor, Department of Information Technologies,

HEC Montreal, Canada.

External member of the thesis committee

Prof. Valérie CHAVEZ-DEMOULIN

Professor, Vice-Dean for Research and Academic Succession, Faculty of Business and Economics (HEC),

University of Lausanne, Switzerland.

President of the thesis committee

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Hippolyte Antoine Marie LEFEBVRE

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________ Date: June 19, 2024

Prof. Christine LEGNER
Thesis supervisor

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Hippolyte Antoine Marie LEFEBVRE

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ___________________________ Date: _____24.06.2024______

Prof. Stéphanie MISSONIER
Internal expert

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Hippolyte Antoine Marie LEFEBVRE

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________ Date: 20.06.2024

Prof. Yash Raj SHRESTHA
Internal expert

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Hippolyte Antoine Marie LEFEBVRE

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________ Date: 21/06/2024

Prof. Elena PARMIGGIANI
External expert

University of Lausanne
Faculty of Business and Economics

PhD in Information Systems

I hereby certify that I have examined the doctoral thesis of

Hippolyte Antoine Marie LEFEBVRE

and have found it to meet the requirements for a doctoral thesis.
All revisions that I or committee members
made during the doctoral colloquium
have been addressed to my entire satisfaction.

Signature: ______________________________ Date: June 19, 2024

Prof. Gregory VIAL
External expert

Acknowledgements

A common cliché about pursuing a PhD is the loneliness it entails. Looking back at the past four years, I never felt alone. Some amazing individuals contributed along the way. I dedicate this work to my incredibly supportive team members—Valérianne, Pavel, Redwan, Liz, Tobias, Konrad—and to my bright academic friends—Alisa, Ramin, Maylis, Khalid, Amir, Anna and all the others with whom I exchanged (a lot of) laughter and (a few) ideas.

Another prevalent cliché about pursuing a PhD is that your success depends largely on your supervisor. While I can't speak for everyone, in my case, it turned out to be true. To my supervisor Christine Legner, I want to extend my deepest gratitude for your guidance. While it sometimes felt like solving a Rubik's cube blindfolded, this process greatly enriched my research undoubtedly shaped my resilience, preparing me for future endeavors. Thank you for your relentless challenges and meticulous scrutiny, ensuring every citation was checked, every argument was problematized. Your constant push forward has been invaluable. I hope that, in turn, I have also succeeded in pushing you forward.

People also emphasize how crucial support from your close ones is. A special shout-out to Bégonia and my family, especially my mom, who encouraged me to pursue a PhD since I was 10, yet was in disbelief when I finally announced that I would do it.

Abstract

In recent decades, the perception of data within corporations has evolved significantly, elevating from a mere operational resource to a strategic asset pivotal for value creation. However, despite some progress, industry reports show that many firms still struggle with scaling the necessary data practices for using data effectively, which hampers their ability to achieve data-driven innovation. This challenge often arises because data practices are not sufficiently developed beyond data experts, limiting the ability of a broader range of employees to competently engage with data in their daily work. Research has yet to explain how companies develop data practices among a widening audience — a capability denoted as democratization — especially how employees can effectively integrate data with domain expertise through context-specific data practices. Simultaneously, the role of data governance, traditionally viewed as a control function overemphasizing data protection, needs to evolve into a coordinating role that facilitates data practices to realize data-driven innovation. Therefore, this thesis elucidates how data democratization and data governance co-evolve toward strategic value creation from data, through two interrelated streams of research. Through three essays, the first research stream grounds data democratization in Information Systems (IS) research, identifying it as a capability rooted in practice and highlighting its socio-technical nature. We emphasize the critical necessity of integrating both generic and situated data practices to achieve true data democratization. We illustrate how these data practices are cultivated through situated learning and practice exchange. Through two essays, the second research stream explains how to govern data effectively to achieve both control and innovation. We introduce systems thinking to position data governance at the intersection of data strategy and data operations within a Viable System Model. We describe the reconfiguration of data governance into archetypes that reflect the evolving strategic role of data. Altogether, our findings significantly advance data management research by providing a clearer understanding of how to scale data practices through the interplay between data democratization and data governance, highlighting their synergistic efforts in driving value creation from data.

Contents of the thesis

Introductory Paper on

SCALING DATA PRACTICES IN MULTINATIONAL FIRMS:

ESSAYS ON DATA GOVERNANCE AND DATA DEMOCRATIZATION

Hippolyte Lefebvre

Department of Information Systems,

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

Table of contents

List of figures

List of tables

1 Introduction

In recent decades, the perception of data in corporations has evolved significantly, elevating it from a mere operational resource to a strategic asset pivotal in value creation (Grover et al., 2018; Günther et al., 2022). Accelerated by Big Data's steep rise, by advanced analytics, and more recently by GenAI, this shift has prompted organizations to scale the practices through which employees engage and re-engage with data (Alavi et al., 2024; T. H. Davenport et al., 2024). In the context of information systems (IS), these practices refer to "*what people actually do with the technological artifact*" (Orlikowski, 2000, p. 408). Therefore, data practices refer to the activities and methods through which employees engage with data in their daily work (Parmiggiani et al., 2023). Data practices typically encompass data conceptualization, data collection, data curation, data consumption, and data control (Chua et al., 2022). By progressively scaling data practices to more teams and business units beyond just data offices composed of experts, thereby reaching those employees "*without 'data' in their title*" across the entire organization (Redman, 2022, p. 1), companies seek to combine business acumen and data expertise to obtain a better data impact (Rutschi et al., 2023; Someh et al., 2023).

However, this effort is frequently challenged due to restricted data access and a scarcity of data competencies among users other than data specialists (Zeng & Glaister, 2018; Awasthi & George, 2020). To address such challenges firms must develop a new capability denoted as data democratization that aims to empower their employees toward working with data (Zeng & Glaister, 2018; Awasthi & George, 2020; Hyun et al., 2020). Despite its clear socio-technical nature, data democratization has been incorporated in a range of fields (e.g., medicine, urban planning); however, it remains scarcely integrated in IS research. Nascent IS literature that conceptualizes data democratization as a capability mainly focuses on isolated aspects (e.g., training programs, platforms, access rights). While useful, these insights which mostly focus on extending data access to non-specialists, lack a broader perspective on how data practices are developed and integrated into the work routines of a progressively growing group of employees.

A prerequisite for scaling data practices is data governance (Grover et al., 2018). Data governance represents "*the leading function of data management as it specifies which decisions need to be made in data management and who makes these decisions*" (Otto, 2011b, p. 1). Therefore, the literature has characterized data governance as a control function (Chua et al., 2022), and examined it through three prominent streams. In the first stream, data governance is considered as exercising authority in managing data quality, while also seeking to make (master) data fit-for-use (Otto, 2011b; Weber et al., 2009a). The second stream, which is analogous to IT asset management and integrates other decision domains than just data quality (e.g., data security), considers data governance as "*cross-functional framework for managing data as a strategic enterprise asset*" (Abraham et al., 2019, p. 425). Mirroring IT governance frameworks, data governance involves structural mechanisms (e.g., roles and responsibilities, decision-making locus), procedural mechanisms (e.g., processes, monitoring), and relational mechanisms (e.g., communication, training) (Abraham et al., 2019; Tallon et al., 2013). The third stream investigates data governance as a matter of work practices. It studies how employees implement data governance frameworks when they use data in their respective working contexts (Benfeldt et al., 2020; Parmiggiani et al., 2023). However, insights from practice (Vial, 2023) show that data governance still fails to fulfill its goal of "*maximizing the value of data assets in enterprises*" (Otto, 2011b, p. 1). To address this difficulty, research must extend data governance's role beyond one of control since this overemphasizes data protection (Vial, 2023). Rather, due to the criticality of data for business innovation, data governance should have a coordinating function, not only maintaining oversight over data practices but also supervising an increasing data scope and value creation scenarios, which involves a growing number of employees (Mikalef et al., 2020; Vial, 2023).

This thesis uses a practice lens to capture the broader context in which data practices are scaled. Through two interrelated research streams, it explores the mutual development of data practices and their integration into a larger framework for creating value from data. The first research stream explains data democratization by considering how others than only the data experts develop data practices. The second research stream explains data governance's dual mandate of balancing data control and data-driven innovation.

This thesis contributes to the emerging data democratization research and revisits the notion of data governance. On the one hand, we show that to democratize their data firms must develop both generic practices and situated data practices, thereby respectively establishing a foundation for all employees and enabling contextual and pragmatic data utilization that meets diverse workplace expectations. On the other hand, we explain how data governance achieves its dual mandate by dynamically unfolding into a viable system at the interplay between data strategy and data operations. Further, our findings show that data governance practices form configurations that can be mapped onto different strategic contexts for data.

The remainder of the paper unfolds as follows; Section 2 establishes the theoretical background of the thesis — examining data democratization and data governance in the literature — and pinpoints the limitations of the existing body of knowledge. Section 3 articulates the overarching research objective of the thesis and its structure, followed by a description of the research setting. Sections 4 and 5 represent each research stream, outlining their individual motivations, research objectives, methodologies, main outcomes, and discussions. Finally, Section 6 explicates what the thesis contributes both theoretically and practically by summarizing the findings, critically analyzing their broader implications and limitations, and suggesting future research directions at the intersection of data governance and data practices.

2 Background

2.1 Data democratization to address the changing role of data

The role data fulfills in enterprises has changed considerably over the past decade (Legner et al., 2020). We observe the change mainly in three chronologically identified phases, as given in Table 1.

Phase	Phase 1: (since the 1980s)	Phase 2: (since the 1990s)	Phase 3: (since the 2010s)
Data's roles	Data as a prerequisite for application development and as an enabler of automation in business functions	Data as an enabler of enterprise-wide business processes and decision-making	Data as an enabler of a firm's business models and value propositions
Data scope	Structured data	Mainly internal, structured data	Large volumes of internal and external data (*big data*), comprising structured and non-structured data sources
Data-related concerns	Data model quality, data availability, data re-use (Gillenson, 1985)	Enterprise-wide data integration, data quality (Goodhue et al., 1992; Grover & Teng, 1991; Ravindra, 1986)	Business value and impacts, data compliance, data privacy, data security (Akter, Wamba, Gunasekaran, Dubey, & Childe, 2016; Constantiou & Kallinikos, 2015; Xie, Wu, Xiao, & Hu, 2016)
Responsibilities for data	Database administrators (Goldstein & McCririck, 1981; Weldon, 1981)	Business process owners, later master data management (MDM) and business intelligence (BI) teams	Chief data officer, data scientists, data analysts

Table 1. Evolution of data roles in enterprises (adapted from Legner at al. (2020))

In the first phase, spanning the 1980s, data was considered an essential component of IT artifacts and served as a foundational element in developing applications and a lever for automating business operations (Grover & Teng, 1991). This phase was characterized by the utilization of structured data within siloed business functions. Collecting data was largely a reactive process driven by specific demands and requirements (Ballou et al., 1998; Goodhue et al., 1988). Data consumption, which primarily occurred in the internal systems, was largely transactional in nature. The output in each case was designed to support operational decision-making and day-to-day management, rather than to provide strategic insight or to drive innovation. The primary concerns centered on the quality of data models, the availability of data, and the facilitation of data reusability. The responsibility for managing data rested predominantly with database administrators, whose roles were defined by a focus on the integrity and maintenance of database systems (Goldstein & McCririck, 1981; Weldon, 1981).

During the second phase of the evolution of data, which began in the 1990s, the role of data within organizations expanded significantly. Data started to enable enterprise-wide business processes and strategic decision-making, thus departing from the system-centric perspective on data (McKeen & Smith, 2007). The scope of data during this era shifted toward mainly internal, structured data, but with a greater emphasis on enterprise-wide integration. Key concerns included ensuring data quality, especially master data quality, and integrating data seamlessly across various business units (Goodhue et al., 1992; Grover & Teng, 1991; Ravindra, 1986). Data responsibilities were taken from the hands of individual database administrators and elevated to be managed by business process owners, master data management (MDM) specialists, and business intelligence (BI) teams. This marked a significant transition toward a more holistic approach to data management within the corporate milieu, recognizing data as a valuable asset in assuring operational efficiency and gaining a competitive edge.

In the third phase, which began in the 2010s, data has become as an enabler of a firm's business models and value propositions. Organizations now harvest and harness substantial volumes of data from a vast array of sources, from IoT devices and online interactions — collectively known as big data — to drive decision-making and craft data-centric business strategies using sophisticated tools in analytics and artificial intelligence (Chen et al., 2021; Wamba, 2022). In this context, collected data is considered a valuable and rare strategic resource that is hard to imitate and organizationally embedded (VRIO), which can help companies gain a long-term competitive advantage (Gupta & George, 2016; Mikalef et al., 2018). The term data monetization (Mehta et al., 2021; Wixom & Ross, 2017) has been coined to describe the different ways in which firms create direct or indirect value from data, such as informational value (e.g., decision-making support), transactional value (e.g., cost efficiency), transformational value (e.g., business models), or strategic value (e.g., market positioning) (Elia et al., 2020).

Despite modest advancements, industry reports over the past five years consistently indicate continuing challenges in deriving value from their data (Bean, 2023). A prominent reason for these difficulties is that only a few specialists are granted data responsibilities (see Table 1). To create value at scale, firms should scale data practices—i.e., the activities and methods through which employees engage with data—beyond just a single unit or function typically composed of data experts only. This expansion should include 'regular' employees who have collaborative roles in their day-to-day work, thereby spearheading innovation from data (Alaimo & Kallinikos, 2022; Benfeldt et al., 2020).

IS research has conceptualized "*the act of opening organizational data to as many employees as possible, given reasonable limitations on legal confidentiality and security*" as a capability known as data democratization (Awasthi & George, 2020, p. 1). This conceptualization has its roots in the idea of democratization, a well-established area of inquiry in social and political sciences that examines transition from authoritarian regimes to representative governance. In such systems, power is distributed widely rather than concentrated, thus promoting individual rights and encouraging all segments of society to participate by reducing or eliminating structural barriers to power (Grugel, 2004). Applied to data, this lens reflects an ideological shift from data elitism, which concentrates all data privileges with data professionals, to data participation, in which all employees, supported by the data elite, are empowered to work with data (see Table 2). This ideal implies that eventually all employees should be granted data rights and responsibilities on data practices (Lycett, 2013; Redman, 2022; Zhu et al., 2019). Data democratization thus implies that all employees have equal rights to participate in data decision-making, whether within their specific working areas or in collaborative projects (Labadie, Eurich, et al., 2020). This inclusive approach is expected to enhance value creation through, for instance, improved quality of decisions, business process efficiency, customer experience, innovation and agility (Grover et al. 2018; Hyun et al., 2020).

Views on democracy (Jansen, 2009)		**Application to enterprise data practices**
Elitism, argued by Lippman	Experts and professionals rule due to possessing the necessary knowledge and expertise to make decisions in the public interest.	Data elitist organizations maintain privileges over data by keeping control in the hands of a minority of data professionals who possess all the knowledge required to work with data.
Participation, argued by Dewey	Citizens actively engage in decision-making processes and are part of a participatory democracy.	Data democratic organizations seek to empower as many employees as possible to work with data, while acknowledging data professional's guidance and expertise.

Table 2. Views on democracies applied to enterprise data practices

Research has mainly focused on isolated aspects of data democratization, such as training programs and data platforms (Awasthi & George, 2020; Labadie, Legner, et al., 2020a). However, while useful, these insights are mostly technical and detached from practice. Therefore, the existing studies are not sufficient to fully comprehend the capability through the expansion of data practices across the organization. To achieve value creation from data at scale, research would need to explain how data democratization materializes in the development and integration of data practices in the work of an ever-growing group of employees (Benfeldt et al., 2020; Parmiggiani et al., 2023).

2.2 Rethinking data governance for data-driven innovation

Data governance began to take shape in the late 1990s, primarily driven by the necessity for improved quality of data assets, as discussed in Section 2.1. Since then, various studies have examined data governance from three prominent streams: 1) Data governance as a matter of data quality management; 2) Data governance as a matter of structural, procedural, and relational mechanisms; and 3) Data governance as a matter of work practices (see Table 3).

Research streams	Key studies	Definition of data governance	Contribution
Data governance as a matter of data quality management	(Weber et al., 2009a)	"the framework for decision rights and accountabilities to encourage desirable behavior in the use of data."	Framework with data quality roles, decision areas, and responsibilities.
	(Otto, 2011b)	"Data Governance refers to the allocation of decision-making rights and related duties in the management of data in enterprise."	Framework with decision areas, roles, and authority.
	(Otto, 2011c)	"A companywide framework for assigning decision-related rights and duties in order to be able to adequately handle data as a company asset."	Framework for DG organization design: organizational goals, organizational form, and organizational transformation.
Data governance as a matter of structural, procedural, and relational mechanisms	(Khatri & Brown, 2010)	"Data governance refers to who holds the decision rights and is held accountable for an organization's decision-making about its data assets."	Framework of decision domains for data governance; locus of accountability through structural mechanisms.
	(Tallon et al., 2013)	"A collection of capabilities or practices for the creation, capture, valuation, storage, usage, control, access, archival, and deletion of information over its life cycle."	Framework of structural, procedural, and relational governance practices, shaped by antecedents and with consequences.
	(Alhassan et al., 2016)	Appropriated from (Otto, 2011c)	A framework of actions (Define/Implement/Monitor), eight governance areas, and five decision domains.
	(Abraham et al., 2019)	"A cross-functional framework for managing data as a strategic enterprise asset."	Conceptual framework for data governance, shaped by antecedents and with consequences.
	(Vial, 2023)	"A system of decision rights and accountabilities for information-related processes." (taken over from Data Governance Institute (2021))	Four research themes for data governance: Embracing data governance without compromising digital innovation; Enacting data governance through repertoires of mechanisms; From data governance to governing data; From systems, through data, to services.
Data governance as a matter of work practices	(Benfeldt et al., 2020)	"Data governance refers to the organization and implementation of rules and responsibilities, which enforce decision making and accountabilities regarding an organization's data assets."	Six enablers for implementing data governance in practice: Perceiving value, Enabling collaboration, Fostering capabilities, Data overview, Local practices, Political ambiance.
	(Parmiggiani & Grisot, 2020)	"A concept that describes an organization's capability to ensure data accessibility, consistency, and usability throughout their lifecycle."	Data governance (instantiated through data curation practices) as a matter of work practice instead of just a matter of asset management.

Table 3. Overview of data governance literature

The first stream associates data governance with data quality management, referring to how data was created, curated and used in standardized business processes and processed in operational systems (e.g., ERP) and analytical applications (e.g., BI). In such circumstances, data governance emerged within the IT governance scope, primarily revolving around the specific contexts of Enterprise Resource Planning (ERP) and data warehouses (Rifaie et al., 2009; Watson et al., 2004). This shift marked a departure from the conventional emphasis on traditional relational and transactional databases, and focused on enhancing master data (Khatri and Brown, 2010; Otto, 2011a, 2011b; Weber et al., 2009). Subsumed under organizational resources (Chua et al., 2022), data governance considers data to be a tangible organizational resource whose quality must be ensured to enable business processes. This perspective places particular emphasis on master data, recognizing its pivotal role for operational excellence. Drawing on Total Data Quality Management methodology which argues that "*an analogy exists between quality issues in product manufacturing and those in information manufacturing*" (Wang, 1998, p. 1), data governance ensures that data is "*fit for use*" (Wang & Strong, 1996, p. 6). This means that the data is not only accurate, but also relevant and timely, making it a reliable foundation for critical decision-making processes. Data governance, from this view point, is a means toward implementing clear accountabilities, standards, and policies designed to effectively ensure the quality of data assets (Otto, 2011c, 2011b; Weber et al., 2009a).

The second stream has expanded the management perspective on data governance's role, describing it through structural, procedural, and relational mechanisms. Based on principles of corporate and IT governance (Kohli & Grover, 2008; Tiwana et al., 2013), these mechanisms position data governance as a framework (see Figure 1) primarily designed to exercise control over the data lifecycle, ensuring data protection and mitigating risks of misuse (Tallon et al., 2013). Structural mechanisms "*focus specifically on the design of formal organizational elements to establish the decision rights and accountability of actors*" (Vial, 2023, p. 4). Examples of these mechanisms include boards and committees that can enforce decision making in structures on a continuum between centralized (e.g., within a group of corporate executives) and decentralized (e.g., in functional governance hubs) (Khatri & Brown, 2010). Representing the earliest data governance research findings, structural mechanisms are regarded as encompassing the most formalized data governance practices (in the form of, e.g., roles and responsibilities, decision-making locus) (Vial, 2023). Procedural mechanisms "*emphasize the operational means that are put in place to ensure compliance with governance principles*" (Vial, 2023, p. 4). They include policies, standards, and procedures, as well as contractual agreements and access management (Abraham et al., 2019). Relational mechanisms, often overlooked in research due to the excessive focus on data protection, "*provide a less formalized means of ensuring that data governance principles are understood and enforced by actors*" (Vial, 2023, p. 4). They include practices that facilitate collaboration and coordination between stakeholders, such as communication and training (Abraham et al., 2019, p. 430).

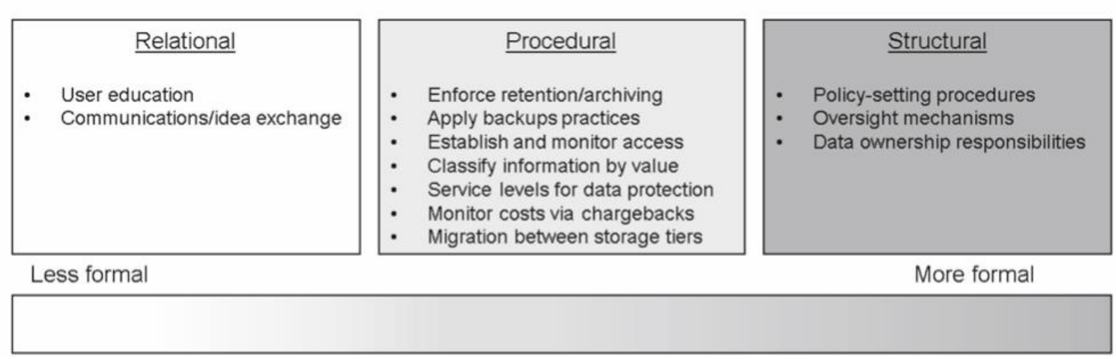


Figure 1. Examples of data governance mechanisms (Vial, 2023)

In a third, and more recent stream, studies suggest "*a shift from data governance as a matter of asset management to data governance as a matter of work practice*" (Parmiggiani & Grisot, 2020, p. 3). Researchers in this stream argue that data governance cannot simply focus on *what* data governance practices should be implemented (through the structural, procedural and relational mechanisms) while ignoring *how* to implement these practices (e.g., curating data following certain governance standards) (Benfeldt et al., 2020; Vial, 2023). As Alhassan (2016) highlighted, data roles and responsibilities should specify how data governance practices are defined, but also how they are implemented and monitored by various employees in different business units. For instance, while data standards may specify objectives for data quality, the tangible data curation is done in executing data operations that are part of the day-to-day job description (e.g., recording a new customer's address). The subsequent job performance then is subject to rigorous monitoring. Therefore, data governance can no longer "*overlook the day-to-day work of users engaged in data governance practices (i.e., working with data, interpreting outcomes, and making decisions)*" (Parmiggiani & Grisot, 2020, p. 2). Although this approach recognizes the potential contributions of those who work with data, it often underestimates the systemic barriers that prevent data governance policies to reach those employees. Consequently, even well-intentioned policies might fail to resonate with or address the daily realities and challenges faced by data practitioners, so that policy can remain misaligned with practical needs (Benfeldt et al., 2020).

To support the scaling of data practices, we need to rethink data governance. Recent insights from practice show that describing data governance only as a control function jeopardizes data driven innovation. Such a description overemphasizes data protection with restrictive policies that prevent non experts from working with data (Vial, 2023). Instead, the non-linear relationship between data governance and firm performance necessitates continuous monitoring and adjustment to strike a balance between under-governing data to stimulate data-driven innovation and over-governing data to ensure data protection (Tallon et al., 2013). Therefore, the mandate of data governance must evolve from its traditional controlling capacity over the data lifecycle to fulfil a more coordinating function which orchestrates the development of compliant and situated data practices (Alaimo & Kallinikos, 2022; Benfeldt et al., 2020). Despite some preliminary insights, there is a critical gap in our understanding of how these governance practices are practically integrated and coordinated across the organization, for instance across functions and regions. Further, although current research acknowledges that data governance enhances firm performance, it often overlooks that it is a "*dynamic element that is implemented and should evolve in conjunction with strategy and operations*" (Vial, 2023, p. 9). Concretely, as data practices are scaled across the organization, data governance must continuously adapt to an evolving data scope, and generally support growing strategic needs and an ever-evolving regulatory landscape (Abraham et al., 2019). Therefore, the fast-paced integration of data practices into organizational routines is incompatible with the rather static view offered by the conceptualization of data governance as structural, procedural, and relational mechanisms. Instead, the dynamic nature of data governance ensures that these mechanisms evolve in response to various inputs. This adaptive approach to data governance is crucial for organizations to remain agile, address emerging challenges, and leverage data as a strategic asset effectively. It fosters a culture of continuous improvement and learning, enabling organizations to innovate while maintaining compliance and managing risks.

3 Thesis overview

3.1 Research objectives and streams

Scaling data practices is essential for unlocking the full value of data. Research indicates that data democratization and data governance must be examined together to effectively address the scaling of data practices. Therefore, this thesis explores the intricate relationship between data democratization and data governance in scaling data practices through two distinct but connected research streams (see Figure 2). Research Stream I investigates data democratization through the development of data practices to include more than only the data experts. Research Stream II seeks to understand how data governance balances control and coordination of data practices. Together, these two streams elucidate the essential interplay between data governance and data democratization, enabling the scaling of data practices that drive value creation.

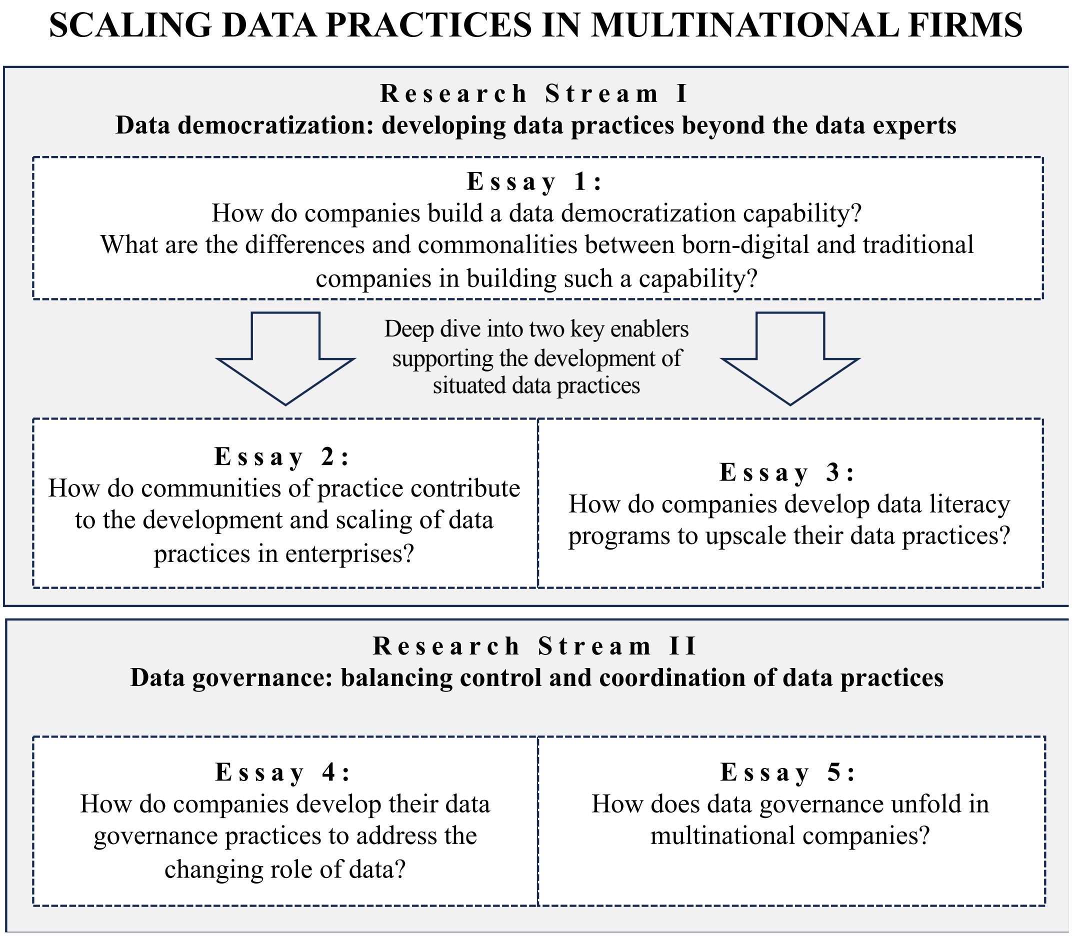


Figure 2. Overview of research streams and essays in the thesis

Research Stream I is titled *Data democratization: developing data practices beyond the data experts*. In three essays, it explains how data practices are developed and integrated into the work of a progressively growing group of employees. Essay 1 explores data democratization as an overlooked topic in IS research and refines its conceptualization as a capability. It suggests five enablers of data democratization initiatives and delineates them at born-digital firms and incumbent companies. Building on this essay's findings, Essays 2 and 3 explain two of these enablers, namely communities of practice and data literacy training, respectively. Essay 2 examines the interplay between three types of communities of practice in a landscape of practice for data democratization, encompassing practices ranging from generic to situated. These communities engage in diverse boundary interactions, enabling practice exchanges across various roles represented in the network. Essay 3 leverages curriculum theory to propose a theory-inspired curriculum for data literacy upskilling in enterprises built on five learning blocks, four of which are highly situated. In this way firms can provide customized learning pathways tailored to the unique needs of specific audiences, encompassing both current data roles and data specialists, groups who have frequently been overlooked in the literature on data literacy. Overall, Research Stream I emphasizes the development of both generic and situated data practices among regular employees for data democratization success.

Research Stream II is titled *Data governance: balancing control and coordination of data practices*. Two essays are dedicated to elucidating data governance's role in facilitating control and coordination of data practices, thereby enabling innovation. First, Essay 4 explores how data governance practices are shaped by different strategic contexts for data. The findings disclose three distinct data governance archetypes as data role transitions from (master) data quality to data monetization. Next, Essay 5 proposes a first attempt to explain data governance through systems thinking. It clarifies how data governance practices dynamically unfold to become a viable system at the intersection of data strategy and data operations. Overall, Research Stream II highlights the pivotal role of data governance in driving value creation from data by orchestrating practices that strategically harness data for innovation.

Essay	Research question	Methodology	Key contributions	Publication status
Research Stream I. Data democratization: developing data practices beyond data experts				
Essay 1: Data Democratization: Toward a Deeper Understanding	How do companies build a data democratization capability? What are commonalities and differences between born-digital and traditional companies in building such a capability?	Explorative multiple-case research design: 8 cases of data democratization initiatives at incumbents and digital natives	Conceptualization and definition of data democratization; A set of five key enablers for data democratization	• Proceedings of the 42nd International Conference on Information Systems (2021) • *Practitioner version*: Published in Harvard Business Review Digital
Essay 2: Examining the Pivotal Role of Communities of Practice for Data Democratization in Enterprises	How do communities of practice contribute to the development and scaling of data practices in enterprises?	Explanatory multiple-case research design: embedded case studies of 45 CoPs at 12 companies	A multilevel landscape of practices for data democratization enacted by 3 key CoPs with different boundary interactions	• Proceedings of the 30th European Conference on Information Systems (2022) • *Extended version*: 1st round revision at the European Journal of Information Systems[1]
Essay 3: Toward A Curriculum for Data Literacy in Enterprises	How do companies develop data literacy programs to upscale their data practices?	Explanatory multiple-case research design: 5 case studies of data literacy curriculum	A theory-inspired and situated curriculum for data literacy in enterprises built upon five learning blocks	• Proceedings of the 57th Hawaiian Conference on System Sciences (2024) *Nominated for best paper award*
Research Stream II. Data governance: balancing control and coordination of data practices				
Essay 4: From Data as a Resource to Data as an Asset - Data Monetization as a New Frontier for Data Governance	How do companies develop their data governance practices to address the changing role of data?	Explorative multiple-case research design:9 cases of data governance approaches with varying strategic contexts for data	Three configuration of data practices in the form of archetypes that reflect the changing role of data in enterprises	• Proceedings of the 29th European Conference on Information Systems (2021) • *Extended version*: revised based on feedback received from first round review[1] and resubmitted to Information & Organization journal
Essay 5: Rethinking Data Governance: A Viable System Model	How does data governance unfold in multinational companies?	Explanatory multiple-case research design: 5 cases of data governance that combine global and local data responsibilities	A Viable System Model depicting data governance as a system at the interplay between data operations and data strategy	• Proceedings of the 32nd European Conference on Information Systems (2024) *Best conference paper award (1st runner-up)* • *Extended version* integrated into the thesis manuscript[1]

Table 4. Thesis structure: research streams and essays

[1] In Appendix, we provide a summary of the article's extensions relative to the conference proceedings.

3.2 Research context, activities and methods

The context of the research this thesis presents is the Competence Center Corporate Data Quality (CC CDQ), an industry-research consortium (Österle & Otto, 2010) active in the field of data management. The CC CDQ brings together a team of researchers and data specialists from approximately 20 multinational corporations (e.g., Nestlé, BASF, and Siemens). Operating as a form of collaborative practice research (Mathiassen, 2002), this consortium is designed to foster communication between researchers and practitioners by concentrating on their mutual interests and "*serving the general knowledge interest as well as knowledge interests that are specific for the participating organizations*" (Mathiassen, 2002, p. 10). As a result, the various research outputs are not only tailored to the diverse industrial contexts of the consortium's members; they also resonate with universal professional practices. This setting further facilitates the accumulation of design knowledge in data management (Legner et al., 2020).

The CC CDQ is composed mainly of large multinational companies from diverse industry sectors such as retail, fast-moving consumer goods, automotive manufacturing, chemical engineering, and pharmaceuticals. Given the varied levels of maturity in their data management initiatives, as well as their distinct goals and challenges, these companies facilitate a rich exchange of experiences and insights. Hence, the consortium setting provides a compelling environment for investigating exploratory and explanatory inquiries (Yin, 2018). It allows the derivation of generalizable and empirically validated knowledge, thereby ensuring theoretical relevance in the form of Type I – Theory for analyzing and Type II – Theory for explaining (Gregor, 2006). Consortium research methodology stipulates that research findings should be diffused within and beyond the participating companies to ensure their practical relevance. For this reason, our research findings were often discussed in the consortium's regular workshop activities, captured in factsheets, and adapted for publication in practitioner articles. The findings have also been presented to a wider audience in executive education programs, and in prominent practitioner outlets such as the Harvard Business Review (Lefebvre, Legner, et al., 2023).

A key aspect of any research is the perspective that the researcher adopts with respect to the "*stakeholders whose interests the researcher treats as being of primary importance*" (Clarke & Davison, 2020, p.483). The research perspective thus influences the research design and the formulation of research questions. Considering our own research goal and questions, the research perspective for this thesis must align with the viewpoint of executives with enterprise-wide data responsibilities (Lee et al., 2014). Both projects—one focused on data democratization and the other on data governance— were sponsored by the CC CDQ steering committee, which comprises data executives from each member company, typically Heads of Data and Chief Data Officers. Accordingly, the present research reflects the perspective of the central data team or central data office, which is tasked with strategically scaling data practices, rather than that of the employees enacting these data practices across the entire organization.

The first project on data democratization spanned the period from January 2021 to December 2022. It sought to build our understanding of how more employees could be empowered to work with data, focusing on the development of data practices beyond data experts. In total, our research activities entailed conducting 15 focus group meetings, which regularly involved over 35 specialists from more than 20 companies engaged in the subject matter. During these focus group discussions, participants could present their data democratization initiatives in detail, as well as add their recommendations and the challenges they faced. The second project, which spanned the period from September 2020 to December 2023, was centered on data governance. Specifically, the research activities concentrated on updating a reference model for data governance to account for the scaling of data practices. This model aimed to serve as a blueprint for companies and was intended to facilitate further benchmarking of participants' progress. These research activities were conducted in the course of 17 focus group meetings which regularly involved over 25 specialists from more than 12 companies engaged in the subject matter. These focus group discussions gave participants opportunities to present their data governance approaches, also giving their recommendations and indicating the challenges they encountered. In parallel, we conducted regular interviews with 17 partner companies' data leadership (i.e., the CC CDQ membership sponsors) to better understand their data governance practices, which helped to inform our benchmarking.

In line with the analysis phase of consortium research, we relied on qualitative research methods in capturing practitioners' knowledge, their motivations, challenges, and needs. This involved first analyzing the state of the art in literature and practice as input for the research activities, and next enriching our insight by adding the practical knowledge from participating companies. Through our focus groups, we could identify cases suitable for further qualitative scrutiny (Morgan & Krueger, 1993), which facilitated the investigation of our topic of interest through various analytical units (Dubé & Paré, 2003). To ensure the richness of each case, and strengthen the generalizability of the results (Yin, 2018), research activities extended beyond the consortium's confines to include further interactions, such as interviews. This research process supported the sampling of multiple cases to answer each research question. In an explorative manner, the data collected for each case was first analyzed individually (within-case analysis) and then compared across different cases (cross-case analysis) to observe commonalities and differences (Yin, 2018). For instance, in Essay 1, we analyzed four cases of data democratization initiatives established at digital-native companies, and four cases at incumbents, to generalize a conceptualization of data democratization. In Essay 2, we analyzed 45 data communities of practice as embedded units of analysis in 12 case companies to build a landscape of practices for data democratization. In Essay 3, we generalize a theory-inspired and highly situated data literacy curriculum by analyzing five cases of training programs. In Essay 4, we analyze nine data governance approaches to generalize archetypes of data governance practices in three key strategic contexts for data. In Essay 5, we analyze five cases of federated data governance to explain how data governance dynamically unfolds into a Viable System Model.

4 Research Stream I. Data democratization: developing data practices beyond data experts

4.1 Motivation and background

In curating the extant literature, Chua et al. (2022) distinguish five types of data practices identified as data conceptualization, data collection, data curation, data consumption, and data control. Table 5 illustrates how the significant advancements in supporting technologies and techniques, as well as the expanding scope of data from structured to unstructured forms have reconfigured these data practices. Our research focuses on what is most current, i.e., the era stretching from 2011 to the present. Thereby, our work corresponds to a period in which data enables firms' business models (see Table 1).

Data practice	**Definition**	**IS research contributions by period**			
		1970s - 1980s	**1990s**	**2000 – 2010s**	**2011- present**
Data conceptualization	The practice of grasping the methodologies for engaging with data and interpreting its implications within a work context.	Requirements elicitation & analysis approaches, specific information requirements, conceptual modeling notations and methods	Requirements elicitation, conceptual modeling methods and evaluations	Conceptual modeling methods and evaluations, ontological foundations	Conceptual modeling methods, representation theory, ontological foundations,
Data collection	The practice of accumulating data and establishing the necessary infrastructure to support data capture and storage.	Transactional databases, EDI, batch processing	Transactional databases, distributed databases	Web mining, web content design	User-generated content, social media, crowd-sourcing, big data extraction
Data curation	The practice of organizing, cataloging, and indexing data to facilitate streamlined access and retrieval.	Relational databases, data warehouses,	Data warehouses, distributed environments	Data warehouses, social media, e-commerce, data integration	Big data storage; longitudinal, dynamic data
Data consumption	The practice of processing and analyzing data to extract meaningful insights and information.	Data requirements, managerial perspective, query languages	Factor analysis, data transformation	Data visualization, web data	Data science, big data processing
Data control	The practice of implementing security measures and governance policies to ensure data integrity and compliance.	Manage data processing, cryptography	EDI audits, quality, integrity, security	Roles, responsibilities, and locus of decision-making	Data ethics, Data governance mechanisms

Table 5. Evolution of data practices (adapted from Chua et al., 2022)

IS Research on data democratization remains limited and fragmented. It has mainly focused on identifying individual means of providing data access to a larger number of employees than before, facilitating data collection (e.g., with data catalogs) and data consumption (e.g., through analytics platforms)(Awasthi & George, 2020; Labadie, Legner, et al., 2020a). Despite these efforts, focusing solely on access and tools fails to address the full scope of data democratization. True data democratization involves widespread participation in data-related activities across various business units, thereby covering a range of data practices. This view emphasizes the critical need for employees to seamlessly integrate data with domain-specific knowledge through contextually relevant data practices (Lycett, 2013), which is due to data only acquiring its value through situated (i.e., contextual) use practices (Parmiggiani & Grisot, 2020), and especially through employees' sense-making processes (Aaltonen et al., 2021; Alaimo & Kallinikos, 2022). Some authors have elaborated on the distinct ontological status of data as a portable and editable token. Consequently, data requires recontextualization every time it is employed for the distinct purpose of creating value (Alaimo & Kallinikos, 2022). Therefore, data practices cannot remain the purview of data experts only because they generally lack the context in which data consumption can create business value (Someh et al., 2023). For this reason, we take a comprehensive approach to data democratization, which should ensure that employees are not just capable of accessing data but are also equipped to shape and actively participate in data practices, and are encouraged to do so. Having identified this as a gap to be effectively addressed, we propose regarding data democratization as a capability rooted in practice, centered on the development of data practices for more users than just data experts.

4.2 Research objectives, methodology, and contributions

This research stream has three objectives. First, we aim to redefine data democratization and explain it as a matter of practice. Next, we aim to investigate the means by which companies develop data practices for users beyond the data experts.

Essay 1 addresses the following two research questions: *How do companies build a data democratization capability? What are the commonalities and differences between born-digital and traditional companies in building such a capability?*

Born-digital firms like Airbnb, Uber, and Netflix emphasize data democratization as a crucial component of their data-driven strategies. These companies integrate data usage across their workforce and embed it in their organizational culture (T. H. Davenport et al., 2020). In contrast, traditional companies, despite substantial IT investments, often struggle to effect the cultural shifts necessary to democratize data and realize its value (Bean, 2021; Kiron et al., 2012). For instance, many executives are not digital-savvy (Shah, 2021), and they overestimate their employees' data skills (T. Davenport et al., 2019).

To investigate data democratization as a complex contemporary phenomenon in both born-digital and traditional firms, we employed a multiple exploratory case study approach (Yin, 2003). We selected eight data democratization initiative cases in eight multinational companies from different industries (four born-digitals and four incumbents) and with different scopes for data. The born-digital companies, which industry experts often highlighted as models of data democratization, openly discuss their initiatives in keynotes, articles by senior employees, and on their websites. This provided rich information for our analysis. Such a diversity of data sources allowed for a comprehensive view and facilitated information triangulation, leading to the development of a preliminary theoretical framework. For the incumbent firms, we gathered primary data through a combination of focus group discussions and expert surveys, all structured around this framework. We meticulously examined each company's data democratization initiative, following with a comparative analysis to highlight the commonalities and differences in their approaches. This approach offered insights into the unique and shared challenges companies face in implementing data democratization.

As a key contribution, we identified five essential enablers for building a data democratization capability: (1) Broader data access, (2) Self-service analytics tools, (3) Development of data and analytics skills, (4) Collaboration and knowledge sharing, and (5) Promotion of data value. We show that incumbents and born-digital firms address these enablers differently and we illustrate these findings with examples from the cases (see Table 6). Our results provide an overarching conceptualization of how companies develop and strengthen their data democratization capability. Thereby we consolidate work presented in previous literature on data democratization, as the extant studies do not provide such a comprehensive overview.

Further, we propose a revised definition of data democratization, going beyond mere data access to cover a broader spectrum of data practices, enabling more employees to engage with data. We thus define data democratization as *the enterprise's capability to motivate and empower a wider range of employees - not just data experts - to understand, find, access, use, and share data in a secure and compliant way*. This definition extends those given in the current body of knowledge—in particular, the definition Labadie et al. (2020) provided in the context of data catalogs—by incorporating motivation (e.g., through corporate communication or sharing platforms) and the necessity to approach data in a secure and compliant way as is highlighted in both the cases and the literature (Awasthi & George, 2020). This acknowledges that while not all employees will be involved equally, data democratization should target a wider and more varied audience.

Enabler	Traditional companies	Born-digital companies
Broader data access	• Controlled approach to data access (donating) • Need to know about available data and sources • Emerging data catalogs	• Universal access • Need to share data internally and externally • In-house data catalogs enhanced with visualizations
Self-service analytics tools	• BI/reporting tools (limited access to relevant roles)	• Analytics experimentation platforms • Enterprise-wide access to reports and visualization
Development of data and analytics skills	• Few (optional) internal training • Focus on data literacy and data contents • Addressed at the role level	• Learning for career development • Internal academy or tailored partnerships using personas • Focus on how to generate insights
Collaboration and knowledge sharing	• Data communities • Boards or committees between data and business	• Collaborate directly on digital tools • Technical specialists sitting in business
Promotion of data value	• Emerging dedicated communication channels • Promote business value from data	• Data in the company values • Stimulate demand for data • Critical thinking and curiosity

Table 6. The five enablers of data democratization

Essay 2 deep dives into the findings presented in Essay 1 and addresses the following question: *How do communities of practice contribute to the development of data practices in enterprises?*

Data democratization should become manifest in a repertoire of data practices. These practices are developed and adapted by different roles within an organization to achieve strategic goals. In this study, we chose communities of practice (CoPs) as a prism through which to observe the collective empowerment process among members engaged in shared data practices (Wenger, 2000). In line with this motivation, we used multiple embedded case studies (Yin, 2018) to investigate how data communities emerge or how they are organized to foster data practices. We divided the data collection into two main phases: First, we used semi-structured interviews with 31 experts from 17 companies to broaden our understanding of data and analytics management practices in enterprises. This included scrutinizing the formal and informal mechanisms for alignment and collaboration between the data organization, business, and IT departments. Next, we organized two focus groups (30 experts from 13 companies, and 16 experts from seven companies) with a narrower scope on CoPs that foster data democratization. We analyzed the final sample of 45 data CoPs as embedded units of analysis in 12 case companies. First, we analyzed each individually against CoP theory, and second, we analyzed the commonalities and differences as disclosed in the individual analyses.

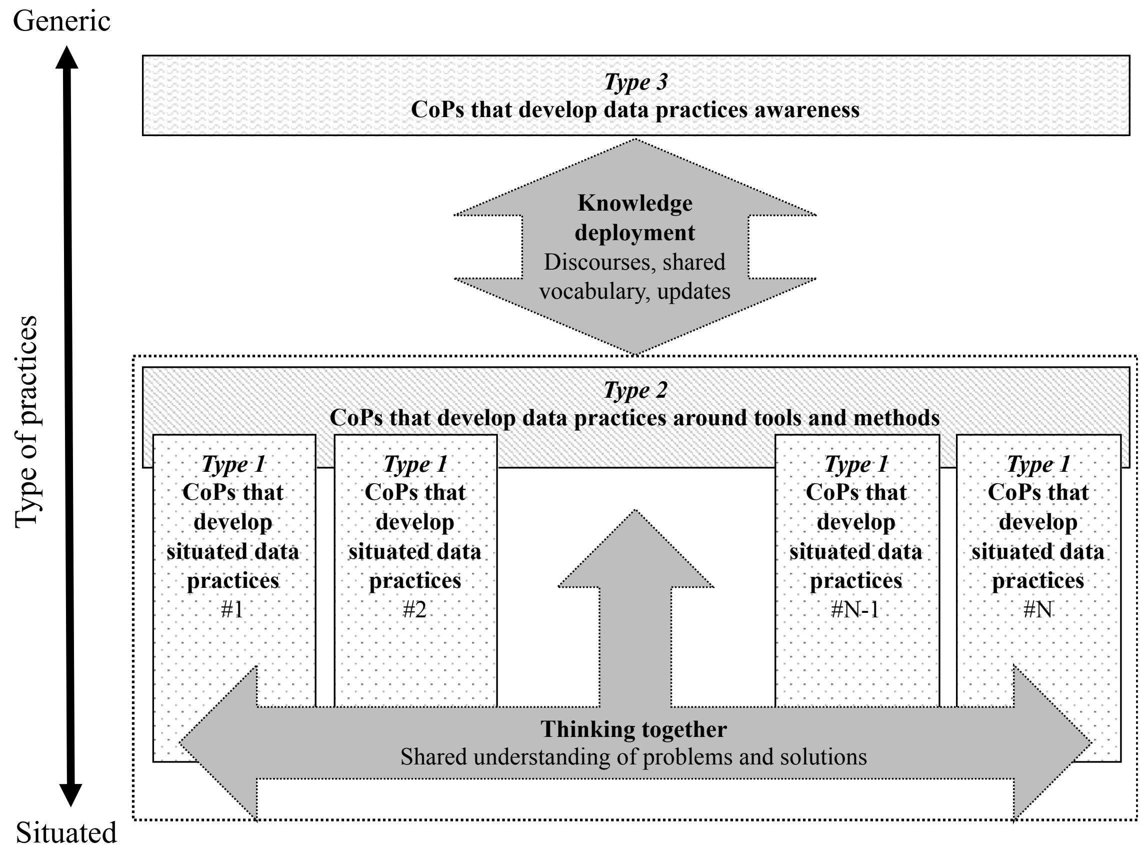


Figure 3. Landscape of practice for data democratization

We find that observing data democratization via the concept of CoPs provides a theory-informed and practically grounded understanding of the informal structures that connect employees working with data across teams, and in a shared data practice. CoPs thereby help overcome existing barriers to data democratization, such as poor collaboration in data projects, limited knowledge sharing with experts, and general lack of competence. Specifically, we disclose a multilevel landscape of practices where three types of CoPs foster data practices at different levels of situatedness —from generic to highly situated— and encounter each other through different boundary interactions to foster data democratization. Type 1 CoPs *develop situated data practices*, for instance by focusing on data-driven innovation or data lifecycle improvements. Endowed with business acumen, the CoP members collectively work toward refining data practices to align them with the demands that enable efficient and innovative data use in a specific work context. Type 2 CoPs *develop data practices with a focus on tools and methods*, for instance by attending to data management methods (e.g., data architecture) or analytics techniques (e.g., reporting tools). Members of such communities aim to gradually enhance their skills by learning from each other's experiences, and especially from the data experts. In this way a CoP collaboratively cultivates new competencies. Type 3 CoPs *develop data practices awareness*. By fostering generic enablement, these CoPs engage participants across the data spectrum, from beginners to experts, aiming thereby to enrich the organization's data culture.

Essay *3* is based on the following research question: *How do companies develop data literacy programs to scale their data practices?*

Prior studies on data literacy have mostly focused on educational settings and identified data-related skills. However, the suggested generic skill catalogs do not account for the highly situated nature of data practices. To address this gap, and to generalize a new curriculum for data literacy (Miles et al., 2014), we opted for multiple case studies which are suitable for capturing rich and diverse insights directly from practitioners' working contexts (Paré, 2004). Using a combination of focus group discussions and expert interviews, we collected rich insights on five data literacy programs with different scopes and target audiences conducted in five multinational companies. We analyzed them individually and compared them using insights from curriculum theory as theoretical framework, relying specifically on Bennet et al.'s (1999) theory which distinguishes between generic and discipline-specific learning outcomes.

Our results show that firms opt for data literacy curricula which offer personalized learning paths that address specific audience needs, tailored to other roles than those of data experts, thus focusing on users who have often been neglected in existing data literacy research. We, therefore, reaffirm the contextual nature of data practices. We present a curriculum for data literacy development (see Figure 4) built on five blocks, each addressing different learning outcomes, i.e., generic skills, disciplinary content, disciplinary skills, workplace awareness, and workplace experience.

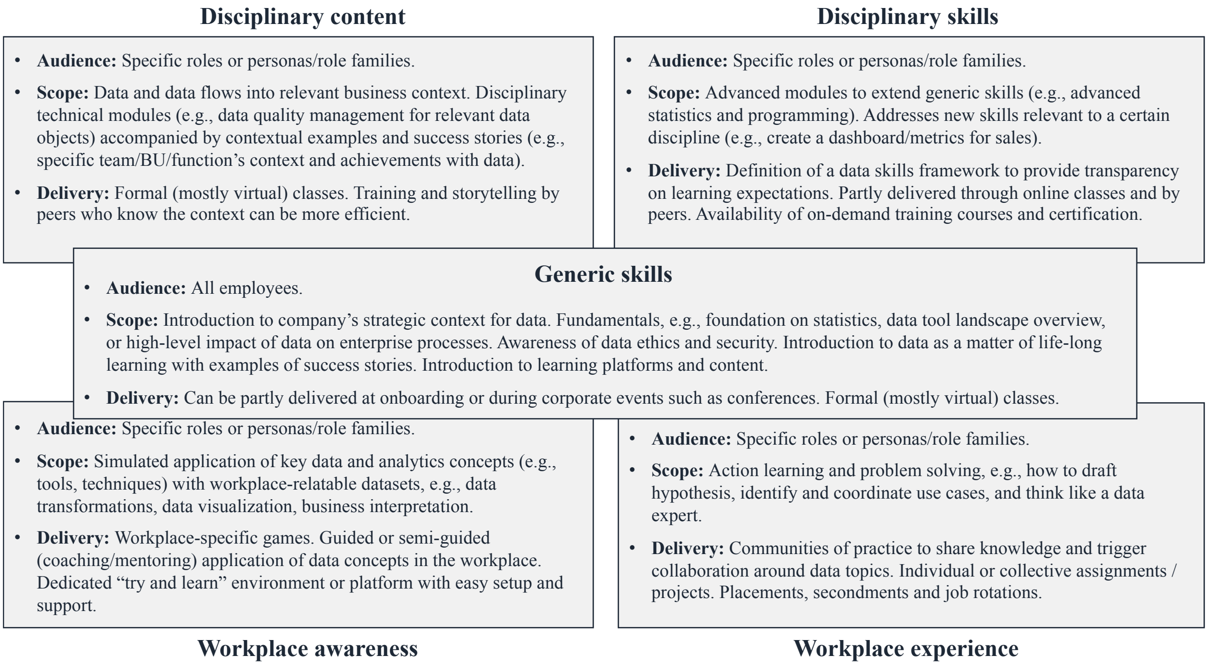


Figure 4. Data literacy curriculum

Generic skills include foundational skills for all employees by providing a set of common baseline skills which are transferrable to various working environments and generally enable employees with different data backgrounds to work, collaborate, and communicate with others about data or in projects. The *disciplinary content* shows how and which data can be used in trainees' specific (business) context. *Disciplinary skills* are outcomes that aim to transform disciplinary content into situated data activities, i.e., they are the skills employees require to use data in their daily work. To develop *workplace awareness* the curriculum aims to support the application of theoretical knowledge in a simulated environment, as authentically as possible. In developing *workplace experience* trainees are invited to take on data responsibilities and commit to a continuous learning journey in active job duty. For each of these five learning outcomes, we indicate target audiences, the scope, and means of delivery. We also highlight the importance of learning formats and experiences that shape the learning experience as closely as possible to the workplace reality. Depending on the scope of the training program, all blocks or only a subset of them can be activated.

4.3 Discussion, limitations, and outlook

The different contributions of this research stream inform on how to develop the data experts' data practices alongside those of other employees. More specifically, we shed light on how companies balance the development of generic and workplace-specific data practices by means of practice-based learning programs (Essay 2) and through CoPs (Essay 3). On the one hand, we confirm that data literacy development cannot be detached from practice. Although all relevant employees should develop a common baseline of generic skills, most data practices should be developed through disciplinary-relatable learning outcomes, which as closely as possible resemble the trainees' workplace reality. Therefore, developing data practices by means of a curriculum requires that cognitive expectations be defined for various data roles that are aligned to the context in which data practices will be performed. On the other hand, we shed light on the prominent role of practice exchange between and across roles in competence development. By crossing boundaries between different CoPs, practitioners gain knowledge of other practices, ranging from generic to highly situated knowledge, developed through legitimate peripheral participation and situated learning (Lave and Wenger, 1991).

This first research stream is not without limitations. The first set of limitations is inherent to our consortium research setting and questions the generalization of our findings to smaller firms. However, we believe that the natural variation in our sample, for instance in terms of industry, size, and maturity, strengthens our theory for multinational companies. Second, as for any analytic and explanatory theories, we believe our research could benefit from empirically testing these theories. Further, certain findings of this research stream, such as the curriculum model, could inform subsequent theories for design and action.

The findings from this research stream hold promising prospects for future inquiry. As our analysis does not fully cover the lifecycle of CoPs, a detailed exploration of their developmental stages could provide insight into how sustained practice exchange supports their evolution. Additionally, refining the curriculum model could enrich theoretical frameworks for design and action, potentially guiding innovative artifact creation. Further, the burgeoning field of generative AI, which was not well-established at the time of these studies, offers fertile ground for examining data democratization. This emerging technology could reshape the five key data practices, suggesting a need to reassess our findings' relevance in this new context. Future research could also delve into how generative AI redefines data consumption, examining the altered cognitive demands needed to leverage this technology effectively, thereby enhancing our proposed curriculum model.

5 Research Stream II. Data governance: balancing control and coordination of data practices

5.1 Motivation and background

As indicated earlier in giving the theoretical background (see section 2.2), prior studies have provided foundational insights into defining data governance's scope and framework by means of three overarching governance mechanisms. This prior knowledge is increasingly being challenged because it acts mainly as a top-down framework emphasizing control over data. Instead, it has become clear that data governance must also coordinate the growing use of data to support innovation across the entire organization. As data is increasingly used to innovate across the organization, more sophisticated data governance capabilities are needed to support growing day-to-day data practices (Aaltonen et al., 2021; Alaimo & Kallinikos, 2022; Benfeldt et al., 2020; Legner et al., 2020), such as actual patterns of data production, use, and reuse (Parmiggiani & Grisot, 2020). Therefore, data governance can no longer be seen as a "*series of mechanisms implemented in organizations, at the expense of understanding the process of governing data*" (Vial, 2023, p. 6).

A first related concern is that the literature does not explain how these mechanisms evolve as the role of data changes. Despite evidence of antecedents to data governance (Abraham et al., 2019; Tallon et al., 2013), such as data strategy, research still mainly considers it as a set of three unchanging data governance mechanisms. This approach overlooks the dynamic nature and evolution of data governance frameworks that could adapt to new challenges and technological advancement (e.g., analytics). Therefore, there is still a lot to learn about how data governance evolves to support innovation.

In addition, given global firms' complex organizational structures, establishing fit-for-purpose data governance for them remains a challenge (Otto, 2011b). To fully appreciate the dual role of data governance, it is essential to examine how such governance coordinates data practices across an organization's core structure composed of functions, divisions, or regions established to support innovation. To be effective, data governance must reach many different parts of an organization and shape the situated data practices through which data acquires its value. That is why researchers have argued for "*a shift from data governance as a matter of asset management to data governance as a matter of* *work practice*" (Parmiggiani & Grisot, 2020, p. 3). Federated data governance models, which combine global and local data governance responsibilities, have been proposed as a solution in rolling out data governance to fit the primary organizational structure (Grover et al. 2018; King 1983). However, so far, no link has been established for understanding how data governance mechanisms materialize at local and global levels. Further, the rather static view of data governance mechanisms does not properly explain the dynamic nature of data governance which must evolve in symbiosis with strategy and operations (Benfeldt et al., 2020). Hence, it is critically important to explain how data governance unfolds in practice (Parmiggiani & Grisot, 2020; Vial, 2023).

5.2 Research objectives, methodology, and contributions

This study aims to develop a better understanding of how data governance balances control and coordination of data practices in large organizations. First, we explore the implications of scaling data practices for data governance. Second, we explain how data governance dynamically unfolds in large organizations with complex organizational structures.

Essay 4 answers the following research question: *How do companies develop their data governance practices to address the changing role of data?*

This research aims to provide a more thorough understanding of how data governance evolves when data's role changes in enterprises and companies move toward more offensive data strategies. Like most studies in the field of data governance (Otto, 2011c; Parmiggiani & Grisot, 2020; Tallon et al., 2013), we adopted a qualitative research design to explore data governance dynamics in their true complexity, shaped by a multitude of internal and external influences (Dubé & Paré, 2003). However, unlike prior studies that focused mainly on single cases, we adopted a multiple case study approach to investigate different strategic and industry contexts for data (Yin, 2003). Using a theoretical replication logic, we selected a diverse set of nine multinational companies representing different industries, strategic contexts, data scope, and experience with data as these are key contingencies for data governance (Abraham et al., 2019; Tallon et al., 2013). In this way we could analyze how their data governance practices are enacted in different organizational contexts, and we could study the variations between them.

	DATA GOVERNANCE ARCHETYPES		
	Archetype I ***Improve master data quality***	**Archetype II** ***Establish enterprise-wide data transparency***	**Archetype III** ***Enable data monetization***
STRATEGIC DATA OBJECTIVES AND DATA SCOPE			
Data strategy	Improve data quality to enable business processes/reporting	Improve data quality to enable business processes/reporting, broaden data access/availability to enable value creation	Improve data quality, broaden data access/availability, monetize data
Data scope	Narrow focus on master and reference data and few data domains (e.g., Supplier, Customer, Product, Material)	Broad focus on any data type and increasing number of data domains (e.g., Finance, HR, Controlling)	Broad focus on any data type including analytical data and stable number of data domains (any relevant)
STRUCTURAL MECHANISMS			
Shape the data organization	Small central data organization aligned with business and IT through projects or master data boards	Growing central data organization with decentral staff allocation or role assignment to business stakeholders, and emerging boards to decide on data governance principles with business	Large, federated data organization relying on divisional, functional and regional data governance hubs. Boards and councils to connect within and across the network
Assign data roles and responsibilities	Only essential data roles (head of data management, data steward, data architect)	Additional central oversight completed with the expansion of data steward and data owner roles into the business	Complete role model addressing strategic (Chief Data Officer), governance (e.g., data quality manager, data documentation manager) and operational roles (e.g., data citizen, data editor)
PROCEDURAL MECHANISMS			
Define and monitor data strategy and investments	Isolated strategy planning activities, investments in data quality improvements and infrastructure	Emerging data strategy planning process, investments in data quality improvements and infrastructure, business case analysis for new data domains	Data strategy planning and control process, pro-active identification, and management of data monetization opportunities
Define and enforce data principles	Creation of standards and data models for master data	Data governance framework and process for data modeling and architecture design	Data and analytics data governance framework, unified data architecture
Manage data operation	Data quality monitoring and support	Data quality monitoring and support, coordinated data lifecycle management	Data quality and use monitoring and support, and data lifecycle management in functions
RELATIONAL MECHANISMS			
Align and collaborate with business & IT stakeholders	Mostly through procedures or extended boards. Collocation with 1-2 data roles in IT functions	Collocation with an extended array of responsibilities for data-related aspects in IT function.	Collocation or even combined with a focus on delivering data and analytics products
Develop and share data knowledge	Few communities for master data. Few training options for non-specialists besides about compliant access and use	Regular updates. Emerging community management. Training in data quality methods and data literacy.	Enterprise-wide promotion of data. Personalized data literacy learning paths with peer coaching.

Table 7. Data governance archetypes

We find that the evolving role of data, and the move toward data monetization go hand in hand with the expansion of existing and the development of new structural, procedural, and relational data governance practices. As a key contribution, we derive three archetypes that characterize typical ways of governing data and reflect the changing role of data toward a strategic asset, which we articulate as (1) improve master data quality, (2) establish enterprise-wide data transparency, and (3) enable data monetization (see Table 7). Strategic management defines archetypes as configurations that are context- specific and are identified based on an array of organizational features (Short et al., 2008). While the two first data governance archetypes focus on defining accountabilities, data standards, and policies on a growing data scope, we find that data monetization involves extending structural practices for strategic decision-making and investment planning. Moreover, expanding data practices in the extended network cannot happen without developing new relational governance practices that foster data literacy and knowledge sharing. Our rich empirical insights thus inform both researchers and practitioners on how companies implement data governance according to different strategic phases.

Essay 5 addresses the following research question: *How does data governance unfold in multinational companies?*

Inspired by systems thinking, we framed data governance as a system dynamically shaped by its environment, and composed of a set of interrelated global and local elements. As a system, it must maintain its dual purpose of controlling and supporting data innovation. This allowed us to use the Viable System Model (VSM) as theoretical lens as it explains a system's ability to maintain its existence in a changing environment (Beer, 1985).

In the context of developing our reference model for data governance (see section 3.2), we organized nine focus groups with 34 high-profiled data experts, where participants provided an overview of their data governance approach, as well as described its evolution over time. Using purposeful sampling (Patton, 1990), we identified five companies' data governance approaches for the subsequent case study analysis. We selected these companies for their diverse characteristics regarding their industry, the goal and scope of their data governance, and different organizational structures that influenced the design of global and local data governance teams. The case companies had implemented federated data governance design decisions, e.g., they had complete role and process models at global and local levels. To gain in-depth insight regarding the five companies' federated data governance approaches, we conducted semi-structured interviews with key informants who had been mandated to oversee enterprise-wide data governance in the case companies. In analyzing our data, we applied abductive reasoning because it allows for embedding empirical findings into an existing theoretical model (Ketokivi & Mantere, 2010). This approach facilitated theorization through a detailed examination of the data by employing inductive coding for categorizing interview data and deductive coding for incorporating the VSM perspective in each case. We then studied the variations between them.

Systems		**Theory** (Beer, 1985)	**Description**	**Data practices**	**Layer**
System-in-focus	**S1**	Describes the different operative units that execute the tasks expected to fulfill the system's purpose.	Represents all business units where data practices are embedded in work practices and performed by providers and consumers of data.	• Data creation • Data curation • Data usage	**Operations**: *Perform data practices*
Meta-system	**S2**	Handles coordination and communication across the different S1s, especially during disturbances affecting the VSM (e.g., environmental fluctuations).	Ensures coordination between data governance teams by assigning data roles and responsibilities and distributing the latest governance principles to the entire network. It also provides data management support, training, and data applications to data providers and consumers.	• Definition of data roles and responsibilities • Data enablement • Data management support • Data documentation and architecture • Data applications management	**Governance**: *Orchestrate data practices*
	S3	Oversees the activities of the system-in-focus (S1) through "day-to-day management" to ensure the smooth delivery of data operations against strategic goals.	Oversees all data practices in the system-in-focus (S1) and ensures that they are performed in line with strategic goals and according to standards and guidelines (e.g., for data collection, storage, use, documentation). Monitors the execution of the data strategy and provides periodic strategic reporting.	• Definition of data standards and guidelines • Performance monitoring and improvement	
	S3*	Complements System 3 act as a compliance system of operative unit (S1).	Performs data-related audits of operative units to ensure compliance with laws, regulations, and standards.	Data compliance auditing	
	S4	Senses data threats and opportunities to the system by scanning the environment.	Senses data opportunities (e.g., trends) and threats (e.g., compliance) that could impact the data organization.	Data threats and opportunities sensing	**Strategy**: *Shape data governance practices*
	S5	Maintains the system's identity by describing the system's norms and purpose.	Provides strategic direction for the entire data activities in alignment with company strategy.	Data strategy definition and monitoring	

Table 8. VSM sub-systems and their application to data governance

As a key contribution, we theorize a VSM for federated data governance designed to adeptly navigate the delicate equilibrium between control and innovation within organizations. The model shows that strategic, governance, and operational data practices unfold on multiple, interconnected levels, corresponding with five distinct subsystems, each with its specialized data practices (see Table 8). Specifically, we give evidence of 13 systemic data practices embedded in strategic (2 data practices), governance (8 data practices) and operational processes (3 data practices).

As shown in Table 8, S1 refers to “operative units” that are typically business functionaries who are providers and consumers of data, and who thus embed data in their work practices. S2, S3, S4, and S5 together form a metasystem including the totality of operational data practices performed in operative units (S1). In this metasystem, S2, S3, and S3* represent the data governance layer (i.e., the data governance teams) that orchestrate data practices. S2 refers to “Coordination” which is managed by the data governance team, be it at the global or local level. This subsystem’s role is to communicate about data governance and to coordinate the network of data providers and consumers (S1). Thereby, it ensures alignment at enterprise-wide level, be it between data providers and data consumers within one operative unit (S1) or between several operative units (e.g., data sharing between customer and sales data domains). S3 refers to “Control monitors,” a subsystem managed by the data governance team who monitors all data practices in S1 and ensures that they are performed in line with strategic goals and according to given standards and guidelines. At the interface of operations and strategy, System 3’s role is pivotal for standardizing data practices, and for delivering and reporting strategy. S3* refers to “Audit” which complements System 3 by auditing operative units’ data practices to thereby ensure that they align with legal requirements, industry standards, internal policies, and data standards and guidelines. S4 and S5 form the strategy layer of the metasystem (through boards and committees) and shape data governance practices. S4 refers to “Intelligence” which coordinates data executives to ensure that the whole system can adapt to perturbations by scanning changes in the environment (e.g., new data trends, use cases) and by proposing mitigation plans. This, in turn, informs S5 which refers to “Policy” and provides the strategic direction for the data activities in alignment with the strategic business priorities.

Further, we find that the various data practices unfold through several local systems-in-focus (i.e., multiple System 1). Consequently, global data governance practices can be distributed by being embedded, and often enriched, in local systems (e.g., regions, divisions, functions). This indicates a recursive logic with two (possibly more, depending on organizational structure) systems-in-focus, which are (1) at level “n”, the totality of corporate data practices governed by global data governance practices, and (2) at level “n+1”, local data practices governed by local data governance practices. The recursive metasystems are designed to mirror, to varying degrees (e.g., in definition of a local data strategy, country-specific regulations sensing), the principal overarching metasystem, thereby ensuring coherence and alignment with the global data governance framework while still accommodating local nuances.

5.3 Discussion, limitations, and outlook

This research stream enriches data governance research by explaining how data governance can maintain its dual purpose of simultaneously balancing control and innovation. Acting as the orchestrator of operational data practices, data governance emancipates from IT governance and undergoes a dynamic evolution characterized by its practices being reconfigured to align with the broader strategic context for data. In this way data governance can be viewed as a system that responds to immediate needs and proactively supports overarching strategic goals.

These two interrelated essays are among the first to position data governance as a dynamic system that mediates the interface between data strategy and data operations. These advances place data governance at the interface where strategic considerations for data intersect with day-to-day operational practices. Our findings introduce a dynamic perspective to conventional portrayals of data governance as a fixed set of structural, procedural, and relational mechanisms inherited from IT governance. Instead, it is through actively enacting these mechanisms in comprehensive repertoires of data governance practices that we attained a more nuanced and profound understanding of data governance's contribution to value generation from data. In contrast, through the application of the VSM, we could demonstrate that data governance encompasses distinct yet complementary roles that fulfil a control function ensuring adherence to policies and standards, and a coordination function facilitating seamless interaction between stakeholders. This dual-role framework casts the pivotal function of data governance as one coordinating data practices across diverse organizational units, thereby catalyzing the formation of an extensive and interconnected network of data practitioners.

Further, we improve the understanding of federated data governance as a suitable model for achieving both control and innovation by displaying it as a recursive viable system, typically aligned to the organization's primary structure. These findings address limitations in prior research which primarily delved into global data governance by examining its structural, procedural, and relational mechanisms, often simplistically distinguishing responsibilities as either global or local (Grover et al., 2018; Otto, 2011c). Additionally, this stream offers precious insights into how companies can achieve federated data governance through recursive data governance practices. Essay 2 reveals that, rather than a simple split between global and local, global data governance practices can be replicated at the local level, with subtle differences in how these practices are executed. Therefore, our findings also contribute to us better understanding the roles of relational mechanisms, which have thus far been overlooked in research (Vial, 2023). We conceptualize these mechanisms as communication channels between various global and local data roles. Essay 1 shows that relational mechanisms must grow when the role of data is changing toward data monetization, usually in tandem with the development of a federated data governance model. In this way local innovation can be achieved taking the local environment into account, and complying with global data governance principles.

This research stream is also not without limitations. Its empirical foundation is rooted in the reality of our consortium's multinational companies with highly specialized and compartmentalized organizational structures. Consequently, we cannot easily generalize our findings to smaller firms, characterized by more versatile roles and limited resources. Furthermore, our focus on the evolution of data governance practices, distinct from the realms of IT governance, could have overlooked critical synergies and tensions between these domains (e.g., related to responsibilities such as data architecture, or to applications). Our investigation of these intersections delineates a compelling trajectory for future research.

By extending systems thinking to data governance, we provide new fertile ground for future scholarly endeavors. Our findings pave the way for exploring other elements constituting each sub-system. Since this study primarily explicated the five sub-systems through their underlying data practices, we suggest that researchers go forward with an in-depth characterization of each subsystem, for instance by identifying their input and output. We also propose that researchers continue examining the VSM's extensive attributes (e.g., communication channels between all subsystems), on which our initial inquiry has merely touched. Also, future research could examine how strategies and operating environments across different industries affect the system's viability. In this context, applying VSM theory principles such as variety and transduction (Beer, 1985), which this study has not extensively covered, offers valuable opportunities to refine the model.

6 Contribution, Discussion, and Implications

6.1 Contributions

Contributing to the broader data management body of knowledge, this thesis clarifies how data practices are scaled through the interplay between data democratization and data governance. The two research streams elucidate this essential interplay, enabling the scaling of data practices that drive value creation. A summary of each research stream's contribution can be found in Figure 5.

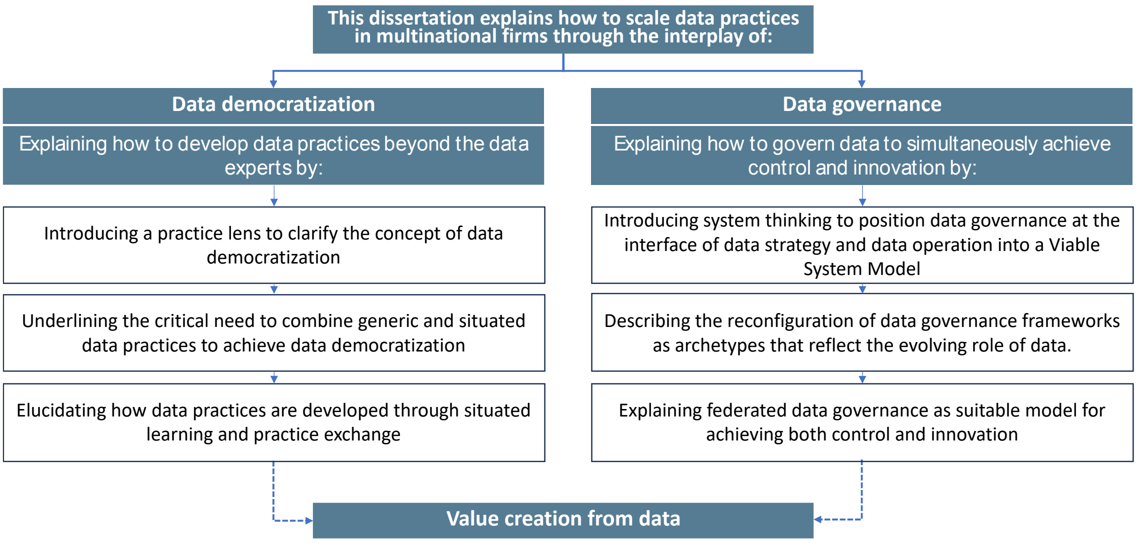


Figure 5. Overview of the dissertation's contribution to data management research

Our findings on data democratization explain how to develop data practices relevant to users beyond the data experts. We first anchor data democratization in IS research as a capability grounded into practice and describe its socio-technical nature. We then underline how the critical need to combine generic and situated data practices to achieve data democratization. Finally, we elucidate how data practices are developed through situated learning and practice exchange.

Our findings on data governance explain how to govern data to simultaneously achieve control and innovation. We first introduce system thinking to position data governance at the interface of data strategy and data operation into a Viable System Model. We then describe the reconfiguration of data governance frameworks as archetypes that reflect the evolving role of data. Finally, we explain federated data governance as suitable model for achieving both control and innovation.

6.2 Discussion

This thesis demonstrates the close interplay between data governance and data democratization to scale data practices in enterprises. It has direct implications for IS research and for practice.

Clarifying the concept of data democratization in enterprises

We explain how data democratization capabilities can be built, going beyond isolated aspects of the entities, combining relevant features to provide a thorough conceptualization. By enriching the capability with a practice perspective, we complement and extend the prominent access-based perspective on data democratization which, until now, has not been able to explain how situated data practices transcend the confines of the data experts' realm. We demonstrate that making data accessible and cultivating robust data practices allows individuals across various organizational levels to uniquely contribute to creating value from data. This democratic approach challenges the traditional expert-centric models of data-handling and decision-making. Overall, we offer a view on data democratization "*influenced by a combination of factors such as data resources, competencies, and practices that work together to drive business results*" (Xu et al., 2023, p. 9).

Informing on how to develop generic and situated data practices

Our findings suggest that such democratization can effectively serve a variety of user groups or personas. Therefore, it is essential to conduct a thorough mapping of all enterprise data practices as they relate to the specific employees who are implementing them (Alaimo & Kallinikos, 2022). As highlighted into our VSM, this mapping enables a clearer understanding of how data practices are deployed and adapted within different segments of the organization.

We could identify that the nature of these data practices varies widely. Some are generic, meaning they can be transferred and applied across multiple contexts, enhancing flexibility and broader applicability; others are situated, designed to address the needs of specific, often unique, contexts, thereby ensuring targeted and relevant application. To further develop and refine these practices, we underscore the importance of CoPs and the implementation of practice-based curricula. Such communities and curricula are pivotal because they promote an environment of active learning and knowledge exchange (Nicolini et al., 2022). Through regular practice exchanges between these communities, individuals not only share, but also evolve their expertise, leading to more innovative and effective data use across the organization. This approach not only enhances individual competency, but also fosters a more collaborative and informed organizational culture.

Rethinking data governance as controller and coordinator of data practices

To summarize and anchor our contribution to data governance research, Table 9 shows the timely academic and practical relevance of the contributions by mapping them onto the research outlook proposed by Vial (2023).

Themes proposed by Vial (2023)	**Opportunity for IS research suggested by Vial (2023, p. 6)**	**Thesis contribution**
Embracing data governance without compromising digital innovation	"*Research can draw attention to the possibility to espouse the two objectives of data governance as paradoxical, fostering the implementation of mechanisms (e.g., data stewards) that can help to reconcile both objectives*"	The VSM (combined with the archetypes) elucidates the dual function of data governance as both a controller and coordinator of data practices.
Enacting data governance through repertoires of mechanisms	"*Research can help to develop a conceptualization of data governance as repertoires of mechanisms that form configurations that contribute to the achievement of organizational outcomes*"	The archetypes and the VSM reveal the intricate relationship between planned organizational outcomes and the requisite configurations of data governance practices.
From data governance to governing data	*"Like strategy, data governance incorporates both planned and emergent components, calling for approaches that are closer to the practice of governing data and its impact on everyday work"*	In the VSM, data governance shapes data practices but also introduces new relational mechanisms that facilitate practice exchanges and the development of data practices.

Table 9. Overview of research stream's contribution to data governance research

Embracing data governance without compromising digital innovation

Our research illustrates how data governance transcends its conventional role as merely a mechanism of control. It now emerges as a pivotal coordinator and enabler of data practices that balance the paradoxical objectives of fostering innovation while ensuring robust data governance. The two essays in this stream portray data governance practices as adaptive and responsive to sensing data-related opportunities and threats (e.g., data strategy), and capable of orchestrating operational practices that deliver innovation. This perspective repositions data governance as a pivotal function that directs operational data practices, ultimately steering them toward the achievement of strategic business goals.

Enacting data governance through repertoires of mechanisms

At the nexus of data strategy and data operations, data governance infuses the necessary dynamism, acknowledging the strategic significance of data and ensuring that governance frameworks are not only about policy and compliance but also about enabling strategic innovation. This is showcased by allocating data practices to both the VSM's control and coordination subsystems, which explains how data governance permeates the entire organizational structure. This thesis further explores how a repertoire of data governance mechanisms, such as architectural standards, decision rights, and roles, can be conceptualized and deployed to form configurations that actively contribute to achieving key organizational outcomes. By situating these mechanisms within various structural components of the organization, data governance effectively becomes embedded.

From data governance to governing data

Our findings show a strong connection between data governance practices and operational data practices, thereby confirming recent studies that argue for a user-centric perspective in data governance frameworks (Parmiggiani & Grisot, 2020, p. 2). For instance, we show that data creation, data curation, and data use are operational data practices that should adhere to data governance principles. We position data governance not just as a policy or framework but as an everyday governance tool that shapes and is shaped by everyday work practices. This approach highlights how governance integrates with the emergent and planned components of organizational strategies, thus aligning data governance more closely with the realities of work practices. It underscores the importance of governing data through practical, elemental mechanisms that directly influence how data is handled daily, making data governance a more tangible and integral part of daily operations in the organization.

Refining the understanding of all enterprise data practices

Our findings have implications for the understanding of data practices. They offer a new perspective on the five essential data practices given in the literature, and especially on data control and data consumption. On the one hand, through applying systems thinking to data governance, this study's findings position data governance as a subsystem at the interface of strategy systems and operational systems. Hence, as Chua et al. (2022, p. 4) anticipated, the control data practices should evolve toward a *coordination* data practice that confronts the "*centralization or decentralization of overall information management functions.*" On the other hand, the resulting VSM addresses the entirety of data practices (strategic, governance, operational) performed in the data organization. Unlike prior studies on analytics governance (Baijens et al., 2020), this integration further reveals that data governance alone does not suffice as a comprehensive meta-system for managing data operations. Rather, it is through integrating data strategy and data governance into the meta-system that operational data practices are performed to achieve innovation at scale. Consequently, our findings have implications for data consumption that should align with new strategic approaches to using data.

6.3 Implications for research

Looking ahead, our findings have major implications which open multiple pathways to investigate more deeply how companies scale data practices toward greater value creation from data. Representing the leading frontiers of our exploratory journey, the following three areas are highly promising for both practice and research.

Further investigation of data governance as a dynamic system

This thesis reframes data governance as a dynamic system that plays a crucial role in supporting organizational objectives and thereby scaling data practices. In an increasingly competitive and data-driven business environment, it is imperative for future research to continue refining and expanding the VSM application in organizations. As an extensive theory, the relevance of VSM should be further studied, going beyond the identification of data practices in each subsystem. One of the areas that would require more attention is the variety (i.e., the extent of change) in the VSM following different external disturbances (e.g., reaction to the EU AI act) which may have an impact on the organizational performance. Therefore, understanding the flow of information and its processing within this model could help us identify where the weak links are. In turn, this would support the refinement of coordination mechanisms, while the practical integration of VSM into the primary structure of organizations has yet to be fully realized. More specifically, continued investigation would provide fertile ground for research to develop methodologies and frameworks that facilitate the implementation of VSM beyond its theoretical foundations.

Deeper understanding of user-centric data governance

Understanding user-centric data governance complements previous research that has shown how roles and responsibilities related to data are applied to three different actions (Alhassan et al., 2016), namely *define*, *implement*, and *monitor*. This responsibility split requires further coordination between data governance professionals who provide the data governance frameworks, and employees who implement them in the context of situated data practices (Parmiggiani & Grisot, 2020), bringing to the fore the importance of relational mechanisms, which to date have mostly been underemphasized (Vial, 2023). Since our inquiry did not delve into the implementation, further research could combine and refine our findings with this perspective using a similar approach (e.g., through practice exchange). Building on our findings, this line of research should include a more detailed investigation of the cognitive evolution and skills development of these new roles, as well as the practice exchange emerging between them.

Inter-organizational data governance for value creation into data ecosystems

Despite the significant potential for value creation from data within broader data ecosystems (Gelhaar et al., 2023), research into inter-organizational data governance remains notably scarce. Although Abraham et al. (2019) suggest that a similar data governance framework can be applied to an inter-organizational context, recent research indicates that the latter is much more grounded in practice (Lefebvre, Flourac, et al., 2023). This shift brings new challenges regarding data sharing and data sovereignty (Abbas et al., 2024). Future studies could delve into the mechanisms by which organizations can jointly manage shared data resources, respecting each other's governance structures, while also striving for mutual benefits. Further, research should focus on the practice exchange between data experts from different companies, exploring how such collaborations can foster innovation, enhance data quality, and lead to more robust governance practices. These have implications that could significantly influence the development of consortiums, partnerships, and alliances centered on shared data operations.

6.4 Implication for practice

From the perspective of practitioners, this thesis makes a significant contribution by offering concrete, actionable guidance on implementing data governance frameworks, with a particular focus on federated models within multinational corporations.

Further, the thesis identifies and outlines specific archetypes of data governance practices. These archetypes serve as adaptable blueprints that facilitate the customization and implementation of data governance setups tailored to the specific needs and objectives of each organization. Such adaptability is crucial for multinational organizations that must navigate varied regulatory environments and business cultures, allowing for the creation of bespoke governance strategies that align with their specific needs and goals.

Additionally, thanks to its qualitative approach, this research offers practitioners a multitude of real-world case studies from firms actively investing in their data governance. These examples serve as valuable reference points for organizations wanting to enhance their own data governance strategies, offering insights into best practices and innovative approaches to data management.

In addition to these strategic and organizational insights, the thesis provides concrete insights toward democratizing data in organizations. Besides informing on concrete enablers firms can invest in, the thesis also elaborates on the relevance of specific data practices at different organizational levels, contributing to the development of an extended data governance framework. This is complemented by a practical curriculum model developed to boost data literacy across the organization, along with recommendations for leveraging key communities toward the exchange of data practices.

7 References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. *Journal of Management Information Systems*, *38*(2), 401–429.

Abbas, A. E., Van Velzen, T., Ofe, H., Van De Kaa, G., Zuiderwijk, A., & De Reuver, M. (2024). Beyond control over data: Conceptualizing data sovereignty from a social contract perspective. *Electronic Markets*, *34*(1), 20.

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.

Alaimo, C., & Kallinikos, J. (2022). Organizations Decentered: Data Objects, Technology and Knowledge. *Organization Science*, *33*(1), 19–37.

Alavi, M., Georgia Institute of Technology, Leidner, D. E., University of Virginia, Mousavi, R., & University of Virginia. (2024). Knowledge Management Perspective of Generative Artificial Intelligence. *Journal of the Association for Information Systems*, *25*(1), 1–12.

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems*, *25*(sup1), 64–75.

Awasthi, P., & George, J. J. (2020). A Case for Data Democratization. *Proceedings of the 26th Americas Conference on Information Systems (AMCIS)*, 23.

Baijens, J., Helms, R. W., & Velstra, T. (2020). Towards a Framework for Data Analytics Governance Mechanisms. *Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference*.

Ballou, D., Wang, R., Pazer, H., & Tayi, G. K. (1998). Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science*, *44*(4), 462–484.

Bean, R. (2021). Why Is It So Hard to Become a Data-Driven Company? *Harvard Business Review*. https://cdn.ampproject.org/c/s/hbr.org/amp/2021/02/why-is-it-so-hard-to-become-a-data-driven-company

Bean, R. (2023, January 30). Has Progress on Data, Analytics, and AI Stalled at Your Company? *Harvard Business Review*. https://hbr.org/2023/01/has-progress-on-data-analytics-and-ai-stalled-at-your-company

Beer, S. (1985). Diagnosing the system for organizations (Repr). Wiley.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, *22*(2), 299–313.

Bennett, N., Dunne, E., & Carré, C. (1999). Patterns of core and generic skill provision in higher education. Higher Education(37), 71–93.

Chen, D. Q., Preston, D. S., & Swink, M. (2021). How Big Data Analytics Affects Supply Chain Decision-Making: An Empirical Analysis. *Journal of the Association for Information Systems*, *22*(5), 1224–1244.

Chua, C., Indulska, M., Lukyanenko, R., Maass, W., & Storey, V. C. (2022, February 14). *MISQ Research Curation on Data Management*. MIS Quarterly. https://www.misqresearchcurations.org/blog/2022/2/11/data-management

Data Governance Institute. (2021). Data Governance: The Basic Information. *The Data Governance Institute*. https://datagovernance.com/the-data-governance-basics/adg-data-governance-basics/

Davenport, T. H., Bean, R., & Wang, R. (2024). *CDO Agenda 2024: Navigating Data and Generative AI Frontiers* (pp. 1–24) [Survey]. Amazon Web Services.

Davenport, T. H., Mittal, N., & Irfan, S. (2020). What Separates Analytical Leaders From Laggards? *MIT Sloan Management Review*.

Davenport, T., Smith, T., Guszcza, J., & Stiller, B. (2019). Analytics and AI-driven enterprises thrive in the Age of With. *Deloitte Insights*. https://www2.deloitte.com/us/en/insights/topics/analytics/insight-driven-organization.html

Dubé, L., & Paré, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597–636.

Elia, G., Polimeno, G., Solazzo, G., & Passiante, G. (2020). A multi-dimension framework for value creation through Big Data. *Industrial Marketing Management*, *90*, 617–632.

Gelhaar, J., Bergmann, N., Müller, P., & Dogan, R. (2023). Motives and Incentives for Data Sharing in Industrial Data Ecosystems: An Explorative Single Case Study.

Goodhue, D. L., Quillard, J. A., & Rockart, J. F. (1988). Managing the Data Resource: A Contingency Perspective. *MIS Quarterly*, *12*(3), 373–392.

Gregor, S. (2006). The Nature of Theory in Information Systems. *MIS Quarterly*, *30*(3), 611–642.

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.

Grover, V., & Teng, J. T. C. (1991). How effective is data resource management? Reassessing Strategic Objectives. *Journal of Information Systems Management*, *8*(3), 16–20.

Grugel, J. (2004). *Democratization: A critical introduction* (Nachdr.). Palgrave Macmillan.

Günther, W. A., Rezazade Mehrizi, M. H., Huysman, M., Deken, F., & Feldberg, F. (2022). Resourcing with data: Unpacking the process of creating data-driven value propositions. *The Journal of Strategic Information Systems*, *31*(4), 101744.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049–1064.

Hyun, Y., Kamioka, T., & Hosoya, R. (2020). Improving Agility Using Big Data Analytics: The Role of Democratization Culture. *Pacific Asia Journal of the Association for Information Systems*, *12*(2), 35–63.

Jansen, S. C. (2009). Phantom Conflict: Lippmann, Dewey, and the Fate of the Public in Modern Society. *Communication and Critical/Cultural Studies*, *6*(3), 221–245.

Ketokivi, M., & Mantere, S. (2010). Two strategies for inductive reasoning in organizational research. *Academy of Management Review*, *35*(2), 315–333.

Khatri, V., & Brown, C. V. (2010). Designing Data Governance. *Communication of the ACM*, *53*(1), 148–152.

Kiron, D., Shockley, R., Finch, G., & Haydoch, M. (2012). Analytics: The Widening Divide. *MIT Sloan Management Review*. https://sloanreview.mit.edu/projects/analytics-the-widening-divide/

Labadie, C., Eurich, M., & Legner, C. (2020). Data democratization in practice: Fostering data usage with data catalogs. *Communications of the 20th Symposium of the Association Information and Management (AIM)*, 1541.

Labadie, C., Legner, C., Eurich, M., & Fadler, M. (2020a). FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs. *Proceedings of the 22nd IEEE Conference on Business Informatics (CBI)*, *1*, 201–210.

Labadie, C., Legner, C., Eurich, M., & Fadler, M. (2020b). FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs. *Proceedings of the 22nd IEEE Conference on Business Informatics (CBI)*, *1*, 201–210.

Lee, Y., Madnick, S., Wang, R., Wang, F., & Zhang, H. (2014). A Cubic Framework for the Chief Data Officer (CDO): Succeeding in a World of Big Data. *MIS Quarterly Executive*, *13*(1).

Lefebvre, H., Flourac, G., Krasikov, P., & Legner, C. (2023). Toward Cross-Company Value Generation from Data: Design Principles for Developing and Operating Data Sharing Communities. In A. Gerber & R. Baskerville (Eds.), *Design Science Research for a New Society: Society 5.0* (Vol. 13873, pp. 33–49). Springer Nature Switzerland.

Lefebvre, H., Legner, C., & Teracino, E. A. (2023). 5 Pillars for Democratizing Data at Your Organization. *Harvard Business Review Digital Articles*, 1–9.

Legner, C., Pentek, T., & Otto, B. (2020). Accumulating Design Knowledge with Reference Models: Insights from 12 Years' Research into Data Management. *Journal of the Association for Information Systems*, *21*(3), 735–770.

Lycett, M. (2013). 'Datafication': Making sense of (big) data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386.

Mathiassen, L. (2002). Collaborative Practice Research. *Information Technology & People*, *15*(4), 321–345.

McKeen, J. D., & Smith, H. A. (2007). Developments in Practice XXIV: Information Management: The Nexus of Business and IT. *Communications of the Association for Information Systems*, *19*, 15.

Mehta, S., Dawande, M., Janakiraman, G., & Mookerjee, V. (2021). How to Sell a Data Set? Pricing Policies for Data Monetization. *Information Systems Research*, *32*(4), 1281–1297.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2020). The role of information governance in big data analytics driven innovation. *Information & Management*, *57*(7), 103361.

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. N. (2018). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and E-Business Management*, *16*(3), 547–578.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.

Morgan, D. L., & Krueger, R. A. (1993). When to Use Focus Groups and Why. In D. Morgan, *Successful Focus Groups: Advancing the State of the Art* (pp. 3–19). SAGE Publications, Inc.

Nicolini, D., Pyrko, I., Omidvar, O., & Spanellis, A. (2022). Understanding Communities of Practice: Taking Stock and Moving Forward. *Academy of Management Annals*, *16*(2), 680–718.

Österle, H., & Otto, B. (2010). Consortium Research: A Method for Researcher-Practitioner Collaboration in Design-Oriented IS Research. *Business & Information Systems Engineering*, *2*(5), 283–293.

Otto, B. (2011a). A morphology of the organisation of data governance. Proceedings of the 19th European Conference on Information Systems (ECIS).

Otto, B. (2011b). Data Governance. *Business & Information Systems Engineering*, *3*(4), 241–244.

Otto, B. (2011c). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. *Communications of the Association for Information Systems, 29*.

Parmiggiani, E., Amagyei, N. K., & Kollerud, S. K. S. (2023). Data curation as anticipatory generification in data infrastructure. *European Journal of Information Systems*, 0(0), 1–20.

Parmiggiani, E., & Grisot, M. (2020). Data curation as governance practice. *Scandinavian Journal of Information Systems*, *32*.

Redman, T. C. (2022). Your Data Initiatives Can't Just Be for Data Scientists. *Harvard Business Review*. https://hbr.org/2022/03/your-data-initiatives-cant-just-be-for-data-scientists

Rifaie, M., Alhajj, R., & Ridley, M. (2009). Data governance strategy: A key issue in building Enterprise Data Warehouse. *Proceedings of the 11th International Conference on Information Integration and Web-Based Applications & Services (iiWAS)*, 587–591.

Rutschi, C., Berente, N., & Nwanganga, F. (2023). Data Sensitivity and Domain Specificity in Reuse of Machine Learning Applications. *Information Systems Frontiers*, *26*, 633–640.

Shah, P. W., Stephanie L. Woerner, and Aman M. (2021). Does Your C-Suite Have Enough Digital Smarts? *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/does-your-c-suite-have-enough-digital-smarts/

Short, J. C., Payne, G. T., & Ketchen, D. J. (2008). Research on Organizational Configurations: Past Accomplishments and Future Challenges. *Journal of Management*, *34*(6), 1053–1079.

Someh, I., Wixom, B., Davern, M., & Shanks, G. (2023). Configuring Relationships between Analytics and Business Domain Groups for Knowledge Integration. *Journal of the Association for Information Systems*, *24*(2), 592–618.

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.

Vial, G. (2023). Data governance and digital innovation: A translational account of practitioner issues for IS research. *Information and Organization*, *33*(1), 100450.

Wamba, S. (2022). Impact of artificial intelligence assimilation on firm performance: The mediating effects of organizational agility and customer agility. *International Journal of Information Management*, *67*, 102544.

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, *41*(2), 58–65.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33.

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.

Weber, K., Otto, B., & Österle, H. (2009a). One Size Does Not Fit All—A Contingency Approach to Data Governance. *Journal of Data and Information Quality*, *1*(1), 1–27.

Weber, K., Otto, B., & Österle, H. (2009b). One Size Does Not Fit All-A Contingency Approach to Data Governance. *J. Data and Information Quality*, *1*(1), 1–27.

Wenger, E. (2000). Communities of Practice and Social Learning Systems. *Organization*, *7*(2), 225–246.

Wixom, B. H., & Ross, J. W. (2017). How to Monetize Your Data. *MIT Sloan Management Review*, *58*(3), 9–13.

Xu, D., Someh, I. A., & Indulska, M. (2023). Towards a reconceptualization of data in organizations: A literature review. *Proceedings of the Thirty-First European Conference on Information Systems (ECIS)*.

Yin, R. K. (2003). *Case Study Research: Design and Methods* (4th ed., Vol. 5). SAGE Publications.

Yin, R. K. (2018). Case Study Research and Applications: Design and Methods (6th ed.). SAGE.

Zeng, J., & Glaister, K. W. (2018). Value creation from big data: Looking inside the black box. *Strategic Organization*, *16*(2), 105–140.

Zhu, S., Gupta, A., Paradice, D., & Cegielski, C. (2019). Understanding the Impact of Immersion and Authenticity on Satisfaction Behavior in Learning Analytics Tasks. *Information Systems Frontiers*, *21*(4), 791–814.

8 Appendix

Essay	Summary of changes
Essay 2: Examining the pivotal role of communities of practice for data democratization in enterprises	**Theoretical foundations**: In this version, we strengthened the theoretical foundations by reviewing the nascent conceptualization of data democratization as capability-building and motivating the need for studying the development of data practices in enterprises. To address this gap, we introduce communities of practice (Wenger, 2002; Nicolini et al., 2022) as a novel theoretical lens to study data democratization. **Findings and contribution**: Compared to the conference publication, we extend the theorization of the multi-level landscape practice for data democratization by: • refining the three types of CoPs with detailed vignettes and summarizing their characteristics in a table to map their alignment with the theoretical framework, • discussing the boundary interactions between the different types of CoPs, • positioning our findings against the three CoP lenses by Nicolini (2022) and discussing that such a clear cut is not so obvious for data communities. In doing so, we connect the discourse on data democratization to the emerging data practices literature and enrich the capability view of data democratization with this practice perspective.
Essay 4: From Data as a Resource to Data as an Asset - Data Monetization as a New Frontier for Data Governance	**Problem statement and motivation**: As suggested by the reviewers, we have put more emphasis on problematization and motivation for our research. The revised version argues that data's changing role – and the move from data quality to data monetization – impact on data governance practices. Existing data governance research has mostly focused on isolated practices, with rather operational focus on data lifecycle and data curation, and the related data roles and responsibilities. However, we lack insight into how companies adapt their governance practices to changing business requirements. **Research questions and contributions**: To address the reviewers' comments, we restate the research question (based on RQ1) as follows: How do companies develop their data governance practices to address the changing role of data? We thereby put emphasis on the specific data practices that implement the general data governance mechanisms and their evolution as data's business criticality and strategic importance are increasing. **Methodology and chain of evidence**: We have significantly reworked the entire methodology section and research process to make it more analytical. We added information about the interview protocol. For data analysis, we use a combination of deductive and inductive coding. To ensure that all corresponding practices have been uncovered, we engaged in a bottom-up approach similar to Gioia method where we first created open codes from quotes, then used axial coding to derive categories (corresponding to data governance practices) leveraging notably prior literature. We added Figure 1 in order to make the chain of evidence more transparent and exemplify the codes with quotes from the interviews. **Theoretical contributions:** We clarify our contribution. First, our findings add a strategic perspective to data governance research by uncovering how data governance practices are expanded when companies move from data quality to data monetization. Second, our study provides evidence that data is governed independently from IT and identifies specific structural, procedural and relational data governance practices. While the evolution of data governance mirrors the move from IT as support function to strategic enabler, the type and nature of data governance practices differs significantly from those for IT.
Essay 5: Rethinking data governance: A viable system model	For the dissertation manuscript, we performed an extension of the conference proceeding as following. We considerably extended Section 6 (Federated data governance as recursive system). We added Table 4 which provides an overview of metasystems for each of the five cases' VSM. This way, we provide a visual representation of the replications of data practices in each local data governance hub, thereby demonstrating the existence of a recursion. This also allows us to motivate the choice of ManufCo as vignette for fur further analysis. Indeed, ManufCo's VSM displays a patent example of recursion, showing that most data practices enacted in the five sub-systems are replicated into data domains.

Table 10. Summary of articles extensions relative to the conference proceeding

Essay 1

Data Democratization: Toward a Deeper Understanding

Hippolyte Lefebvre, Christine Legner, Martin Fadler

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

Published in the proceedings of the 42nd International Conference on Information Systems (2021)

Abstract: *Owing to a lack of access and skill, most of the data that companies are creating today is unused, even though it is widely viewed as a strategic asset. To overcome this obstacle, enterprises are establishing data democratization initiatives that can empower employees to use data and extract additional business value from them. However, IS research on data democratization has been scarce and has yet to explain how companies build their data democratization capability. Leveraging a multiple case study in-volving eight companies, we identify five enablers of data democratization: (1) Broad-er data access, (2) Self-service analytics tools, (3) Development of data and analytics skills, (4) Collaboration and knowledge sharing, and (5) Promotion of data value. As academic contribution, our findings clarify the concept of data democratization and shed light on the differences between traditional and born-digital companies. For practitioners, our study delivers actionable insights to tailor their data democratization initiatives.*

Keywords: Data democratization, Data literacy, Data & analytics skills, Data value, Self-service analytics

Table of contents

List of tables

1 Introduction

More and more data is being generated every day—mainly by large enterprise systems and through online social graphs, open data, and digitized devices and applications (Baesens et al., 2016). Although the value-creation opportunities of exploiting these vast data resources are considered significant (Grover et al., 2018), companies are only using a small fraction of the data they collect (IDC 2020). An important reason for this is the lack of data education among non-specialists, who also have limited access to data and analytics tools (Gualtieri et al., 2016; IDC, 2020). Thus, empowering a wider range of employees to access and use data (Zeng & Glaister, 2018)—also called data democratization (DD)—is a priority for many firms. For example, Airbnb launched a universal data discovery and exploration tool, *Dataportal*, in response to the growing data volumes and fragmented system landscape. In addition to enhancing data access, the company also set up an educational program, called *Data University*, to improve data literacy among its employees.

While DD is recognized as an important concept in research and practice, studies on it are scarce, and we still lack a comprehensive definition and proper conceptualization. DD is generally considered a capability (Awasthi & George, 2020; Zeng & Glaister, 2018), but it is unclear what this capability comprises and how it is built. The few studies that investigate DD either rely on a very brief case analysis (Awasthi & George, 2020) or focus on specific aspects. For instance, Labadie et al. (2020) focus on data catalogs as platforms for DD and emphasize the increasing access to data. Others look into DD culture and investigate how it correlates with the adoption of Big Data and Analytics (Hyun et al., 2020). DD is explicitly mentioned by born-digital firms, such as Airbnb, Uber, and Netflix, as an important pillar of their data-driven strategies. In born-digital companies, the use of data is mainstreamed across the workforce and integrated with the organizational culture (T. H. Davenport et al., 2020). By contrast, traditional firms, despite their considerable investments in IT and software capabilities, often struggle with the cultural change that is required to democratize data and generate value from it (Bean, 2021; Kiron et al., 2012). For instance, executives are often not digital-savvy (Shah, 2021), and their perception of what their employees can do and what they know about data often does not reflect the employees' actual capabilities (T. Davenport et al., 2019). This raises questions related to whether and to what extent traditional companies embrace DD.

To address these gaps in research, we ask the following research questions:

RQ1: How do companies build a data democratization capability? RQ2: What are the differences and commonalities between born-digital and traditional companies in building such a capability?

To study DD as a complex contemporary phenomenon within real-life contexts (Benbasat et al., 1987; Yin, 2003), we analyzed case studies from eight companies that have ongoing DD initiatives. Our sample comprises four born-digital and four traditional companies. Based on our insights from within and cross-case analysis, we lay the foundation for an academic conceptualization of DD: First, we define DD as the enterprise's capability to motivate and empower a wider range of employees—not just data experts—to understand, find, access, use, and share data in a secure and compliant way. Second, we identify the five enablers characterizing DD approaches: *Broader data access, Self-service analytics tools, Development of data and analytics skills, Collaboration and knowledge sharing, and Promotion of data value*.

The remainder of the paper is structured as follows. We examine the DD literature and identify the research gap. Then, we detail the case study method and our research process. Next, we introduce the eight case study narratives. Subsequently, we present our findings and analyze the differences and commonalities between traditional and born-digital companies. And finally, we summarize and discuss our findings, and provide an outlook on future research.

2 Background

Our analysis of extant literature in IS and other disciplines reveals that DD has been addressed in a scattered manner. DD is defined differently between studies (see 11) that have different points of interest (e.g., FAIR principles, data sharing, big data analytics) and that target different audiences for DD. However, from these studies, we can extract the key aspects and identify common dimensions concerning how DD is defined (see Table 12). These include greater data access and development of data and analytics skills, as well as collaboration and knowledge sharing. We also detail each of these dimensions in the following sections.

2.1 Data democratization in IS and other disciplines

In IS research, DD has only been addressed in recent years and is primarily associated with broader access to an organization's data. Awasthi and George (2020, p. 1) characterize DD as "*the act of opening organizational data to as many employees as possible, given reasonable limitations on legal confidentiality and security*." They argue that DD can be derived from data philanthropy and open data and further define it as "*intra-organizational open data*." Following that logic, the authors position the sharing of data, skills, and responsibilities as the central thrust of DD. Thus, users should be empowered with the right skills and tools to perform their own analysis (Awasthi and George 2020). Labadie et al. (2020) investigate data catalogs as platforms that support DD in the enterprise context and point out that data access and sharing should be controlled and compliant. They define DD as "*the process of empowering a group of users—not just data experts—to find, access, and use data by removing obstacles to data exploration and sharing in enterprises*" (Labadie et al. 2020, p. 201). Besides highlighting the challenges that broadening the data audience entails, the authors point out that there are several approaches available to companies to address the FAIR (Findable, Accessible, Interoperable, and Reusable) principles (Labadie et al., 2020). A recent study by Hyun et al. (2020) considers DD through the lens of organizational culture and investigates its impact on decision-making based on big data analytics. Generally, both basic (e.g., reporting) and advanced analytics (e.g., predictive models) have a positive effect on organization agility (i.e., how data informs decision-making in a timely manner). However, this effect is moderated by an "*organizational culture that values the willingness to share information and the acceptance of diversity*" (Hyun et al. 2020, p. 42).

Democratization culture associated with advanced analytics positively moderates a firm's agility, whereas the opposite effect is observed when basic analytics is used (Hyun et al., 2020).

Field	Source	Definition	Method	Research scope
IS	Awasthi et George (2020)	*"the act of opening organizational data to as many employees as possible, given reasonable limitations on legal confidentiality and security"*	Four case studies	Data democratization for competitive advantage based on the resource-based view and resource-dependent theories
	Hyun et al. (2020)	*"Organizational culture that values the willingness to share information and the acceptance of diversity"*	Survey of 304 managers	Moderating effect of firm's culture on the impact that big data analytics has on agility
	Labadie et al. (2020)	*"The process of empowering a group of users—not just data experts—to find, access, and use data by removing obstacles to data exploration and sharing in enterprises"*	Three case studies	Data catalogs as platforms to democratize enterprise data to a broader audience
Non-IS	Treuhaft (2006)	*"enabling community actors to access data and to use it to build community capacity to effect social change"*	Two case studies	Impact of Internet and data intermediaries (e.g., GIS) as data suppliers to empower communities
	Bellin et al. (2010)	*"The ability of users to access all data using well-defined and easily used analytic patterns to answer unexpected questions without requiring preauthorization or special additional resources"*	Single case study	Software implementation at a medical clinic facilitating physicians' access to patient data and supporting statistical comparison between patient groups

Table 11. Overview of main definitions associated with data democratization in literature

Interestingly, other disciplines have considered DD earlier than IS. They have also looked beyond the firm's boundaries and considered the needs of external actors. In medical research, DD has been illustrated through the need for healthcare providers to convert data from an electronic medical record (EMR) system into insights to support the treatment of patients or from medication surveillance to achieve better healthcare quality goals (Bellin et al., 2010). Thereby, practitioners can compare their patients' data with deidentified data from other patients and eventually draw better conclusions or diagnostics. Thus, DD not only aim at improving internal processes and making better decisions, but it also benefits patients. Accordingly, DD has been defined as *"the ability of users to access all data using well-defined and easily used analytic patterns to answer unexpected questions without requiring preauthorization or special additional resources"* (Bellin et al. 2010, p. 1366). In urban planning research, DD is discussed as a means to empower low-income urban neighborhoods with the data and tools (e.g., GIS) required to make the right decision for their community (Treuhaft, 2006). Here, the purpose of democratization has been defined as "*enabling community actors to access data and to use it to build community capacity to effect social change*" (Treuhaft 2006, p. 5). In their work, the authors build on an older study by Sawicki and Craig (1996), who characterize DD in three ways: It broadens the locus of computing power and data access, grants non-specialists access to applications, and makes tools easier to use for less-skilled employees.

1.2 Key dimensions of data democratization

1.2.1 Greater data access and tools for non-specialists

All the studies in 11 associate DD with granting a broader, non-specialist audience greater access to data and tools. However, enabling employees to find and use data remains a challenge in practice, and companies need to support them (Awasthi & George, 2020; Labadie et al., 2020).

Field	**IS**			**Non-IS**	
Source	Awasthi et George (2020)	Hyun et al. (2020)	Labadie et al. (2020)	Treuhaft (2006)	Bellin et al. (2010)
D1: Greater access to data and tools granted to non-specialists	Controlled access is given to novices following security and compliance checks	More access to information	Self-service access to data catalog given to more user groups. Focus on FAIR principles	Data resources and applications brought to communities	Access to EMR and reports supported by transparent policies and practices
D2: Development of data and analytics skills	Continuous education of data skills and responsibilities	-	-	Development of community skills to use data	Users trained to treat datasets with care
D3: Collaboration and knowledge sharing	-	Share opinions, information, and knowledge; Communication and active interaction	Data catalogs to foster collaboration between data-related roles	-	Medical tool (EMR) favoring data exchange and access by doctors; use of a wiki
Target audience	Technical specialists and non-specialist employees	All employees.	Nine distinct groups of users	Community actors in an urban area	Clinicians and authorized healthcare administrators

Table 12. Dimensions of data democratization based on prior studies

Data catalogs have been identified as suitable and helpful platforms to democratize data by cataloging and providing an overview of the enterprise's data assets (Labadie et al., 2020). Eventually, democratization can fully happen only if "*the locus of applications must move closer to the citizenry*" (Sawicki and Craig 1996, p. 512). Therefore, users need to be able to use data and applications themselves without restrictions (Labadie et al., 2020; Sawicki & Craig, 1996). However, due to the sensitivity of data, establishing an adequate democratization culture (Hyun et al., 2020), being aware of regulations (e.g., HIPAA), and having proper policies and training are prerequisites to compliant access (Awasthi & George, 2020; Bellin et al., 2010).

1.2.2 Development of data and analytics skills

Existing studies highlight the need to develop data and analytics skills as an important dimension of DD. To relieve data specialists of the most basic tasks, companies must provide their employees with the right set of skills that allow them to manipulate and analyze data in their domain of expertise (Awasthi & George, 2020; Bellin et al., 2010). Before employees can master analytics and derive meaningful insights (Hyun et al., 2020), they need to read and understand the data (Awasthi & George, 2020; Sawicki & Craig, 1996; Treuhaft, 2006). In other words, users must be data-literate. Companies need to support casual users to become more confident when using data and doing data analysis (Treuhaft, 2006). Groups might also be wary of data because of historical, hurtful, and unilateral data-based decisions against them (Sawicki and Craig 1996). Finally, employees need to be skilled at monitoring and controlling the quality of their data (Bellin et al., 2010).

1.2.3 Collaboration and knowledge sharing

Across all studies, collaboration and knowledge sharing are emerging as means of collective empowerment. The diversity of knowledge and opinions and the willingness to share information throughout the company via open communication and active interaction characterize a data democratization culture (Hyun et al., 2020). Hence, a DD culture should be differentiated from a collectivistic culture that focuses exclusively on interactions within groups (Hyun et al., 2020). DD also requires the development of trust between data consumers and data providers (Treuhaft, 2006). Although the definition provided by Bellin et al. (2010) suggests that users can derive insights autonomously and without support, the authors highlight the importance of wikis as collaboration tools to share insights and knowledge with their colleagues. Tools providing search functionality, as well as experimentation environments, also allow users to inform their colleagues promptly of any errors or quality issues in the datasets (Bellin et al., 2010).

2 Methodology

We use a multiple case study research design (Yin, 2003) which is appropriate when few research looked into the topic of interest (Benbasat et al. 1987). Case studies enable us to probe DD in a natural setting (Benbasat et al., 1987; Yin, 2003) and are "*well-suited to capturing the knowledge of practitioners and developing theories from it*" (Benbasat et al. 1987, p. 370). Multiple case studies also improve external validity while supporting analytical generalization (Yin, 2003). A cross-case analysis involving between four and 10 case studies is considered an adequate basis for analytical generalization (Eisenhardt, 1989).

2.2 Case selection and data collection

Based on these guidelines, we selected eight companies with ongoing DD initiatives: four born-digital companies and four traditional companies (Table 13). By selecting a set of companies that vary in terms of industry, date of registration, and digital capabilities, we can analyze the differences between companies' approaches to building DD capabilities. Born-digital firms, which first appeared in the late 1990s, base their core business model on advanced digital capabilities and human capital (Panetta, 2016; Tumbas et al., 2017). We identified Airbnb, Uber, Netflix, and Techtrav (alias) as born-digital companies suitable for our study. These companies are often mentioned by industry experts as reference for DD and report publicly on their DD initiatives. They provide a relative breadth of content made publicly available in various formats (e.g., keynotes, articles by senior employees, company website), which allows us to get a good overview of their DD initiatives and triangulate information, thereby ensuring construct validity (Yin, 2003). In addition to these sources, Techtrav also provided us with additional internal materials during a presentation and a discussion with its Senior Director for Data Science. We provide the complete list of sources used in the Airbnb, Uber, Netflix, and Techtrav cases in Table 16 in the Appendix. For Airbnb, we also included the earlier analysis by Awasthi and George (2020) to enhance our understanding of these cases.

Traditional or "big old" companies seek to develop new digital capabilities to sustain their business models (Pfaff and Hasan 2011; Sebastian et al. 2020). They have recently started DD, but they rarely report publicly about their initiatives. To select suitable cases, we ran focus groups and surveyed data experts from 12 global companies participating in research activities on the topic of data democratization. To collect comparable information, we designed the questionnaire against the dimensions identified both in the literature and at born-digital companies. We queried companies on whether they have a data democratization initiative, what the goals and expected benefits are, which challenges they have encountered, and what the different approaches used and their relevance are. From the nine responses that we received, we removed cases that declared that data democratization was not currently relevant to their company or that they had no ongoing data democratization initiative. This left us with four companies (Sporto, Manuf, Packa, and Automo) that are currently running a DD initiative in various stages. We also conducted interviews with these four companies to confirm our understanding of their initiatives and get additional details (e.g., name of data catalog solution). Case narratives were eventually submitted for review and approval.

Company (year) of registration		**Industry**	**Revenues/ employees**	**Goal of data democratization initiative**
Born-digital companies	Airbnb (2008)	Online accommodation booking platform	\$1B–50B/ ~15,000	Get rid of "*tribal knowledge*" Democratize data beyond technical teams as "*it wouldn't be possible to have a data scientist in every room*" Free up time for data scientists to work on higher-impact projects
	Uber (2009)	Transportation and food delivery service	\$1B–50B/ ~7,000	Use data to inform all decisions in the company by enabling access to infrastructure and advanced tooling in a community-driven environment
	Netflix (1997)	Online video streaming platform	\$1B–50B/ ~20,000	Ensure confidence in data to inform every employee's intuition and strategic decisions
	Techtrav* (1996)	Online travel shopping	\$1B–50B/ ~12,000	Stimulate data demand and the ongoing evaluation of data assets to always have the best insights to share internally and externally
Traditional companies	Sporto* (1949)	Consumer sports goods	\$1B–50B/ ~60,000	Ensure data is shared and accessible to everyone in the company. Foster a data culture focused on data quality
	Manuf* (1915)	Manufacturing, automotive	\$1B–50B/ ~150,000	Increase access to data and develop data analytics capabilities in business to improve operational excellence with assistance from the data enablement strategy released in 2020
	Packa* (1951)	Packaging, food processing	\$1B–50B/ ~25,000	Discover data sources and empower users with the right skills for "*increased productivity, effectiveness and reduced time to insight in many decision-making situations*" Unify all business strategies through the enterprise-wide data strategy released in 2019 to tackle data monetization
	Automo* (1883)	Manufacturing, automotive	\$1B–50B/ ~90,000	Provide enough access to data without having a data "*Wild West*" Develop a data-sharing culture and a data literacy program in line with the 2021 IT and digitalization strategy

Table 13. Overview of case companies

2.3 Within and cross-case analysis

For our case analysis, we followed Eisenhardt (1989) and looked at "*within-group similarities coupled with intergroup differences*." By using pattern-matching, we could identify commonalities as well as differences, depending on the type of company (traditional companies vs. born-digital companies). The first researcher conducted the within-case analysis for all the cases by starting with the four born-digital firms and coded the results according to the analysis framework derived from the literature. The second researcher performed another independent and through review of the cases. To develop our cross-case analysis, we used pattern-matching to identify common and differential enablers used by companies to democratize data. Both researchers refined the final set of key enablers and provided a comparative understanding between born-digital and traditional firms. To gather feedback on our analysis and discuss the specificities of traditional companies, we discussed the final five key enablers with a focus group of 21 experts from 11 multinational companies in February 2021. We received confirmation that the enablers were relevant to their companies' DD initiatives and provide a complete picture of their DD activities.

3 Case analysis

In the following, we introduce the cases and detail the corresponding data democratization initiatives. We first show an overview of the case studies' data against our analysis framework composed of the three dimensions of data democratization previously identified (**Table 14**).

Case	**D1: Greater access to data and tools for non-specialists**		**D2: Development of data and analytics skills**	**D3: Collaboration and knowledge sharing**	**D4: Promotion of data value**	**Target audience**
	D1.1: Broader data access	**D1.2: Self-service analytics**				
Airbnb	*Dataportal* provides content and supports data exploration	Extended analytics toolset upon request	Internal *Data university* program	*Knowledge repository* informs about project progress/success	Promotion of data in teams by data scientists	All, divided into three personas
Uber	*Databook* allows access to data and documentation	Data science workbench for self-experimentation	Tailored (external) training program; Learn by experimenting	*Data science workbench* designed as a collaboration platform	Data scientist in business teams; Data awareness boot camps	All, divided into four personas
Netflix	*Metacat* is a federated data catalog	Universal access to reports and visualizations.	Learning and development team; Learn by experimenting;	Information not publicly available	Being data-driven is in firm's work charter	All employees
Techtrav	Data is accessible to everyone in the data catalog (Alation)	Access to reporting tools for everyone	Internal academy: from basic to advanced data science	Forced collaboration through service interfaces	Employees are encouraged to trigger new data lifecycles	All, divided into four groups
Sporto	Access policies are in place to use the Collibra data catalog	Exploring how to roll out self-service analytics	Optional learnings are externally sourced (e.g., LinkedIn)	Data communities collaborate to publish data on data catalogs and break silos	Awareness session for new joiners; Dedicated channels (e.g.,videos)	All employees, but with restrictions
Manuf	To break silos, a data catalog is under consideration	BI capabilities only for specialists	Training for everyone is coming soon (e.g., PowerBI)	Communities only between technical users for now	Planned, but nothing yet	Mostly technical roles but expanding
Packa	Access to data is restricted by class; Selection of data catalog	Self-service based on the job. Planned expansion	Enterprise Data Literacy project	Communities for experts; Data management committees	Internal podcast	All employees, but with restrictions
Automo	Data access is controlled by policies; Infogix data catalog	Access to analytics tools by role relevance	Data literacy program by role under development	Digital transformation council; Network of data coordinators	Corporate newsletters, but events mainly for specialists	All employees, but with restrictions

Table 14. Overview of case studies

3.1 Airbnb

As the leading online lodging platform, Airbnb has been a data-driven company from its inception and rapidly developed a powerful infrastructure for analyzing the data collected through its only booking platform. Over time, the Hive data warehouse grew to more than 200,000 tables. Although every single employee was expected to be able to use this data, this proved to be challenging due to tribal knowledge (i.e., information only known by certain groups of users). This incentivized the company to democratize data beyond technical teams as "*it wouldn't be possible to have a data scientist in every room*" (Product Lead, Data, and AI/ML). To foster data exploration, discovery, and trust, the company gave all employees at the company access to information so that everyone would be empowered to make data-informed decisions. In particular, data access issues were addressed by using a single source of truth, updated access rights, documented data, and a request process for data and tools (e.g., Excel, BI, Dataportal, Knowledge repository). Dataportal, the company's data catalog, provides certified content to enable trust and guide users in their quest for data exploration by providing contextual information, collaborative interface, and lineage functionality. Airbnb further introduced three personas to define different levels of data and tools proficiencies: the typical (technical) data power user called *Daphne Data*; the less data-literate but more resource-oriented *Manager Mel*; and *Nathan New*, who is typically not comfortable with data or has recently joined the team. A data education program called *Data University* was established to empower employees with the right data skillset to make data-informed decisions. The curriculum is designed around three levels focused on data awareness, data collection and visualization, and data at scale. The rapid success of the program relieved the data scientists of basic technical tasks since ad hoc requests could now be addressed by newly trained employees on their own. As highlighted by an experienced data scientist at Airbnb, "*When business partners can answer their own questions using basic SQL queries and dashboards, it frees up significant time for data scientists to work on higher-impact projects that are crucial for the strategy and direction of their partner teams.*" *Data University* was then revamped as *Data U Intensive* in 2018 to focus training on specific datasets that are relevant to each team. The company's *Knowledge repository* tool informs users in real time about projects' progress while ensuring peer-reviewed contributions and documentation. In this way, the company set out to diffuse insights beyond creators and planned recipients. Data scientists are also allocated to each of these teams, where they act as teachers and advocate for data and data products.

3.2 Uber

Owing to the company's sharp and rapid growth as a transportation platform, Uber has developed quite a complex data landscape that includes processing trillions of Kafka messages per day and storing petabytes of data in HDFS across its different services. Access to data has become too complicated, making collaborations between teams difficult. This was not compatible with Uber's vision for data to inform every decision in the company. Therefore, it was decided to "*design tools accessible for less technical people but encouraging the development of experts skills*" (Product Manager for Experimentation). In response, the company defined four personas as role families and attributed a set of technical capabilities to them. *Reliable Rebecca (RR)* makes decisions based on business insights. While she can create basic visualization roles (e.g., marketing managers, entry-level analysts, and product managers), she also relies on more complex insights provided by data experts. *Monitoring Matt (MM)* focuses on a set of metrics through dashboarding and makes decisions that are usually related to regional operations. He is familiar with more advanced analysis (e.g., advanced SQL). *Analyst Anna (AA)* is typically a data analyst or operations manager who is skilled at using SQL, knows about R and Python, and is responsible for building and customizing the dashboards that inform RR and MM. *Inventor Ivan (II)*, the most technical persona, deals with the data platform's software development and builds the data pipeline, data science, and machine learning, which requires expert programming and statistical knowledge. The company relies on two main in-house tools to support its DD initiative while ensuring cost and compliance requirements: *Databook* and *Data science workbench*. *Databook* allows access to data and documentation not only for engineers and data scientists but also for operation teams. Metadata sourced from different systems (e.g., Hive, internal systems, user input, Cassandra) can further be analyzed through embodied statistics and visualization to help users derive the necessary knowledge. By granting access via browser only and targeting the II, AA, and MM personas, *Data science workbench* aims to "*democratize data science by enabling access to reliable infrastructure and advanced tooling in a community-driven learning environment*" (Product Manager for the *Data science workbench*). This working and productivity-driven tool offers users an environment in which to experiment and collaborate with data with dedicated computer power and an easy setup (e.g., pre-configured Jupyter notebooks) while supporting knowledge sharing through an embodied collaboration-oriented feature. Interactive workspaces support data exploration, preparation, and ad hoc analyses, while advanced dashboards display business metrics and provide insights generated by advanced analytics models. The tool also supports business process automation. Uber partnered with the database modeling tool Vertabelo to develop SQL courses tailored to its own datasets. Since casual users are encouraged to perform their own analysis using the workbench, data scientists have more time to act as coaches, leading to the creation of communities between technical and non-technical users. Also, data scientists act as ambassadors for data in their teams to stimulate demand. Eventually, new product managers in the company are also taking a two-day boot camp to learn how data experimentation can support their work.

3.3 Netflix

Netflix's Vice President for Data Science and Analytics has stated that "*no company in the world is more data-driven than us.*" As displayed on its career website, the company encourages employees to make their decisions independently while sharing as much information as possible. Among its values, Netflix also highlights that data should inform every employee's intuition. In 2017, 700 billion unique user events were processed each day via Kafka and more than 60 petabytes of data were stored in the data warehouse. Its recommendation engine is core to its positioning as a leader in the entertainment business and requires the best data quality possible. The company has further identified data culture and breaking down all barriers to data usage as being critical to its success. Thus, every person in the company has access to data, reports, and visualization (e.g., Tableau) and can perform queries. This is powered by the company's in-house data discovery tool *Metacat*, which acts as a federated data catalog within the company infrastructure. *Metacat* acts as an access layer to data and metadata while ensuring compatibility and interoperability between the different data sources (e.g., RDS, Amazon redshift) and providing an abstraction layer for different data queries (e.g., Hive, Pig, Spark). The tool also makes it possible to track any change to the metadata and enables auto-suggest and auto-complete of SQL queries in the company data visualization platform *Big data portal*. Thus, the tool supports not only broader access to data but also data and analytics teams seeking to increase their project efficiency and impact. To ensure confidence in data, the company identifies three role layers: data producers (e.g., data engineers, analytics engineers, data visualization engineers), primary data consumers (business analysts, research and quantitative scientist, ML scientists), and data consumers who rely on data quality to make decisions (e.g., executives for the more strategic questions, product managers in the content team deciding what to pay for a certain movie based on audience predictions, algorithm engineers to decide what should be the best content and interface to be displayed, and software engineers to optimize content's bandwidth consumption). Technical specialists on the Data Science and Engineering team are also further broken down into personas: *Analyst, Engineer,* and *Visualizer.*

3.4 Techtrav

Techtrav's vision is to enhance travel with data. In particular, the company specializes in matching the right travel options to the right customer and relies on data for its recommender system. As its business model depends on partners (hotel, air, vacation rental) using the platform, the company has to share key data and trends with them so they can make better decisions. With growing data volumes, the company faced challenges storing and managing data due to the lack of standards and the use of different tools to access siloed data across different locations. A strategic initiative was launched to standardize the approaches and make data accessible through one tool using common naming conventions and harmonized governance structure. To this end, the company selected a data catalog provided by Alation. Employees are encouraged to constantly evaluate their business, derive insights, and develop new products by doing experiments or using AB tests, among others. At Techtrav, users are classified into four groups: *Business*, which consumes metrics, dashboards, and reports for product development; *Analyst,* who mainly uses BI tools (e.g., Tableau) to build dashboards, studies trends, and runs tests; *Data science* for data modeling and implementing automated decision-making, who have superior technical skills (e.g., R, Python, Databricks); and *Engineering*, which encompasses data engineering activities and software engineering for application development. Each group can initiate a new data lifecycle by raising a need (e.g., reporting, compliance, analytics), leading to an experimentation phase with different solutions tested and compared before being put into production. As Techtrav strives to always use the best data available, datasets can expire if no value is generated from them, or they are simply not used enough. Regular migrations require the company to establish clear communication between the different groups regarding the data lifecycle. Furthermore, enforcing the use of service interfaces, which also have to be designed for external use, as the only inter-process communication allowed the company to break down the silos and boost collaboration. The continuous training of users is done through an internal academy that is available to all employees and addresses a range of skills from data basics to data science.

3.5 Sporto

Sporto is a global leader in the sports goods industry but has been facing challenges in its distributed data landscape. In turn, these challenges have generated complexity and data quality issues. Sporto seeks to foster a data culture that creates trust in data and ensures data quality from source to usage instead of performing downstream quality fixing on data that is usually siloed. For Sporto, DD is about establishing "*a sustainable link between creators and consumers*". The company relies on collaboration between communities and their willingness to share data on the data catalog (provided by Collibra). In concrete terms, communities—i.e., groups of data producers—can propose that their data be onboarded to the data catalog, which creates a link with existing content and the data consumers. This process required the company to make changes to its data access policies. Under the leadership of the Data Catalog Community Governor, for whom "*Data is not only for geeks*," the DD initiative also fosters data literacy by offering a complete view of data between creators and consumers. In addition, Sporto has started rolling out self-service tools for BI (e.g., MicroStrategy) and aims to launch similar tools for data science to experiment with data in business (e.g., with Alteryx). Data awareness training and introductions to the company's data quality journey are an integral part of the new employee onboarding process, which happens twice a month. Key messages focusing on data quality challenges are embodied in dedicated educational videos produced internally. A wide range of LinkedIn training materials about data and analytics are used and organized around topics such as AI, digital marketing, data science, and digital literacy skills or related soft skills (e.g., critical thinking, strategic planning, problem solving, collaboration). They are further addressed by proficiency levels (beginner, intermediate, advanced). Associated development paths are then suggested to employees by bundling several courses. Collaboration platforms such as Yammer encourage knowledge sharing and collaboration. Promotional videos (e.g., sprint to data quality) are used, as well as dedicated internal marketing channels (e.g., Newsletter), to advocate and spread the initiative message. In particular, the key messages are: Take pride in every data footprint you make, understand the nature of data and be considerate of its creators, and find ways to collaborate and cooperate to resolve data quality issues.

3.6 Manuf

Manuf is an automotive supplier. Process optimization and regulatory management are crucial for it to take advantage of supply chain capabilities (Koch, 2015). Data is distributed across more than 100 ERPs and more than 3,000 applications on multiple platforms. The company's vision to be data-driven is supported by a "*data enablement strategy*" released in 2020, which focuses on operational excellence and digital transformation by empowering business users to make data-informed decisions. As summarized by the Head of Master Data Management, "*it details how to go from business capabilities and business use cases to data and analytics capabilities. We want to share, transform, and reuse data to improve quality, efficiency, and profitability with the highest security standards*." For now, the data is still very siloed (not shared as required), and existing communities (e.g., Master Data community) involve mainly data experts. Training is only relevant for specific applications and only for users actively using those applications. Zoom and SharePoint are used as collaboration platforms between users. A dedicated data enablement team working closely with the business tackle these challenges and is composed of employees with the highest data knowledge. They started by building a capability map to identify gaps and crucial data capabilities to have a successful data-driven business. As a result, a data catalog is being implemented to broaden access and provide contextual information to less data-literate professionals to create a data marketplace. The Data Enablement team is also working on building competencies and data culture starting with data sharing and data virtualization. PowerBI training is also being rolled out to encourage users to perform their own analysis. The self-service initiative is embedded in the transition to SAP S4Hana and SAP Analytics cloud, which should enable further BI capabilities.

3.7 Packa

At Packa, data and analytics initiatives are driven by the enterprise strategy called "*Company 2030*," which directs the overall business transformation and focuses on operational excellence. Data-generated benefits are expected in terms of increased productivity, effectiveness, and faster decision-making. The strategy, released in 2019, integrates and unifies all the specific data strategies available so far (Master Data, BI, Marketing, Engineering). As also highlighted by the Director of Data Management (DDM): "*We have a lot of data. We need to get to the point where we can monetize data related to our customers*." Currently, all data sources are being discovered and analyzed to build a complete overview of the data landscape in a cross-domain data repository. Access to data is granted by data class (e.g., machine data) and documented in access policies. The company will select a data catalog solution and expects to start the setup at the end of 2021. Before rolling out a DD initiative, Packa will establish a data governance model detailing standards, ownership, and responsibilities for the seven data domains. These domains will focus on data quality and particularly on avoiding data duplication, which is a core issue for the development of analytics products as highlighted by the Director of Business Information Management: "*The data science team spends most of its time cleaning the data and doesn't really do big data science projects*." Thus, the company insists on developing awareness (i.e., the "why" of data) and data literacy through the Enterprise Data Literacy project. For now, communities are still quite exclusive to data experts (e.g., Master Data Yammer community with more than 100 users). The roll-out of self-service tools is part of the agenda and supported by the "*Avoid data duplication*" initiative but is currently limited to the extended data and analytics organization. Thus, BI capabilities are fostered in a network that includes more than 500 employees with PowerBI/SAP access rights. New AI capabilities are also explored by using Alteryx. No marketing and communications channels are used to promote data, but the internal podcast, as well as general change management activities, is spreading key messages about the data initiative.

3.8 Automo

For Automo, there are two main reasons to adopt DD: "*Providing enough access to data without having a data Wild West [and] providing enough data and analytics know-how in the required depth for users to be data-driven*" (Data and Analytics Governance Advisor). Data must be ubiquitous, but the increased availability requires governance and security guidelines. A shift in mindset across the organization is required to address the need for a data culture. The company will renew its data and analytics strategy in 2021, which is integrated with the broader IT and Digitalization strategy. While resources to scale up DD are limited, Automo has been implementing a data governance team with corresponding virtual organization. The data governance team is located in the CEO function within the Strategic Digitalization department. Data governance has been rolled out across 47 data domains defined by business objects and functions. It is now establishing several tools and programs, all of which form part of the firm's top data initiatives: the enterprise data catalog provided by Infogix; the enterprise data platform (enterprise data lake with high-performance analytics capabilities in Microsoft Azure Cloud); and a data literacy program. While no holistic training program exists yet, a project called Fit4Digitalization is currently being set up to increase awareness and develop skills in several digitalization topics through a consolidated learning program. In addition to digital methods and software development know-how, the program includes a third area for data-related training ("*data handling & understanding*") and addresses the following topics: data analysis, data science, digital twin, semantic model, and data management. This is further broken down between basic (e.g., introduction to data management), advanced (e.g., consume a data catalog), and experienced knowledge (e.g., create analytics on PowerBI). While every employee needs some foundation in data, most of the training is role-specific (e.g., semantic modeling for IT, data management processes for experienced data citizens). Furthermore, new roles at the company are established with a definition of the expected skillset level. Automo promotes data through newsletters and its intranet, and through an intensive community and stakeholder management (e.g., 200 people in the decentral and global network of data coordinators). Automo also runs special events which address data specialists, managers, and stakeholders. Strategy alignment between the data governance team and business units happens at the digital transformation council twice a year and in many stakeholder meetings.

4 Results

Our cross-case analysis uncovered a total of five enablers or catalysts of DD, which are used in all cases: *Broader data access, Self-service analytics tools, Development of data and analytics skills, Collaboration and knowledge sharing, Promotion of data value*. We find that DD initiatives not only cover the dimensions identified in the literature (Table 12) but also highlight the need for communicating the business value of data and its importance as a strategic asset. This finding led to an extension of the initial analysis framework with the fourth dimension D4 *Promotion of data value* (Table 14). As a result, we defined data democratization as *the enterprise's capability to motivate and empower a wider range of employees—not just data experts—to understand, find, access, use, and share data in a secure and compliant way.* This definition extends the current body of knowledge—in particular, the definition provided by Labadie et al. (2020) in the context of data catalogs—by incorporating motivation (through corporate communication or sharing platforms, for instance) and the necessity to approach data in a secure and compliant way both highlighted in the cases (e.g., Manuf, Automo) and the literature (Awasthi & George, 2020). It also recognizes that data democratization might not involve all employees in the same way but could target a broader audience, user groups, or personas. Our analysis also reveals that the dimension D1 *Greater access to data and tools for non-specialists (D1)* can be broken down into two underlying enablers: *Broader data access (D1.1)* and *Self-service analytics tools (D1.2).* This split highlights that greater access to data should address both the data sources as well as the analytics tools and platforms, that are designed for non-specialist users.

In Table 15, we provide an overview of the enablers with a comparison between traditional and born-digital companies. Companies might bundle several enablers together; however, we address them separately. For instance, born-digital companies might offer self-service tools that empower users, stimulate learning, and foster knowledge sharing. We argue that we need to delimit these concepts as they might be considered separately, especially by traditional firms willing to start their DD journey. Below, we discuss each enabler in detail.

4.1 Broader data access

Broader data access is an expected enabler that emerged from our study. All companies noted the need to provide access to a greater number of employees. To democratize data, users who do not have access to information, which is usually siloed due to a complex data landscape and many applications, need to have a central and documented source of truth. In all our case companies, greater access to data is materialized mainly with data catalogs, which are recognized as privileged platforms to break down silos. While all companies from our sample seek greater access to data for a broader audience, their individual approach might differ depending on the scope and goal of the democratization initiative or the target audience. Traditional companies grant access to data in a controlled manner, usually on request or based on policies. Born-digital companies seek to provide access to data to all employees as all decisions should be informed by data. Moreover, digital firms tend to develop their own in-house data catalogs to better fit their needs and provide enhanced visualization and query features. While these findings are consistent with existing literature, we note a shift from the need to know (e.g., break silos, identify sources, have a complete view between produces and consumers) at traditional firms to the need to share data (e.g., externalization of data, sharing platforms) at born-digital companies.

Enabler	Traditional companies	Born-digital companies
Broader data access	• Controlled approach to data access (donating) • Need to know about available data and sources • Emerging data catalogs	• Universal access • Need to share data internally and externally • In-house data catalogs enhanced with visualizations
Self-service analytics tools	• BI/reporting tools (limited access to relevant roles)	• Analytics experimentation platforms • Enterprise-wide access to reports and visualization
Development of data and analytics skills	• Few (optional) internal training • Focus on data literacy and data contents • Addressed at the role level	• Learning for career development • Internal academy or tailored partnerships using personas • Focus on how to generate insights
Collaboration and knowledge sharing	• Data communities • Boards or committees between data and business	• Collaborate directly on digital tools • Technical specialists sitting in business
Promotion of data value	• Emerging dedicated communication channels • Promote business value from data	• Data in the company values • Stimulate demand for data • Critical thinking and curiosity

Table 15. The five enablers of data democratization

4.2 Self-service analytics tools

All case companies included self-service analytics tools in their DD initiatives. Nevertheless, such tools (e.g., Tableau, MicroStrategy, PowerBI, Alteryx) are still only available to a narrow audience at traditional companies. Although these companies are planning to expand the user base, access rights are mostly given by the job or role level for now. At born-digital firms, the use of self-service BI is part of the required technical skills that most employees should already have. As part of their development in the company, employees can also access in-house experimentation platforms freely or on request to help them develop their skills for further analysis at their own pace. While being consistent with our analysis of existing literature on data democratization, our results provide a more nuanced approach to self-service tools. In particular, our findings align with specialized literature on classic self-service analytics (e.g., BI) by pointing out that access should be controlled and segregated depending on competencies while ensuring enough flexibility to empower users and stimulate their creativity (Alpar & Schulz, 2016; Michalczyk et al., 2020). By contrast, we find that experimentation platforms are empowered with a set of basic to advanced capabilities and simple interfaces to encourage both specialists and non-specialists to try and learn. Overall, our results show that companies might not all require the same level of capabilities to be data- driven. Thus, they need to figure out which capabilities they expect from casual users (Michalczyk et al., 2020).

4.3 Development of data and analytic skills

All case companies in our study either already have training programs or plan to set them up to provide a broader audience with the required skillset to access and use data. Beyond breaking up silos, a key motivation for case companies to democratize data revolves around data citizens' ability to interpret data and perform their own data analysis to support decision-making. However, approaches differ. To empower users with these competencies, born-digital companies build tailored training programs using personas and enhanced peer-to-peer training while traditional firms mainly use external training materials or are just kicking off enterprise data literacy initiatives. Traditional firms are mainly engaging with data awareness and data quality management content, while born-digital firms focus on developing more advanced analytical competencies. This is consistent with prior literature, which argues that, in the context of analytics, companies can set up internal academies (Ghasemaghaei, 2019) and users can be divided into groups based on the data they need, as well as their skills (Alpar and Schulz 2016). Research has demonstrated that employees with the right analytics competencies enable firms to improve their decision-making quality, while employees unable to perform their tasks are likely to postpone or avoid performing the required analyses (Ghasemaghaei, 2019). Without sufficient data literacy and a basic understanding of data and how to use and protect it, big data projects aimed at empowering users and citizens are likely to fail

(D'Ignazio & Bhargava, 2015). Today, data literacy is a central competence required in enterprises and society (Schüller, 2020), involves a continuous learning journey (Sternkopf & Mueller, 2018), and describes the ability to read, work, analyze, and argue with data (Carlson et al., 2011; D'Ignazio & Bhargava, 2015).

4.4 Collaboration and knowledge sharing

Our results show that knowledge sharing is an important enabler of DD and that companies encourage collaboration between data specialists and non-specialists. In the context of decision-making based on data, knowledge sharing refers to the dissemination of the output of a data analysis process (Ghasemaghaei, 2019; Grover et al., 2018). Born-digital companies foster collaboration on digital data discovery or experimentation platforms: They design their data catalogs and analytics platforms with advanced collaboration features so that technical teams can directly support other users or share data with them. They can also use service interfaces so that teams expose their data and functionalities to others—both internally and externally. At traditional companies, collaboration and knowledge sharing are still limited and mostly addressed through dedicated emerging data communities, which are sometimes reserved for specialists or committees aimed at aligning business and data teams. Previous research has shown that knowledge sharing among employees is a key mediator for successful decision-making based on data and analytics (Ghasemaghaei, 2019; Hyun et al., 2020). Collective empowerment of employees with data has been tackled by research on data and analytics governance as the latter has been identified as a cornerstone to get a return on investment into big data technological solutions (Mikalef et al., 2018). In that context, collaboration mechanisms (e.g., communities) or other platforms that exchange ideas and establish shared perceptions (Baijens et al., 2020; Tallon et al., 2013) support the firm in sharing knowledge.

4.5 Promotion of data value

From our analysis, it is clear there is a need for the company to stress and promote the value of data among its employees. All case companies use or plan for dedicated promotion channels in order to raise awareness about the strategic importance of data. Born-digital companies have been using data-driven decisions in their business for a long time. To ensure the quality of and have the best data-enabled products, born-digital companies enforce their data culture and communicate about the compulsory continuous development of their workforce. At born-digital companies, where data is part of the company's DNA, employees are encouraged to trigger new data lifecycles for product development. This is why companies also need to communicate and stimulate demand for data. This is usually done at traditional firms through corporate channels (e.g., events, newsletters, data strategies) and by focusing on communicating key messages about the business value of data—rather than the actual use of data—before even considering the development of any other enablers. Moreover, data scientists can be empowered within product teams to advocate for data and stimulate demand for analytics products at the local level. Our results extend prior DD literature by emphasizing enterprise-wide promotion of data value as additional capability that is needed for the firm to establish DD.

5 Summary and discussion

Our results uncover unique insights about the way companies proceed with their data democratization initiatives. Based on case studies from traditional firms and born-digital companies, we identify the following five enablers for data democratization: *Broader data access, Self-service analytics tools, Development of data and analytics skills, Collaboration and knowledge sharing, and Promotion of data value.* Our results thereby provide an overarching conceptualization of how companies develop and strengthen their data democratization capability. Previous IS literature on data democratization had not provided such a comprehensive overview but rather focused on a subset of enablers. From comparing born-digital to traditional companies, we provide a nuanced view on data democratization and find that both groups use the same enablers, but with different emphasis. Born-digital companies have data as key element in their company values, and their data democratization initiatives address the growing gap between data specialists and non-specialists. Traditional companies need to promote the value of data and break data silos with data democratization. They grant access to data in a controlled and more targeted manner and put more emphasis on data quality and data governance as key milestone in their journey toward data democratization.

In this study, we also introduce a revised definition for data democratization as *the enterprise's capability to motivate and empower a wider range of employees—not just data experts—to understand, find, access, use, and share data in a secure and compliant way.* Compared to previous definitions, this incorporates the motivation for data and the necessity to approach data in a secure and compliant way. It also recognizes that data democratization might not involve all employees in the same way but could target certain user groups or by personas. In this way, our definition joins previous research (Labadie et al., 2020) that argued that when democratizing data, one size does not fill all.

Our findings lay the foundation for an academic conceptualization of data democratization and investigate data democratization as IS social-technical phenomenon (Sarker et al., 2019). They also contribute to the understanding of the data-driven enterprise (Berndtsson et al., 2018) and position data democratization at the beginning of an enterprise's data journey. Our results fit within other research that has recently examined pivotal concepts for big data analytics adoption and, in particular, the capabilities related to organizational design, such as competencies and data-driven culture (Dremel et al., 2020), as well as big data literacy (D'Ignazio & Bhargava, 2015).

Although research on data democratization in the IS discipline is still scattered, we acknowledge that some analogies can be drawn with knowledge management literature that studied the democratization of knowledge. On the Web, platforms like Wikipedia have democratized access to an astonishing amount of information while incentivizing users to create, contribute, and share this knowledge (König, 2013). In enterprises, wiki-based knowledge management systems, along with organizational cultural changes, have been identified as a solution to improve collaboration and trust between management and workers (Hasan & Pfaff, 2007). We argue that data democratization complements the democratization of knowledge, by providing the data and analytics foundation for the creation of knowledge. However, the relationship between data and knowledge democratization is still unexplored and provides interesting opportunities for future research.

Our study comes with certain limitations. As our sample includes only large multinational companies with a certain amount of experience in data management, our results might not apply to smaller companies. For the born-digital companies, we relied on publicly available sources that could lack exhaustivity and completeness compared with other companies in our sample. Although each case company detailed how it developed its data democratization initiative, getting a more precise understanding of the sequencing of implementing the enablers or developing a maturity assessment might be a relevant theme for future research.

Besides having implications for IS theory, our results also inform practitioners how other companies democratize data, which can help them develop their own initiatives.

6 References

Alpar, P., & Schulz, M. (2016). Self-Service Business Intelligence. *Business & Information Systems Engineering*, *58*(2), 151–155.

Awasthi, P., & George, J. J. (2020). A case for Data Democratization. *Proceedings of the 26th Americas Conference on Information Systems (AMCIS)*, *23*.

Baesens, B., Bapna, R., Marsden, J. R., Vanthienen, J., & Zhao, J. L. (2016). Transformational Issues of Big Data and Analytics in Networked Business. *MIS Quarterly*, *40*(4), 807–818.

Baijens, J., Helms, R. W., & Velstra, T. (2020). Towards a Framework for Data Analytics Governance Mechanisms. *In Proceedings of the 28th European Conference on Information Systems (ECIS)*.

Bean, R. (2021). Why Is It So Hard to Become a Data-Driven Company? *Harvard Business Review*. https://cdn.ampproject.org/c/s/hbr.org/amp/2021/02/why-is-it-so-hard-to-become-a-data-driven-company

Bellin, E., Fletcher, D. D., Geberer, N., Islam, S., & Srivastava, N. (2010). Democratizing information creation from health care data for quality improvement, research, and education-the Montefiore Medical Center Experience. *Academic Medicine*, *85*(8), 1362–1368.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369–386.

Berndtsson, M., Forsberg, D., Stein, D., & Svahn, T. (2018). Becoming a data-driven organisation. *Proceedings of the 26th European Conference on Information Systems (ECIS)*, 43.

Carlson, J., Fosmire, M., Miller, C., & Sapp Nelson, M. R. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Portal: Libraries and the Academy*, *11*(2), 629–657.

Davenport, T. H., Mittal, N., & Irfan, S. (2020). What Separates Analytical Leaders From Laggards? *MIT Sloan Management Review*.

Davenport, T., Smith, T., Guszcza, J., & Stiller, B. (2019). Analytics and AI-driven enterprises thrive in the Age of With. *Deloitte Insights*. https://www2.deloitte.com/us/en/insights/topics/analytics/insight-driven-organization.html

D'Ignazio, C., & Bhargava, R. (2015). Approaches to Building Big Data Literacy. *Proceedings of the Bloomberg Data for Good Exchange Conference*, 6.

Dremel, C., Engel, C., Wulf, J., & Mikalef, P. (2020). Looking Beneath the Surface—Concepts and Research Avenues for Big Data Analytics Adoption in IS Research. *Proceedings of the 41st International Conference on Information Systems (ICIS)*, 18.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review*, *14*(4), 532–550. JSTOR.

Ghasemaghaei, M. (2019). Does data analytics use improve firm decision making quality? The role of knowledge sharing and data analytics competency. *Decision Support Systems*, *120*, 14–24.

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.

Gualtieri, M., Yuhanna, N., Kisker, H., Curran, R., Purcell, B., Christakis, S., Warrier, S., & Izzi, M. (2016). *The Forrester Wave: Big Data Hadoop Distributions, Q1 2016*. Forrester Research.

https://www.forrester.com/report/The%20Forrester%20Wave%20Big%20Data%20Hadoop%20Distributions%20Q1%202016/-/E-RES121574

Hasan, H. M., & Pfaff, C. (2007). *Democratising organisational knowledge: The potential of the corporate wiki*. 1–19.

Hyun, Y., Kamioka, T., & Hosoya, R. (2020). Improving Agility Using Big Data Analytics: The Role of Democratization Culture. *Pacific Asia Journal of the Association for Information Systems*, *12*(2), 35–63.

IDC. (2020). *Rethink data* [Consultancy Report]. IDC.

Kiron, D., Shockley, R., Finch, G., & Haydoch, M. (2012). Analytics: The Widening Divide. *MIT Sloan Management Review*. https://sloanreview.mit.edu/projects/analytics-the-widening-divide/

Koch, M. (2015). *Big data and analytics in the automotive industry. Automotive analytics thought piece.* Deloitte.

König, R. (2013). WIKIPEDIA: Between lay participation and elite knowledge representation. *Information, Communication & Society*, *16*(2), 160–177.

Labadie, C., Legner, C., Eurich, M., & Fadler, M. (2020). FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs. *Proceedings of the 22nd IEEE Conference on Business Informatics (CBI)*, *1*, 201–210.

Michalczyk, S., Nadj, M., Azarfar, D., Maedche, A., & Gröger, C. (2020). A State-of-the-Art Overview and Future Research Avenues of Self-Service Business Intelligence and Analytics. *Proceedings of the 28th European Conference on Information Systems (ECIS)*, 19.

Mikalef, P., Krogstie, J., Wetering, R. van de, Pappas, I. O., & Giannakos, M. N. (2018, January 3). Information Governance in the Big Data Era: Aligning Organizational Capabilities. *In Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS).*

Panetta, K. (2016). *10 Management Techniques from Born-Digital Companies*. //www.gartner.com/smarterwithgartner/10-management-techniques-from-born-digital-companies/

Sarker, S., Chatterjee, S., Xiao, X., & Elbanna, A. (2019). The Sociotechnical Axis of Cohesion for the IS Discipline: Its Historical Legacy and its Continued Relevance. *MIS Quarterly*, *43*(3), 695–719.

Sawicki, D. S., & Craig, W. J. (1996). Bridging the Gap for Community Groups. *Journal of the American Planning Association*, *62*(4), 512–523.

Schüller, K. (2020). *Future Skills: A Framework for Data Literacy* (53). https://hochschulforumdigitalisierung.de/sites/default/files/dateien/HFD_AP_Nr_53_Data_Literacy_Framework.pdf

Shah, P. W., Stephanie L. Woerner, and Aman M. (2021). Does Your C-Suite Have Enough Digital Smarts? *MIT Sloan Management Review*. https://sloanreview.mit.edu/article/does-your-c-suite-have-enough-digital-smarts/

Sternkopf, H., & Mueller, R. M. (2018, January 3). Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations. *Proceedings of the 51st Hawaii International Conference on System Sciences (HICSS).*

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.

Treuhaft, S. (2006). The Democratization of Data: How the Internet is Shaping the Work of Data Intermediaries. *UC Berkeley: Institute of Urban and Regional Development.*, 75.

Tumbas, S., Berente, N., & vom Brocke, J. (2017). Born Digital: Growth Trajectories of Entrepreneurial Organizations Spanning Institutional Fields. *Proceedings of the 38th International Conference on Information Systems (ICIS)*, 13.

Yin, R. K. (2003). *Case Study Research: Design and Methods* (4th ed., Vol. 5). SAGE Publications.

Zeng, J., & Glaister, K. W. (2018). Value creation from big data: Looking inside the black box. *Strategic Organization*, *16*(2), 105–140.

7 Appendix

Company	Sources used
Airbnb	https://medium.com/airbnb-engineering/how-airbnb-democratizes-data-science-with-data-university-3eccc71e073a; https://medium.com/airbnb-engineering/democratizing-data-at-airbnb-852d76c51770; https://medium.com/airbnb-engineering/how-airbnb-is-boosting-data-literacy-with-data-u-intensive-training-a6399dd741a2; https://www.slideshare.net/neo4j/graphconnect-europe-2017-democratizing-data-at-airbnb; https://airbnb.io/projects/knowledge-repo/; https://medium.com/airbnb-engineering/scaling-knowledge-at-airbnb-875d73eff091; Awasthi and George (2020)
Uber	https://www.youtube.com/watch?v=q-omIVIuBcw; https://eng.uber.com/databook/; https://www.youtube.com/watch?v=3Jd8uTO-e5A; https://eng.uber.com/pm-bootcamp/; https://uber-academy.vertabelo.com/
Netflix	https://channels.theinnovationenterprise.com/presentations/netflix-big-data-analytics-culture-freedom-responsibility; https://netflixtechblog.com/analytics-at-netflix-who-we-are-and-what-we-do-7d9c08fe6965; https://www.youtube.com/watch?v=nMyuCdqzpZc https://jobs.netflix.com/culture
Techtrav	Company website; Company press releases; Internal presentation by Senior Director – Data science

Table 16. Set of sources used for case companies Airbnb, Uber, Netflix and Techtrav

Essay 2

Unpacking Data Democratization - How Communities Help Organizations Develop and Scale their Data Practices

Hippolyte Lefebvre, Christine Legner

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

Published in the proceedings of the 30th European Conference on Information Systems (2022)

Extended version under review at the European Journal of Information Systems

Abstract: *To leverage data as a strategic asset for innovation, companies must integrate data into all organizational activities, a shift known as data democratization. Nascent literature has conceptualized data democratization as capability-building, with an emphasis on broader data access and data use beyond data specialists. However, this perspective fails to explain how data practices are developed among a diverse group of employees, each bringing a unique mix of technical and business-domain expertise. Through the theoretical lens of communities of practice (CoP), we observe data democratization as a collective empowerment process among members engaged in shared data practices. Based on insights from 17 companies, we analyze 45 CoPs and arrange them on a spectrum from highly situated to generic. We generalize three CoPs for data democratization and depict their interplay into a landscape of practice: (1) CoPs that develop situated data practices; (2) CoPs that develop data practices around tools and methods; and (3) CoPs that develop awareness about data practices. Our research enriches the capability view of data democratization by clarifying the crucial role of practice exchange for capability realization. For practitioners, we offer insights on utilizing CoPs to cultivate data practices with strategic, governance and operational orientations.*

Keywords: Data democratization; Community of practice; Data practices; Data and analytics capabilities; Data literacy; Data value

Table of contents

List of tables

List of tables

1 Introduction

Data is nowadays widely acknowledged as an enabler for value creation in enterprises (Jones, 2019; Mikalef et al., 2020), such as informational value (e.g., decision making support), transactional value (e.g., cost efficiency), transformational value (e.g., business models) or strategic value (e.g., market positioning) (Elia et al., 2020; Günther et al., 2017). To harness this competitive edge, firms must embed data seamlessly into every aspect of their organizational activities, encompassing a wide variety of work practices (Grover et al., 2018; Günther et al., 2022). This evolution marks a transition from data utilization being the exclusive domain of specialists to a broader, more inclusive role, empowering all employees to work with data, a paradigm shift also denoted as data democratization (Zeng & Glaister, 2018; Hyun et al., 2020; van Giffen & Ludwig, 2023).

Despite being a topic of considerable relevance for value creation and innovation with data, research on data democratization remains limited and scattered across different disciplines such as healthcare (Bellin et al., 2010; Y. Wang et al., 2022), urban planning (Sawicki & Craig, 1996; Treuhaft, 2006), and policymaking (Chenarides, 2024). Only a few initial efforts have attempted to conceptualize data democratization from an Information Systems perspective (Awasthi & George, 2020; Lefebvre et al., 2021), defining it as "*the firm's capability to integrate data across the firm and enable a wider range of employees to access and understand data where it is needed at any given time*" (Zeng & Glaister, 2018, p. 20). This discourse has predominantly focused on the various technical and human resources necessary to facilitate data access and use by a larger number of employees (Awasthi & George, 2020; Hyun et al., 2020). In this vein, studies on data democratization have concentrated on identifying the capability's underlying resources such as data accessibility, the provision of self-service analytics platforms, and the implementation of comprehensive training initiatives (Awasthi & George, 2020; Labadie, Eurich, et al., 2020).

However, while significant, the mere focus on resources neglects the fact that data value is realized when employees make sense of the data in a certain working context (Alaimo & Kallinikos, 2022; Stein et al., 2019). More specifically, this perspective fails to explain how data practices are developed among a diverse group of employees, each bringing a unique mix of technical and business-domain expertise (Someh et al., 2023; Galliers et al., 2017). Although research has emphasized on the relevance of practice exchange for data-driven innovation (Aaltonen et al., 2023; Davidson et al., 2023), we still know little about how data democratization materializes into a repertoire of data practices, and how these data practices are cultivated by different roles to achieve organizational goals.

The concept of community of practice (CoP) by Wenger, McDermott and Snyder (2002a) offers a promising theoretical lens to study data democratization as a matter of practice. CoPs are "*groups of people who share a concern or a passion for something they do and learn how to do it better as they interact regularly*" (Wenger-Trayner & Wenger-Trayner, 2005). CoPs establish connections among practitioners, enabling them to collectively learn from one another and integrate new knowledge into their own work practices. Therefore, CoPs can be used a prism through which one can observe the collective empowerment process among members engaged in shared data practices (Wenger, 2000). Hence, we formulate the following research questions:

RQ: *How do CoPs contribute to the development and scaling of data practices in enterprises?*

Our study is embedded into a collaborative research practice, where we partner with 17 companies seeking to democratize their data. Using a multiple embedded case studies approach (Yin, 2018), we identify 45 communities involving data practice exchanges at 12 companies and analyze each community through the lens of the CoP theoretical framework. Our findings unveil a multilevel landscape of practices where three types of CoPs foster data practices at different levels of situatedness —from generic to highly situated— and encounter each other through different boundary interactions to foster data democratization. Type 1 CoPs *develop situated data practices*. Endowed with both technical and business acumen, their members collectively focus on refining data practices to align them with the demands that enable data to be used efficiently and innovatively. Type 2 CoPs *develop data practices around tools and methods*. Members aim to gradually enhance their skills by learning from each other's experiences and collaboratively cultivating new competencies that none of them possess individually. Type 3 CoPs *develop data practices awareness*. By fostering generic enablement, they engage participants across the data spectrum, from experts to beginners, with the goal of enriching the organization's data culture.

This study makes an important contribution to the conceptualization of data democratization, and thereby to the broader discourse on value creation from data. Specifically, our work enriches the emerging IS body of knowledge on data democratization by framing it as a capability brought to life through data practices. We emphasize the pivotal function of CoPs, which leverage organizational resources to foster the development and expansion of data practices. Furthermore, we explain how the CoPs and their interplay help employees navigate and refine the integration of data into work practices. The resulting landscape of practice nurtures both situated and generic data practices among a wide variety of data roles. For practitioners, our findings, augmented with rich examples and vignettes, provide actionable guidance on how to establish pertinent CoPs and thereby scale their data practices.

The remainder of the paper is structured as follows. First, we examine and provide a synthesis of the relevant literature and identify the research gap. Then, we detail the methodology, as well as our research process. Next, we present the landscape of practice of practice for data democratization and explain in detail each of the CoPs identified. Finally, we discuss our findings, draw conclusions, and provide an outlook on future research.

2 Theoretical Background

2.1 Data democratization as paradigm shift for enterprises

The recognition of data as a critical asset for value creation within enterprises is increasingly prevalent in today's business landscape (Mikalef et al., 2020; Grover et al., 2018; Jones, 2019). When effectively harnessed, data has the power to enhance firm's overall performance (Mikalef et al., 2019; Wamba et al., 2017), for instance, by driving innovation, and by enhancing operational efficiencies (Günther et al., 2017; Vial, 2023). Acknowledging that value creation from data is contingent upon making sense of the data in a relevant business context (Aaltonen et al., 2023; Alaimo & Kallinikos, 2022; Günther et al., 2022), firms have engaged recently into a paradigm shift often referred as data democratization, which posits that the responsibility of handling data can no longer be confined exclusively to data specialists (Zeng & Glaister, 2018; Hyun et al., 2020; Lefebvre et al., 2021). Despite their technical proficiency, these professionals may not have the comprehensive business understanding needed to fully acknowledge the value of data (Someh et al., 2023). Data democratization thus implies that eventually more employees should be granted data rights and responsibilities (Lefebvre et al., 2021; Zeng & Glaister, 2018), thereby enriching data with business domain knowledge (Galliers et al., 2017; Someh et al., 2023).

Data democratization has long been studied in research fields such as healthcare (Bellin et al., 2010; Y. Wang et al., 2022), urban planning (Sawicki & Craig, 1996; Treuhaft, 2006), and more recently policymaking (Chenarides, 2024). Yet, IS research on data democratization in still in its early stages. The few IS studies conceptualize data democratization as a capability-building endeavor centered on the elimination of different organizational or technical restrictions in using data, and sharing access to data or tools (Zeng & Glaister, 2018; Hyun et al., 2020; Lefebvre et al., 2021). Typically, they explore ways to facilitate data discovery and data access, for instance with data catalogs or data marketplaces which allow employees to browse data in a library-like experience before "shopping it" and requesting access to it (Labadie, Legner, et al., 2020). They also suggest to onboard more business users on analytics platforms (e.g., Business Intelligence (BI) solutions) so that they can perform basic analysis and create reports by themselves in a user-friendly environment (Zeng & Glaister, 2018). To address the competence gap, companies invest in trainings and other skill-building initiatives that foster a shared understanding of basic data concepts (or data literacy) – that is, making sure people can read, work, analyze, and argue with data (Awasthi & George, 2020). While acknowledging nuances between incumbent firms and digital natives, the literature also identifies intangible resources that organizations must manage to successfully democratize data, such as collaboration and knowledge sharing, and the promotion of data value (Lefebvre et al., 2021).

2.2 Beyond capability-building: data democratization as a matter of practices

While the capability building perspective on data democratization provides an interesting starting point, it overlooks that value creation relies considerably on the successful ability of employees to integrate both data and domain knowledge (Alaimo & Kallinikos, 2022; Günther et al., 2022). Therefore, data democratization transcends mere data access, and implies fostering environments where technical expertise and domain expertise converge toward a shared data goal. As companies integrate data insights into all aspects of business operations, they need to re-evaluate data practices throughout the entire organization (Parmiggiani et al., 2023; Zeng & Glaister, 2018).

Data practices in enterprises are typically performed across three main categories of data roles aligned with data lifecycle and come with their own challenges: Data creator; data custodian; and data consumer (see Table 17). Data creators create, collect or source data as initial input to the organization. Often unaware of the consequences of poorly executed data creation, they either ignore or lack competences and guidelines on how to execute these practices correctly, especially during business process execution (Chua et al., 2022; Hazen et al., 2014). Additional issues are the lacking technical and managerial support needed to improve these practices (H. Zhu et al., 2014). Data custodians manage the principles and infrastructure for processing and storing data. They have expertise in the fields of data storage, data maintenance, data protection, data governance, and data curation. Despite their specific expertise around data storage and data maintenance, data custodians often lack knowledge about the business context to properly perform these practices, for instance when cleaning data (Parmiggiani & Grisot, 2020). Data consumers perform the following data practices: data discovery, data visualization, data sharing, data analytics, data interpretation. This role has been mainly endorsed by data experts due to their data access privileges, and their understanding of a complex tooling environment (e.g., for data discovery and data analysis). Interestingly, the different roles could benefit from further collaboration and alignment to address shared challenges (e.g., difficulty to educate about workplace-specific competences, isolated teams) because the transmission of knowledge across internal firm boundaries improves value creation and decision-making from data (Hyun et al., 2020; Zeng & Glaister, 2018). Therefore, data democratization should develop data practices and stimulate interactions and knowledge sharing between various groups of isolated users or with specialists from shared practices.

Data roles	Data creator	Data custodian	Data consumer
Description	People or other sources who create or source data as initial input to the organization.	People who manage the principles and infrastructure for processing and storing data.	People who consume data in the way that they integrate, aggregate, present and interpret it.
Key data practices	Data collection, Data sourcing, Data creation.	Data storage, Data maintenance, Data protection, Data governance, Data curation.	Data discovery, Data visualization, Data sharing, Data analytics, Data interpretation.
Typical challenges	Lack of awareness and guidelines (e.g., on the quality of data entries) for data creation during business process execution. Lack of support.	Lack of business context when managing data. Scattered data specialists' teams in different organizational areas.	Lack of visibility on available data. Complex toolset. Lack of access to data. Lack of support. Poor trust in data quality.
	Lack of workplace-specific competences. Siloed teams. Poor knowledge sharing between different teams.		
Sources	(H. Zhu et al., 2014) (Hazen et al., 2014) (R. Y. Wang, 1998) (Chua et al., 2022)	(Grover et al., 2018) (Parmiggiani et al., 2023) (Vial, 2023) (Chua et al., 2022)	(Labadie, Legner, et al., 2020) (Mikalef et al., 2020) (Lennerholt et al., 2021) (Chua et al., 2022)

Table 17. Data practices and associated challenges for key data roles

2.3 Communities of practice as relevant theoretical lens to study data democratization

To better understand data democratization as a phenomenon deeply rooted in practice, we propose using the concept of community of practice by Wenger (1998). Communities of practice (CoPs) revolve around the idea that learning and innovation are fostered through collective empowerment around shared practices rather than a unilateral transfer of tacit knowledge between community members (Nicolini et al., 2022; Wenger, 1998). Here, practices are viewed as "*recurrent, materially bounded and situated action engaged in by members of a community*" (Orlikowski, 2002, p. 256). This theoretical lens is particularly suitable for our research context because it underscores the social and collaborative aspects of developing data practices. This complements the array of resources an enterprise must deploy to democratize data effectively.

CoPs have three main intended outcomes: Learning and sharing knowledge (e.g., to develop competences in practice, to circulate knowledge across organization and practice boundaries) ; innovating (e.g., to improve work practices, to generate ideas, to solve problems); and defending and perpetuating interests (e.g., to protect expert's social position, to control change) As highlighted in Figure 6 , three essential characteristics identify and characterize communities of practice (Wenger et al., 2002a). First, members should share a common domain of interest. Second, they benefit from mutually engaging, regularly interreacting, sharing, and learning together in a community setting. Third, members work on developing a shared repertoire of resources, or practice, that they will be able to implement in their working area. Thus, the CoP are sustained if members share common goals for the domain of interest and improve their own practice. The CoP can be distinguished from other group types such as regular work teams, where practice is defined by requirements and tasks assigned – hence, from project teams too. While CoP can be viewed as networks because they connect their members, networks involve more passive participation and focus on sharing rather than collectively developing a shared practice (Brown & Duguid, 2001). Dube et al.(2006) further describe CoPs by their structuring characteristics such as their life span, their creation process, their degree of formalism, their size, their geographic dispersion and their leadership. These characteristics are particularly relevant to examine how CoPs spread across organizational boundaries to reach remote people, and to understand why they emerged to develop or refine practices.

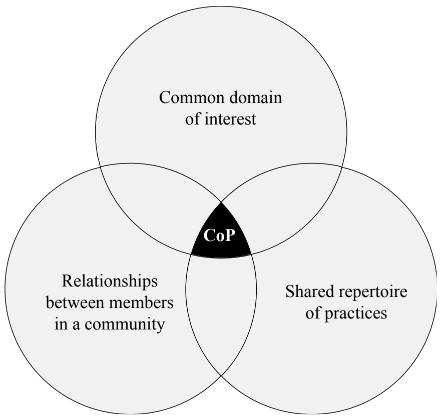


Figure 6. Three characteristics of a Community of Practice

As organizations become more complex, members may need to belong to several CoP that encounter each other, with boundaries delimitating members' inclusion (Nicolini et al., 2022). As a result, CoP can be observed from a landscape perspective showing their interconnections in the form of boundary practices (i.e., specific practices that develop to maintain a connection between two CoPs), boundary encounters (i.e., immersion of a CoP into the action of another CoP) and peripheries (facilitation activities for outsides to join the community) (Carlile, 2004). In such a system, members might explore the applicability of their practice in a "totality of local communities," that is, other CoP, requiring them to cross their own initial boundaries (Carlile, 2004). In this way, they also become knowledgeable about other practices and identify as a member of a larger body of knowledge not limited to their own local practices. Landscapes of practice (LoP), thus, represent the same body of knowledge made out of interconnected practices with clear identities, well-defined boundaries, and knowledgeable members (Wenger, 1998). LoP ignore organizational structures by focusing on practice only while weaving both boundaries and peripheries on the different CoP belonging to it. Such a view is likely to reflect the reality of work and learning, as practitioners must be knowledgeable beyond their local practice to perform their own tasks well (Pyrko et al., 2019). This means that communities belonging to the landscape are "*accountable to one another in terms of their respective practice-based knowing*" (Pyrko et al., 2019, p. 483). Local practices are regularly renegotiated based on practices observed within other communities in the landscape and formatted by more general practices applicable to the different landscapes, i.e., applicable to various bodies of knowledge in the organization (Wenger, 1998). Overall, a difference should be noted between situated practices (i.e., practices fostered within members' own working area) and generic practices that influence both the whole community landscape (Pyrko et al., 2019).

CoP theory offers a suitable lens to study data democratization as a collective empowerment process grounded into practice. By gathering practitioners from diverse teams around a shared domain of interest, data communities stimulate the exchange of practices between business departments that might not be aware of each other's initiatives and would otherwise lead to redundancy, unrealistic expectations, and wrong communication (Baijens et al., 2021; Vial, 2023). In this way, members of shared practice – often geographically separated especially in large organizations – collectively learn from other members in a community setting and apply this knowledge in their own working area. Overall, only anectodical evidence shows that communities as informal structures are an important means to connect across practices (e.g., data-related practices and business-related practices) (Lefebvre et al., 2021; Baijens et al., 2021).

3 Methodology

In line with this motivation, we use multiple embedded case studies (Yin, 2018) to study how data communities emerge or how they are organized to foster data practices. Such method is particularly relevant to capture rich insights about the phenomenon of interest in its natural setting and to capture as much knowledge as possible from practitioners. By connecting our inquiry with CoP theory, we aim to offer up new theoretical insights and explanations for data democratization (Fisher et al., 2021).

3.1 Research context and process

Our study is embedded in a multi-year collaborative practice research program (Mathiassen, 2002) that aims at advancing data and analytics management at large multinational companies. As part of our research program activities, we engaged with 17 companies actively engaged into democratizing their data. These firms have a wealth of experience in data management and operate in a range of industries. Collectively, they form a sample comprising a wide variety of data communities as sub-units for our analysis. This approach was intended to ensure a broad spectrum of insights and perspectives in our study (Miles et al., 2014). In that context, we conducted our study through three subsequent research steps (see Table 18).

3.2 Data collection

The data collection consisted in two steps. First, we aimed to understand each company setting (e.g., data organization, data roles and responsibilities, mechanisms for alignment and collaboration) and inquired about data practices performed across different roles. This step also allowed us to collect a first set of relevant communities. Second, during two focus groups, we focused on collecting and documenting communities dealing with data practices, discussing how they foster data practice exchange.

From September 2020 to October 2021, we conducted a first series of semi-structured interviews with 31 data and analytics management experts from these 17 companies to gain an overview of their data and analytics practices. The interviews were conducted by two researchers via videoconference using MS Teams and were scheduled for 90 minutes (actual range of duration: 66–90 min). The interview guideline was structured around five topics (see Table 24 in appendix): *Business drivers and data strategy*, *Relevant scope of data and analytics*, *Data and analytics organization*, *Data and* *analytics roles*, and *Alignment and collaboration*. To ensure the interviewees had an overview of both global and local practices relating to data and analytics management, we selected managers with data-related leadership and oversights responsibilities and at least three years of experience in the company (see Table 25 in appendix). All interviews were recorded and documented. We also completed the documentation of this primary data by reviewing existing documentation and searching for relevant public sources (e.g., keynotes, press articles, company website). Thus, we could triangulate the information documented during the interview and ensure validity (Yin, 2018). We eventually sent the documentation back to the interviews for review, to confirm our understanding, and to address the remaining open questions.

Research step	**1 – Semi-structured interviews**	**2 – Focus groups**	**3 – Analysis and theoretical integration**
Period	09/2020–10/2021	09/2021–10/2021	10/2021–11/2021
Objective	Understand data practices, and mechanisms for alignment and collaboration between the data organization, business, and IT Map data practices to various roles at each company	Discuss how data communities connect and foster practice exchange between different roles or groups Identify and document typical CoPs for data democratization, and explain how they foster data practices	Analyze the collected communities as subunits of analysis into embedded case studies at each firm Generalize a set of relevant CoPs for data democratization
Main activities	90-minute semi-structured interviews with 31 experts from 17 companies	Focus group with 30 experts from 13 companies Focus group with 16 experts from seven companies	Within-case analysis against CoP theoretical framework Cross-case analysis of 45 CoP from 12 companies to identify generalizable patterns
Outcome	List of data practices at large organizations Mechanisms for alignment and collaboration between data, business, and IT departments	45 communities from 12 companies qualifying as CoPs	Generalized set of typical data CoPs Landscape of practice for data democratization

Table 18. Research process

We then offered companies a follow-up discussion around CoP relevant to data democratization to get additional details about the main CoP observed, as well as interesting new cases of communities that we had identified in the meantime. Focus group 1 happened in September 2021 as part of a workshop with 30 practitioners from 13 companies from various industries and that manifested interest in the topic. Using an

online collaborative whiteboard platform (Miro.com), participants were invited to describe examples of communities of practice that foster data democratization in their company against the three criteria introduced by Wenger *et al.* (2002). Reflecting on the learnings from Research activity 1, we also asked the experts about structuring characteristics from Dube *et al.* (2006) as we expected most of these CoP to be virtual.
For companies that did not participate in the semi-structured interviews and to ensure additional validity, we followed up with them to collect further documentation of their described CoP. For instance, we could obtain documents such as community procedures or examples of community meetings documentation. Necessary clarifications could also be made during focus group 2, which happened in October 2021 in the form of a Web session with 16 practitioners from seven companies. The research team presented the preliminary results of this study, and most participants confirmed they were relevant in the context of data democratization as they support broader data use and learning.

3.3 Data analysis

Overall, we were able to identify 40 data and analytics CoP in the first step and 10 in the second step of our research process. Owing to the participation of several companies in both research activities, five CoP were mentioned twice, leading to a total of 45 CoP from 12 companies identified during the whole research process. One of the challenges we faced relates to practitioners' understanding of communities versus teams or informal networks requiring further clarification with certain companies. As a result, some of the candidate communities could not qualify for the study and were removed. In fact, examples collected from companies H and Q did not qualify as CoP.

We coded each sampled community's intended outcomes as well as the three fundamental characteristics of CoP: *Domain of interest*, *Community*, *Shared practice* (within-case coding). In addition, we coded them against an extended theoretical framework derived from Dube et al. (2006) which lays out additional structural characteristics of CoPs (see Table 19). We focused on the five following essential structuring characteristics: *Size, Community leadership, Life span, Creation process, Degree of formalism*. We purposely ignored characteristics related to maturity and lifecycle as we focus on drafting a landscape of communities that are only emerging in most companies. Thus, we also did not include characteristics dealing with members' enrolment process.

	Characteristic	Description	Coding attributes
Main function (Nicolini et al., 2022)	Intended outcome	Describes CoP's intended outcome i.e., its purpose	Inductively derived
	Mechanism	Underlying mechanisms supporting the intended outcome i.e., how practices are exchanged	Inductively derived
CoP's main characteristics (Wenger et al., 2002a)	Domain of interest	Describes members' common domain of interest	Inductively derived
	Community	Describes member base and their relationships	Inductively derived
	Shared practice	Describe the share practice i.e., the share repertoire of resource and practices	Inductively derived
CoP's additional structuring characteristics (Dube, Bourhis, and Jacob, 2006)	Orientation*	Informs about the general purpose of the community i.e., whether it rather supports a strategic use case or rather focus on operational practices	*Strategic*; *Governance*; *Operational*
	Life span	Informs about the life span of the community	*Temporary*; *Permanent*
	Creation process	Describes how the community was created either be deliberate (e.g., established by management) or emerge among a group of employees	*Spontaneous*; *Intentional*
	Boundary crossing	Informs about the extent of boundary crossing between members of a same community	*Low*; *Medium*; *High*
	Degree of formalism	Relates to different levels of formal recognition and integration in the enterprise structure	*Unrecognized* (i.e., invisible to most employees); *Bootlegged* (i.e., visible only to a specific group); *Legitimized* (i.e., officially sanctioned as valuable entity); *Supported* (i.e., receiving direct resources); *Institutionalized* (i.e., official status and functions)
	Community leadership	Describes how leadership is assigned through formal structures and responsibilities in a governance model. Depending on the needs or expertise, community leaders might also emerge naturally	*Clearly assigned*; *Negotiated based on expertise*
	Size	Informs of the number of people involved in the CoP, hence is a proxy for how data citizens are engaged.	*Small (few people/intimate)*; *Medium*; *Large (e.g., thousand people)*
	Geographic dispersion	Describes how physically dispersed community members are	*Small (e.g., province)*; *Medium*; *Large (e.g., continent)*
	*We extend the work of Dube, Bourhis, and Jacob (2006) by integrating *governance* as a third orientation between operational and strategic levels as introduced by Fadler and Legner (2021)		

Table 19. Theoretical framing of a CoP use for coding

The two researchers performed a joint deductive coding to ensure a common understanding of each sampled community. Although coding was partly achieved with the help of practitioners during the focus groups, the research team ensured a posteriori that coding was accurate (e.g., documentation coherent with the structuring characteristic). Any inconsistencies or unclarity observed by the research team led to a request for clarification directly to the practitioner. As a basis for discussion, we could leverage the contextual understanding collected during the expert study research activity concerning companies' strategies, their current data initiatives, or their organizational structure for data. Moreover, we could collect information about the roles and responsibilities as well as headcounts which helped us to better apprehend the coding.

Next, we conducted the cross-case analysis to understand commonalities and differences across the whole sample, and to generalize our findings. Using pattern-matching (Yin, 2018), we iteratively searched for similarities between codes (e.g., all codes "Master data" for CoP's domain of interest). We then created types or groups of codes and examined cases for shared configurations. We were able to identify (1) a generalizable set of typical communities of practice relevant to the context of data democratization and (2) patterns in the way companies form their CoP across domains and practices, and with various audiences.

4 A landscape of practice for data democratization

Our findings provide evidence of three types of CoPs as key mechanisms to build and scale data practices. Together, and due to their members' interaction, the identified CoPs form a multilevel landscape of practice for data democratization. In the landscape of practice for data democratization (Figure 7), we differentiate the levels of situatedness of each identified CoP along a spectrum. This spectrum spans from situated practices, i.e., practices fostered within members' own working area, to generic practices that influence the whole community landscape (Pyrko et al., 2019).

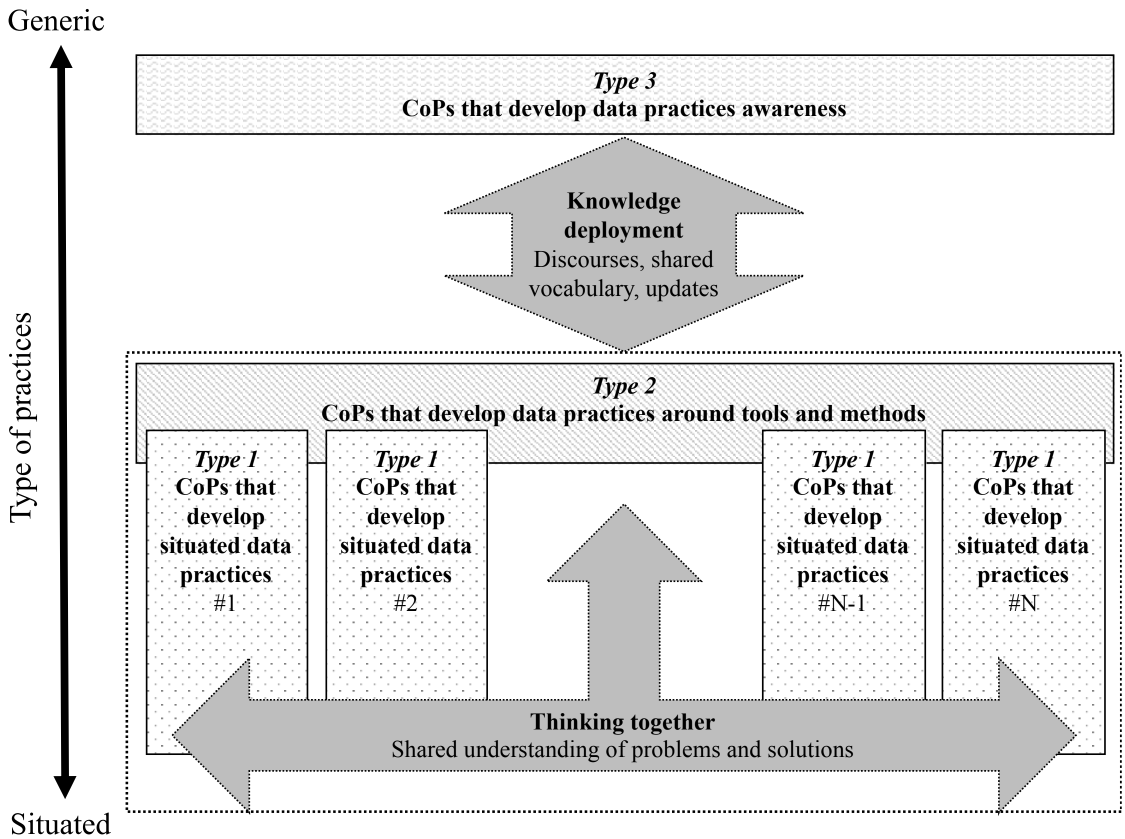


Figure 7. Landscape of practice for data democratization

We unveil three types of CoPs for data democratization. Type 1 CoPs *develop situated data practices*. They highly depend on business context for deriving use cases and managing data lifecycle accordingly. Type 2 CoPs *develop data practices around tools and methods.* They are less situated because they cross organizational boundaries to develop transferable data practices around tools and methods. Type 3 CoPs *develop data practices awareness.* They deploy generic practices by spreading awareness to a large audience that comes from cultural field other than data but in a less specialized manner. Generic practices might then resonate with diverse cultural fields in the enterprise.

Each of the three CoP types for data democratization (Table 20) is characterized by specific intended outcomes, mechanisms, boundary objects, and level of managerial support. It is also associated with typical domains of interests, the community membership and set up, and eventually the shared data practices. In the following, we follow this structure to explain each CoP type and illustrate them with a vignette of our case base.

CoP type	Type 1: Develop situated data practices		Type 2: Develop data practices around tools and methods			Type 3: Develop data practices awareness
Intended outcomes	• Develop new applications for data in business, • Improve data lifecycle in compliance with general methods and tools.		• Jointly develop frameworks (standards, methods, and tools), • Develop data methods and tool competences.			• Develop firm's data culture, • Inform about strategic data context.
Mechanisms	• Joint sense making and ideas sharing, • Incremental improvement of local data practices, • Situated data literacy curriculum.		• Vicarious learning (learn from others' experience with data tools and methods), • Generic and situated data literacy curriculum, • Knowledge encoding into artifacts.			• Legitimate peripheral participation, • Storytelling, • Generic data literacy curriculum
Boundary objects	• Shared data processes • Role in the lifecycle of a data object or data domain		• Artifacts: tools, documents, models			• Discourses • Common language
Managerial support	• Data office advises on membership based on strategic concerns, • Sponsor and domain owners monitor various success metrics.		• Data office provides visibility and support to the community, • Data office coordinates communications between communities and monitor participation.			• Data office makes the community entertaining, • Executive sponsors fund the community's activities.
Domain of interest	Data-driven innovation	Data lifecycle	Data quality and management	Data modelling and architecture	Reporting and analytics	Global data awareness
Community	• Small to Medium-sized (10+ members), and generally focusing on a single function or department, • Composed of data and business leaders e.g., Chief Data Officer, process owners, data architects • Leadership based on sponsorship	• Medium-sized (20+ members), and moderately dispersed across regions, departments, and functions, • Composed of typical data management roles e.g., data stewards, data architects, data domain managers. • Leadership based on ownership	• Medium to large (50+ members), and slightly dispersed across regions, departments, and functions, • Composed of typical data quality and data management roles e.g., data stewards, data managers, data owners • Leadership based on technical expertise	• Medium to large (50+ members), and moderately dispersed across regions, departments, and functions, • Composed of data management experts e.g., data/enterprise architects, data modelers • Leadership based on technical expertise	• Medium-sized (20+ members), and slightly dispersed across regions, departments, and functions, • Composed of typical data quality and data management (e.g., BI experts, data analysts, data scientists) • Leadership based on technical expertise	• Large communities (100+ members) highly dispersed across regions, departments, and functions, • Composed of all employees interested in the data transformation. • Leadership by the data office
Shared data practices	Data innovation, Data opportunities sensing, Data strategy definition, Data consumption, Data productization	Data curation, Data creation, Data productization, Data support, Data sharing, Data documentation	Data curation, Data creation, Data governance	Data modelling, Data architecture, Data storage, Data documentation, Data protection	Data collection, Data sourcing, Data discovery, Data analytics, Data interpretation, Data visualisation	Data protection, Data discovery, Data sharing, Data support
Case companies	A, D, F	B, D, E, I, K, R, T	C, D, F, I, J, K, L, M, N, O, P	A, C, D, E, F, J, K	D, P, T	G, U

Table 20. Overview of the identified types of CoP and their characteristic

4.1 Type 1 CoP: Develop situated data practices

The first type of CoP uncovered develop situated data practices, and thereby embeds data into workplace activities. It brings together various data roles (data creators, data custodians, data consumers) interested in developing innovative applications for data in business, and in improving data lifecycle in compliance with general methods and tools. For this, Type 1 helps with infusing business context into data management (e.g., domain's data access rules) and with using data for strategic purposes (e.g., data productization) through joint sense making and ideas sharing. For instance, we observe temporary CoP popping up at a strategic level when business, IT, and data stakeholders need to discuss and align on the development of a new strategic data and analytics capability in business. In fact, the two underlying domains of interests (data-driven innovation, and data lifecycle) are closely related through a data provider – data consumer relationship. On one hand, the data lifecycle CoP's shared practices revolved around the creation, curation, documentation, productization and sharing of data. On the other hand, data-driven innovation CoP's shared practices typically include consuming data based on a strategic use case. As a result of their contextualization, Type 1 CoPs are of small to medium sizes, not much dispersed, and bring together only members with a direct role in the matter of concern. Their leadership is not really negotiated based on expertise but rather guided by team leaders, sponsors, or hierarchy. The data office may advise on membership and support the sponsors and domain owners into monitoring various success metrics (e.g., value creation estimate, data product consumption).

Vignette 1: Data Catalog community at Company K. Company K is a leading fashion and retail company with more than 55,000 employees worldwide and is implemented in 160 countries. The company releases more than 80,000 articles every year. Each article has more than 400 data points fed by more than 2,000 data creators and generated from 73 products systems. For instance, each sports article reference might be referred to by further attributes (e.g., color, size). In early 2021, the new company strategy released highlights of its ambitious e-commerce goals. Currently, the business model is mainly wholesale and will progressively shift to consumer business. The company was strongly impacted by COVID-19, leading to a surge in its e-commerce sales. This change of business model requires more data, enhanced data quality, and data management. In addition, the wholesale channel also requires further data quality as wholesale partners need it for their e-commerce too (e.g., accuracy of description). Eventually, from an operational perspective, the new strategy targets that 90% of the products should be sustainable. This leads to the collection of new data objects to capture the overall sustainable footprint (e.g., water and electricity used in factories).

Community	Data Catalog community at Company K
Intended outcome	Improve considerably data discovery and use
Mechanisms	Joint sense making and ideas sharing on data-driven innovation
Domain of interest	Trustful sharing of data products
Community	Data domain owners as providers; Analytics teams as consumer.
Shared practices	Data innovation, Data opportunities sensing, Data productization, Data sharing, Data documentation
Orientation	Operational
Life span	As the company uses the data mesh concept, communities alike are key to establish relationship between data domains hence the community is likely to be sustained
Creation process	Established naturally by the central data team alongside the release of the data catalog
Boundary interactions	Medium, two groups from different departments: data providers with ownership of the data and consumers who find and access the data
Degree of formalism	Institutionalized (i.e., officially recognized as highly valuable)
Community leadership	Appointed community governor from the central data management team
Geographic dispersion	Analytics teams are situated within the central data team
Size	Rather low

Table 21. Overview of Data Catalog community at Company K

To enable data-driven decision-making from their huge amount of data, Company K seeks to establish an enterprise data culture of awareness, credibility, and trust, combined with a strong data quality improvement initiative. A growing central data team (120 FTEs) handles data management, data quality, data platform, BI, and analytics. In business, dedicated data owners manage master data (and few transactional data) in data domains. A decentralized analytics team in the sales department focuses more on the fast-moving analytics products (e.g., product recommendation on e-commerce). The company seeks to foster data sharing and collaboration between the different data and analytics teams by offering a 360° view of data. In short, data democratization at company K means establishing a sustainable link between data creators and data consumers in their data mesh. These two groups then form a temporary community of practice around specific data products. Their exchange is organized through the data catalog (provided by Collibra) community and facilitated by a dedicated formal role in the central data management team: a Data Catalog community governor, for whom "*data is not only for geeks.*" More precisely, the latter orchestrates the onboarding of the required data onto the data catalog.

In this way, the company can address data siloes generated by product systems and drive data quality necessary for its ambitious analytics use cases (e.g., in-season product forecasting).

4.2 Type 2 CoP: Develop data practices around tools and methods

The second type of CoP uncovered develops data practices around tools and methods. It brings together various data roles (data creators, data custodians, data consumers) interested in the joint development of frameworks and in developing data methods and tool competences. They seek to grow their technical expertise by learning from their peers (vicarious learning) or by sharing learning outcomes in a data literacy curriculum. Members' pooled knowledge also supports the development of shared data artifacts (e.g., data quality rules, dashboards, data models). More specifically, Type 2 CoPs self-organize around the following domains of interests: data quality and management, reporting and analytics, and data modeling and architecture. Although members are often data specialists (e.g., data architects, data managers) who ensure community leadership based on their expertise, Type 2 CoPs are open to less-experienced profiles who want to benefit from best practices (e.g., business employees who have been assigned data ownership responsibilities). Unlike Type 1 CoPs which may interact with each other, Type 2 CoPs are more independent and more specialized. This is why the data office generally helps with providing visibility and support to the different Type 2 CoPs (including participation monitoring), and coordinates communications between communities. Due to Type 2 CoPs' geographical dispersion across several business functions and locations, preferred communicated channels are e-mails, conference calls, or collaboration platforms such as Yammer. However, for more technical domains (e.g., data modeling and architecture), Type 2 CoPs' may consist of a smaller number of members closer to the data office.

Vignette 2: Data Quality community at Company P. Company P is a global leader in technologies, network, and telecommunications solutions and has more than 200,000 employees worldwide and in more than 170 countries. To scale up data use for the entire organization and foster a data-driven culture, the company has set up a Chief Data Office with about 50 full-time employees (FTEs) reporting to the CIO (board level). The Chief Data Office's responsibilities include deploying the governance framework, rolling out procedures, and maintaining the information architecture across 100+ business objects. Its oversight spans over 10 function areas, for instance, finance and logistics, that have their own data management and data quality teams. Depending on the functions, such teams might sometimes be bigger than the central team. Thus, data is widely democratized in functions.

Community	Data Quality Community at Company P
Intended outcome	Develop methods and standards to improve data quality scoring
Mechanisms	Best practices exchange with equivalent stakeholders from other functions on data and metadata quality standards, and data quality tooling
Domain of interest	Data quality
Community	Data stewards, data experts, business analysts
Shared practice	Data curation, Data creation, Data governance
Orientation	Governance
Life span	Sustained, at least till DQ index objectives are achieved
Creation process	Purposely created by the Chief Data Office
Boundary crossing	Medium, expert from 10 functions
Degree of formalism	Institutionalized (i.e., officially recognized as highly valuable)
Community leadership	Chief Data Office
Geographic dispersion	Global, over 10 functions (e.g., finance, logistics)
Size	Total of 1000+ employees

Table 22. Overview of Data Quality community at Company P

To monitor data quality across functions, a corporate data quality index across 22 data domains (built along functions and divisions) is measured twice a year, signed off by the CFO and CIO, and presented to the board members. A minimum score of 60% is currently set as the threshold for requirements completion versus a list of data quality rules. The assessment is performed against 1200 data objects provided with requirements and ownership on both the data and its metadata. To enable this global effort for better data quality, a large community has been established by the Chief Data Office as essential to foster cross-domain alignment about data quality. This community is institutionalized (i.e., officially recognized as highly valuable) and comprises more than 1,000 data stewards, data experts, and business analysts. The members are geographically dispersed all over the world but gather monthly to exchange and learn how data quality can be improved in source systems. They also represent their local practice in each of the data domains and seek to increase their domain quality index. By learning about practices in the most successful domains and presenting their challenges to others, members expect to improve their own domain data quality. Community members meet every month, but further exchanges happen through a group chat and a dedicated wiki supported by a dashboard to monitor quality improvements. The latter is built upon an unified enterprise-level metadata center based on Informatica which supports the implementation of data quality rules and automatic generation of measurement reports.

4.3 Type 3 CoP: Develop data practices awareness

Type 3 CoPs show the most generic data practices. Their main expected outcomes are the creation of a data culture centered on a corporate understanding of firm's strategic data context. This type of CoP is then special in the sense that it involves less situated practice exchange but aims to develop data practices awareness by disseminating updates (e.g., FAQs), best practices, stories, and training (i.e., general awareness about data) among, for instance during large corporate events. They are piloted by a core team usually located in the central data or analytics organization and seek to reach any employee interested in data to create a data culture down to the operational level. Hence, one of the challenges of these types of CoP is to sustain engagement of a large (typically 100+), culturally heterogeneous, and geographically dispersed member base with limited accountability on their own data practice. As lurkers or guests, most of the participants remain in the CoP's periphery and might never become full members. However, if engagement is sustained, members can integrate these mostly generic practices into their local practices. For instance, Type 3 communities might be highly relevant to foster shared language and vocabulary concerning data. As a result, the core team invests a lot of its time and resources, and dedicated FTEs might be assigned for continuous community entertainment.

Vignette 3: Data Mobilization at Company U. Company U is a leading provider of global logistics solutions and has more than 70,000 employees across 2,100 locations worldwide that seek to achieve financial excellence by 2024. In the company's business strategic plan, three specific goals are targeted: maximize cost and performance transparency; enable instant data availability; and drive digitalization across all finance operations. Harmonization of master data is a key initiative launched in 2021 to support this journey toward excellence. For the CFO who kicked off the global data mobilization, "Master data management (MDM) is not a head office project. It involves all of us." There must be a culture of completeness and correctness (i.e., data quality), especially at the operational level, to enable the digitalization of the finance function.

Community	Data mobilization at Company U
Intended outcome	Increase of company-wide data awareness and data management best practices to support strategic plan
Mechanisms	Exchange of basic best practices for data management, awareness of trainings, news, and updates. Presentation of success stories and updates on projects. "Emotionalize data" to raise data culture and trigger ownership. Inform about support channels
Domain of interest	Data culture and corporate data management
Community	The core data management team (8 FTEs) and anyone in the company willing to attend.
Shared practice	Data protection, Data discovery, Data sharing, Data support
Orientation	Operational
Life span	The community is recent (2021) and is expected to be sustained as the audience keeps growing
Creation process	Orchestrated by the central data team with support from the CFO
Boundary interactions	High, invitees are considered as attendees or lurkers from any departments while the core group is composed of the central data management team
Degree of formalism	Institutionalized (i.e., officially recognized as highly valuable). Funds are allocated to event management, FTEs, and dedicated channels.
Community leadership	The core data team leads the community with dedicated staff for active management of community channels and events preparation
Geographic dispersion	Wide, as it is a global initiative
Size	400+ members

Table 23. Overview of Data Mobilization at Company U

To support this vision, the global data management teams, which consist of eight FTEs and are already well-connected to their regional counterparts, developed the data mobilization to engage with the whole company about master data management. These sessions, which take place twice a month, welcome any data citizen interested in learning more about master data (and particularly business partner data). The global team then provides awareness sessions, project updates, guests presentations, and newly available training sessions. Exchanges are fostered through quizzes and prize winners, Q&A sessions, and feedback/requests collection supported by a dedicated community platform called "MDM knowledge café". As the leader of the community, the global MDM team focuses on sustaining interest by creating excitement about the topic and not by just addressing operational matters. "You have to be attractive and keep them entertained," says the Head of MDM. For instance, themes are associated with each session (e.g., Halloween sessions discussed horror topics, including cyber challenges). Between sessions, continuous engagement is organized through the company collaboration platform where the global team collects members' new topics of interest that could inform upcoming sessions. In less than a year, attendance grew from about 100 attendees at the first session to more than 400 in October 2021, with 70% of participants coming from non-MDM functions.

4.4 Boundary interactions

As show in Figure 7, our results show that the CoP landscape relies on epistemic boundaries of different "flexibility." Thus, we observe practitioners crossing boundaries and becoming knowledgeable about other practices. We can then provide each CoP type with an interpretation of their boundary interactions (Wenger, 1998). Thanks to our cases, we could notably extract concrete examples of boundaries surrounding membership or not into the CoP. For Type 1 CoP, boundary objects are rather linked to business processes, data domains (e.g., sales data), and specific data objects (e.g., employee data). Due to their close data practices exchanges, communities belonging to Type 1 mainly around *boundary practices*, for instance through the data productization practice which formalizes the data consumer-provider relationship around an innovative use case. For Type 2 CoP, boundary objects are mainly artifacts such as tools, frameworks, or models. Type 2 CoP interact with Type 1 CoP by transferring data practices around tools and into local data practices, thus relying on *boundary encounters*. Consequently, we further label knowledge exchanges, which serve to generate a shared understanding of the problem and solutions between Type 1 and Type 2 CoP, as *thinking together.* Type 2 CoPs can then be seen as orthogonal to Type 1 CoPs by bringing their expertise into situated data practices. As a result, some data roles might be members of different CoP types. As depicted in the landscape, these interactions lead to a "grid view" of Type 1 CoPs versus a panel of technical experts from Type 2 CoPs. Owing to the specific set of activities conducted in Type 3 CoP, most of its participants are guests who do not exchange about their practice as much as in the other two types. Moreover, we find that Type 3 CoP focus on deploying knowledge on how to work with data. Type 3 CoP develop generic practices which "*can take the form of exchanging facts or stories at the various layers of CoP periphery*" (Pyrko et al., 2019, p. 489). Hence, the interactions of Type 3 CoP with the rest of the landscape happen through *peripheries* and enable legitimate peripheral participation (Lave & Wenger, 1991), mainly with discourses as boundary objects.

5 Discussion

The depicted landscape of practice for data democratization addresses all three CoP's main functions mentioned in the literature (Nicolini et al., 2022) (learning and sharing knowledge, innovating, and defending and perpetuating interests), yet to various extents.

On one hand, we find that the primary objective of the different CoP identified is about *Learning and Sharing Knowledge* across different organizational and practice boundaries, notably through legitimate peripheral participation which illustrates the evolution of such competence building through the prism of community membership. Novice members progressively grow towards mastery by regularly engaging, interaction and collaborating with a core group of experts willing to listen and share their knowledge and experience (Lave & Wenger, 1991). Situated learning theory explains that learning is more efficient in a collaborative and authentic context, i.e., where domains of interest shared by participants are based on real-life applications. Thereby, learning is rather a social phenomenon which contrast with individual abstract learnings detached from its applications and associated experts. In other words, situated learning is heavily relying on apprenticeship which hopefully will lead them from novice to mastery in the subject matter (Gherardi, 2000; Nicolini et al., 2022). This aligns with studies that demonstrated the significance of practice-based learning into data literacy curricula thereby ensuring the right level of cognition applied to data practices across different roles (Lefebvre & Legner, 2024; Micheli et al., 2020; S. Zhu et al., 2019).

On the other hand, we observe that the landscape cannot be fully explained with just a single lens and that the CoP also leverage underlying mechanisms from *Innovating* and *Defending & Perpetuating Interests* lenses. For instance, Type 1 CoPs might also emerge to stimulate innovative use cases with the goal to create competitive advantage and strategic business impact e.g., to support the development and consumption of new data and analytics products. Hence, in this case the CoP is less about learning through practice exchange, but rather about improving existing practices through strategic initiatives or brainstorm about new ideas to create business value from data and analytics. Hence, the CoP brings strategic business roles into data discussions and facilitates the collection of requirements. If needed, technical assistance and experience are provided through Type 2 CoP which traditionally embed legacy data experts thought the latter might be also benefit of the business context and insights offered through Type 1 CoP. Conversely, Type 3 are by nature quite generalist, have many members, and aim at a large scope of practice exchanges. Activities include raising awareness, discussing best practices, or collecting ideas. We further observed that Type 3 CoP are organized and sustained by a governing group of individuals, hence adding nuance to existing research stating that "*generic practices do not have a clear core group*" (Pyrko et al., 2019, p. 489). Therefore, the core group – oftentimes composed of members from the central data management team – exercises control over the democratization of data by carefully and continuously redefining the scope of tolerated data practices. Hence, this Type 3 CoP can also be observed through mechanisms from the *Defending & Perpetuating* lens. Interestingly, Type 3 CoP are only observed at Companies G and U which are among the least advanced companies in terms of data and analytics management practices in our case base. This might indicate that these communities produce the maximum effect at the beginning of the data democratization journey, when data value needs to be promoted among a very large audience. In fact, no Type 1 and 2 communities were observed at Company G and U which might indicate a sequencing of communities' roll-out by management (top-down implementation and legitimization).

6 Contribution and implications

6.1 Contribution

Our results enrich the prevailing capability perspective on data democratisation (Awasthi & George, 2020; Lefebvre et al., 2021; Zeng & Glaister, 2018). More specifically, our contribution lies in the clarification of data practices as crucial for the conversion of organisational resources into the capability. Observing data democratisation via the concept of communities of practice provides a theory-informed and practically grounded understanding of the informal structures that connect employees working with data across teams, and around a shared data practice. We unveil three types of communities relevant for data democratisation with their own characteristics and their specific contribution to data practices. Together, these CoPs co-exist and interact in a landscape of practice through different types of boundary interactions. We also shed light on the type of practice (on a spectrum from situated to generic) promoted by each type of CoP in the landscape and highlight the importance of certain key data roles for data democratisation (e.g., data steward, data owner, data architect).

6.2 Implications for research

This study has implications for the larger discourse on value creation from data. Although companies can improve value creation through strategic resource allocation, our results show that the majority of data practices grow and proliferate through the exchange of data practices, often beyond direct managerial influence. This underscores the significant role of data experts as mentors, guiding the development and expansion of data practices. However, based on the structuring characteristics of the three types of CoP, we acknowledge that CoPs may need to be legitimised as the network grows. Although CoP might start as spontaneous, their growth could lead to a need for further coordination and support from a management authority. This finding aligns with research on CoP that mentioned the struggle faced by growing communities to remain self-managed (Barrett et al., 2004; McDermott & O'Dell, 2001; Swan et al., 2002).

Our findings also show that data democratisation is stimulated through practice exchanges with strategic, governance, and operational orientations. Hence, data practices are pervasive and extends beyond the mere operational scope, leading to a dynamic reshuffling of firm's informal structures to enable the growing network of data users (Abraham et al., 2019; Peppard, 2018). Therefore, our findings also inform data governance research. We observe that CoP related to data governance act as relays of strategic intents, and diffuse the standards and definitions required to operate data lifecycle i.e., to govern data. This implies that data governance principles, which are typically communicated as policies and guidelines, can be diffused through practice exchange between highly situated CoPs and those centred on methods. This approach effectively counters critiques that data governance is too abstract or removed from day-to-day operations (Alaimo & Kallinikos, 2022; Alhassan et al., 2016; Benfeldt et al., 2020). This interconnectedness ensures that data governance is not just a set of guidelines, but a "*dynamic element that is implemented and should evolve in conjunction with strategy and operations*" (Vial, 2023, p. 9).

6.3 Implication for practitioners

For practitioners, we provide valuable insights on which data communities could be relevant to support data democratisation in their own company. Our vignettes illustrate real-life scenarios on which they can gain insights. Although we do not provide guidelines for sequencing the roll-out of the different CoP, firms may use the different structuring characteristics to judge their applicability (e.g., by assessing if key data and analytics roles are defined). Hence, when trying to democratise their data, firms should reflect on their data governance model as enabler of the landscape while providing a frame for the emerging data practices.

6.4 Limitations

As we engaged only with multinational firms with large headcounts and global operations, our analysis cannot be generalised to smaller structures. In fact, small and medium-sized enterprises might have more straightforward ways of exchanging, especially if they are all located on the same premises.

Moreover, most of this research was done during Covid-19 period including lockdowns, hence during a time when companies were rethinking the way employees work and collaborate (Tim et al., 2023; Weinert & Weitzel, 2023). We could observe an acceleration in the development of CoPs to account for new ways of working that engages geographically dispersed people, among others. Researchers could then follow-up on our study and investigate to what the extent the landscape of practice has been impacted, especially when CoPs for data democratisation go virtual. For instance, while smaller communities are more likely to facilitate shared understanding and practice exchange, larger communities with hundreds to thousands of members in different departments and time zones may lead to a dilution of members' contribution.

Since our study does not encompass a thorough examination of CoPs' lifecycle, delving into their development stages presents interesting avenues for future research (e.g., through a longitudinal case study). Specifically, investigating the role of management throughout each phase of a CoP expansion presents a valuable research opportunity. This inquiry could reveal key insights into how management actions can either facilitate or hinder the growth and effectiveness of CoPs, ultimately impacting organisational knowledge sharing and innovation.

7 References

Aaltonen, A., Alaimo, C., Parmiggiani, E., Stelmaszak, M., Jarvenpaa, S., Kallinikos, J., & Monteiro, E. (2023). What is Missing from Research on Data in Information Systems? Insights from the Inaugural Workshop on Data Research. *Communications of the Association for Information Systems*, *53*(1), 4–19.

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.

Alaimo, C., & Kallinikos, J. (2022). Organizations Decentered: Data Objects, Technology and Knowledge. *Organization Science*, *33*(1), 19–37.

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems*, *25*(sup1), 64–75.

Awasthi, P., & George, J. J. (2020). A Case for Data Democratization. *Proceedings of the 26th Americas Conference on Information Systems (AMCIS)*, 23.

Baijens, J., Huygh, T., & Helms, R. (2021). Establishing and theorising data analytics governance: A descriptive framework and a VSM-based view. *Journal of Business Analytics*, 1–22.

Barrett, M., Cappleman, S., Shoib, G., & Walsham, G. (2004). Learning in Knowledge Communities: Managing Technology and Context. *European Management Journal*, *22*(1), 1–11.

Bellin, E., Fletcher, D. D., Geberer, N., Islam, S., & Srivastava, N. (2010). Democratizing information creation from health care data for quality improvement, research, and education-the Montefiore Medical Center Experience. *Academic Medicine*, *85*(8), 1362–1368.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, *22*(2), 299–313.

Brown, J. S., & Duguid, P. (2001). Knowledge and Organization: A Social-Practice Perspective. *Organization Science*, *12*(2), 198–213.

Carlile, P. R. (2004). Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries. *Organization Science*, *15*(5), 555–568.

Chenarides, L. (2024) 'Creating Engagements: Bringing the User Into Data Democratization', *Harvard Data Science Review* [Preprint], (Special Issue 4).

Chua, C., Indulska, M., Lukyanenko, R., Maass, W., & Storey, V. C. (2022, February 14). *MISQ Research Curation on Data Management*. MIS Quarterly. https://www.misqresearchcurations.org/blog/2022/2/11/data-management.

Clarke, R., & Davison, R. M. (2020). Research Perspectives: Through Whose Eyes? The Critical Concept of Researcher Perspective. *Journal of the Association for Information Systems*, *21*(2), 483–504.

Davidson, E. *et al.* (2023) 'Future directions for scholarship on data governance, digital innovation, and grand challenges', *Information and Organization*, 33(1), p. 100454.

Dube, L., Bourhis, A., & Jacob, R. (2006). Towards a Typology of Virtual Communities of Practice. *Interdisciplinary Journal of Information, Knowledge, and Management*, *1*, 069–093.

Elia, G., Polimeno, G., Solazzo, G., & Passiante, G. (2020). A multi-dimension framework for value creation through Big Data. *Industrial Marketing Management*, *90*, 617–632.

Fadler, M., & Legner, C. (2021). Toward big data and analytics governance: Redefining structural governance mechanisms. *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Fisher, G., Mayer, K. and Morris, S. (2021) 'From the Editors—Phenomenon-Based Theorizing', *Academy of Management Review*, 46(4), pp. 631–639.

Galliers, R.D. *et al.* (2017) 'Datification and its human, organizational and societal effects: The strategic opportunities and challenges of algorithmic decision-making', *The Journal of Strategic Information Systems*, 26(3), pp. 185–190.

Gherardi, S. (2000). Practice-based Theorizing on Learning and Knowing in Organizations. *Organization*, *7*(2), 211–223.

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.

Günther, W.A. *et al.* (2022) 'Resourcing with data: Unpacking the process of creating data-driven value propositions', *The Journal of Strategic Information Systems*, 31(4), p. 101744.

Günther, W.A. *et al.* (2017) 'Debating big data: A literature review on realizing value from big data', *The Journal of Strategic Information Systems*, 26(3), pp. 191–209.

Hazen, B.T. *et al.* (2014) 'Data quality for data science, predictive analytics, and big data in supply chain management: An introduction to the problem and suggestions for research and applications', *International Journal of Production Economics*, 154, pp. 72–80.

Hyun, Y., Kamioka, T., & Hosoya, R. (2020). Improving Agility Using Big Data Analytics: The Role of Democratization Culture. *Pacific Asia Journal of the Association for Information Systems*, *12*(2), Article 2.

Jones, M. (2019). What We Talk about When We Talk about (Big) Data. *The Journal of Strategic Information Systems*, *28*(1), 3–16.

Labadie, C., Eurich, M., & Legner, C. (2020). Data democratization in practice: Fostering data usage with data catalogs. *Communications of the 20th Symposium of the Association Information and Management (AIM)*, 1541.

Labadie, C. *et al.* (2020) 'FAIR Enough? Enhancing the Usage of Enterprise Data with Data Catalogs', in *Proceedings of the 22nd IEEE Conference on Business Informatics (CBI)*. Antwerp, Belgium, pp. 201–210.

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation*. Cambridge University Press.

Lefebvre, H., & Legner, C. (2024). Toward a Curriculum for Data Literacy in Enterprises. *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS)*.

Lefebvre, H., Legner, C., & Fadler, M. (2021). Data Democratization: Toward a Deeper Understanding. *Proceedings of the 42nd International Conference on Information Systems (ICIS)*.

Lennerholt, C., Van Laere, J., & Söderström, E. (2021). User-Related Challenges of Self-Service Business Intelligence. *Information Systems Management*, *38*(4), 309–323.

Mathiassen, L. (2002) 'Collaborative Practice Research', *Information Technology & People*, 15(4), pp. 321–345.

McDermott, R., & O'Dell, C. (2001). Overcoming cultural barriers to sharing knowledge. *Journal of Knowledge Management*, *5*(1), 76–85.

Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, *7*(2).

Mikalef, P. *et al.* (2019) 'Big Data Analytics Capabilities and Innovation: The Mediating Role of Dynamic Capabilities and Moderating Effect of the Environment', *British Journal of Management*, 30(2), pp. 272–298.

Mikalef, P., Pappas, I. O., Krogstie, J., & Pavlou, P. A. (2020). Big data and business analytics: A research agenda for realizing business value. *Information & Management*, *57*(1), 103237.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.

Nicolini, D., Pyrko, I., Omidvar, O., & Spanellis, A. (2022). Understanding Communities of Practice: Taking Stock and Moving Forward. *Academy of Management Annals*, *16*(2), 680–718.

Orlikowski, W. J. (2002). Knowing in Practice: Enacting a Collective Capability in Distributed Organizing. *Organization Science*, *13*(3), 249–273.

Parmiggiani, E., Amagyei, N.K. and Kollerud, S.K.S. (2023) 'Data curation as anticipatory generification in data infrastructure', *European Journal of Information Systems*, 0(0), pp. 1–20.

Parmiggiani, E., & Grisot, M. (2020). Data curation as governance practice. *Scandinavian Journal of Information Systems*, *32*.

Peppard, J. (2018). Rethinking the concept of the IS organization: Rethinking the concept of the IS organization. *Information Systems Journal*, *28*(1), 76–103.

Pyrko, I., Dörfler, V., & Eden, C. (2019). Communities of practice in landscapes of practice. *Management Learning*, *50*(4), 482–499.

Sawicki, D.S. and Craig, W.J. (1996) 'The democratization of data', *Journal of the American Planning Association*, 62(4), p. 512.

Someh, I., Wixom, B., Davern, M., & Shanks, G. (2023). Configuring Relationships between Analytics and Business Domain Groups for Knowledge Integration. *Journal of the Association for Information Systems*, *24*(2), 592–618.

Stein, M. *et al.* (2019) 'Datification and the Pursuit of Meaningfulness in Work', *Journal of Management Studies*, 56(3), pp. 685–717.

Swan, J., Scarbrough, H., & Robertson, M. (2002). The Construction of `Communities of Practice' in the Management of Innovation. *Management Learning*, *33*(4), 477–496.

Tim, Y. *et al.* (2023) 'Digital Resilience: A Conceptual Framework for Information Systems Research', *Journal of the Association for Information Systems*, 24(5), pp. 1184–1198.

Treuhaft, S. (2006). The Democratization of Data: How the Internet is Shaping the Work of Data Intermediaries. *UC Berkeley: Institute of Urban and Regional Development.*, 75.

van Giffen, B., & Ludwig, H. (2023). *How Siemens Democratized Artificial Intelligence*.

Vial, G. (2023). Data governance and digital innovation: A translational account of practitioner issues for IS research. *Information and Organization*, *33*(1), 100450.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365.

Wang, R.Y. (1998) 'A product perspective on total data quality management', *Communications of the ACM*, 41(2), pp. 58–65.

Wang, Y., Blobel, B. and Yang, B. (2022) 'Reinforcing Health Data Sharing through Data Democratization', *Journal of Personalized Medicine*, 12(9), p. 1380.

Weinert, C. and Weitzel, T. (2023) 'Teleworking in the Covid-19 Pandemic: The Effects of Life-Work Conflict on Job Outcomes and the Role of the IT Telework Environment', *Business & Information Systems Engineering*, 65(3), pp. 309–328.

Wenger, E. (1998). *Communities of Practice: Learning, Meaning, and Identity (Learning in Doing: Social, Cognitive and Computational Perspectives)*. Cambridge: Cambridge University Press.

Wenger, E. (2000). Communities of Practice and Social Learning Systems. *Organization*, *7*(2), 225–246.

Wenger, E., McDermott, R. A., & Snyder, W. (2002a). *Cultivating Communities of Practice: A Guide to Managing Knowledge*. Harvard Business School Press.

Wenger-Trayner, E., & Wenger-Trayner, B. (2005). *Introduction to communities of practice*. https://www.wenger-trayner.com/introduction-to-communities-of-practice/

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6th ed.). SAGE.

Zeng, J., & Glaister, K. W. (2018). Value creation from big data: Looking inside the black box. *Strategic Organization*, *16*(2), 105–140.

Zhu, H. *et al.* (2014) 'Data and Information Quality Research: Its Evolution and Future', in H. Topi and A. Tucker (eds) *Computing Handbook, Third Edition*. Chapman and Hall/CRC, pp. 16-1-16–20.

Zhu, S., Gupta, A., Paradice, D., & Cegielski, C. (2019). Understanding the Impact of Immersion and Authenticity on Satisfaction Behavior in Learning Analytics Tasks. *Information Systems Frontiers*, *21*(4), 791–814.

8 Appendix

Section	Questions
1 - Drivers and strategy	What are the drivers for data and analytics in the company? Do you have a data and/or analytics strategy? If yes, since when and what is its focus? What is the business value and benefit created by data and analytics?
2- Scope	Which data domains do you distinguish? How do you define them? Which data types are established or emerging? Which data and analytics products do you deliver?
3 - Data and analytics organisation	What organisational form has been chosen (line function, shared service etc.)? Is the central team/department part of the primary organisation and - if so - where is it located in the organisational structure? What are the responsibilities, headcount, structure and composition of data and analytics teams? Are there any boards and committees for data and analytics? What is their role?
4 - Processes	Which data management processes have you established? Which steps / tasks are taken over by the central / decentral data organisation? Which analytics processes have you established? Which steps / tasks are taken over by the central / decentral data organisation?
5 - Alignment and collaboration	How do you align and collaborate with business stakeholders? How do you align and collaborate with IT stakeholders? How align and collaborate between data and analytics? Which data / analytics communities exist? How do you engage with them?

Table 24. Semi-structured interview protocol

Case	Industry	Revenue/ employees	Key informants	Research activities	Examples of CoP
A	Public transportation	$1B–$50B/~35,000	Product owner data strategy; Enterprise Architect for Data & Analytics	1	Analytics capability community; Data science community
B	Manufacturing, chemicals	$1B–$50B/~5,000	Head of Corporate Data Management	1, 2	Master Data Lunch; Master data material community
C	Packaging, food processing	$1B–$50B/~25,000	Director of Global Master Data Strategy	1, 2	BI community; MDM community
D	Manufacturing, automotive	$1B–$50B/~90,000	VP Data & Analytics Governance	1, 2	Data domain manager round table; Global Data science and AI community; Enterprise Architecture community SAP analytics; Data quality circle;
E	Consumer goods	$50B–$100B/~350,000	Master Data Lead; Group Manager Data and Analytics Products & Services	1, 2	Master Data community; Analytics communities for specific tools (e.g., PowerBI)
F	Manufacturing, automotive	$1B–$50B/ ~150,000	Head of Master Data Management; Head of Advanced and Self-Service Analytics	1, 2	Master Data Management community; Data Science and Analytics Experts groups
G	Pharmaceutical	$1B–$50B/ ~70,000	Global Data Lead; Enterprise Solutions Architect Analytics Lead	1	Monthly global communication of data practice
H	Consumer goods, retail	$1B–$50B/ ~30,000	Vice-President: Data & Analytics	1	Data domain working groups
I	Consumer goods, retail	$100B–$150B /~450,000	Head of Enterprise Data Management	1, 2	Data sharing community; GS1 community
J	Chemicals	$50B–$100B/~120,000	Product Manager Data Governance & Stewardship	1	Reporting & Analytics community; Data steward community
K	Fashion and retail	$1B–50B/~60,000	Head of Data Quality; Data Catalog Community Governor	1, 2	Data Quality community; Data catalog community
L	Pharmaceutical, chemicals	$1B–$50B/~100,000	Head of Enterprise Master Data	1, 2	Master Data Management community
M	Pharmaceutical devices	$1B–$50B/ ~65,000	Senior Manager Business Analytics	1, 2	Master Data Management community
N	Adhesive & beauty products manufacturing	$1B–$50B/~20,000	Director Master Data & Product	1	Data Expert community linking regional hubs
O	Outdoor power products manufacturing	$1B–$50B/~20,000	Senior Director Business Transformation Data Management	1	Data governance community
P	Technology & networks	$100B–$150B /~200,000	Head of Corporate Data Management	1, 2	Data modelling community; Data quality community
Q	Pharmaceutical	$1B–$50B/~70,000	Enterprise Data and Analytics Operations Cluster Chair; Finance Data Director	1, 2	Supply Chain Master Data Team; Customer data team
R	Software development	$1B–$50B/~100,000	Solution Advisor Expert	2	Material master data community; Customer master data community; Governance learning series
T	Network & telecoms	$1B–$50B/~100,000	Head Information Operations Management	2	Information and Data Architecture community
U	Logistics operations	$1B–$50B/~70,000	Head of Global Master Data Management	2	Data mobilisation

Table 25. Companies involved in the research process and example of communities

Essay 3

Toward a Curriculum for Data Literacy in Enterprises

Hippolyte Lefebvre and Christine Legner

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

Published in the proceedings of the 57th Hawaiian Conference on System Sciences (2024)

Abstract: *To create business value from data, firms need a data literate workforce capable of reading, working, analyzing, and arguing with data. Prior studies on data literacy have mostly focused on educational settings and identified data-related skills. However, the suggested generic skill catalogs do not account for the highly situated nature of data practices. In this paper, we delve into five data literacy programs at multinational companies and examine their unique scope and characteristics. We leverage curriculum theory to analyze the different curriculum components and how they foster workplace data practices. As a contribution to data literacy research, we propose a theory-inspired and situated curriculum for data literacy in enterprises built upon five learning blocks, namely generic skills, disciplinary content, disciplinary skills, workplace awareness, and workplace experience. We also disclose each block's target audience, scope, and delivery mode and thereby inform practitioners on how to build their own curricula.*

Keywords: Data literacy, Data competences, Data skills, Data practices, Curriculum

Table of contents

List of figures

List of tables

1 Introduction

Firms increasingly recognize the strategic potential data has and they seek to bring an increasing number of employees on board to engage in data-related activities. Expanding data practices to beyond the data expert domain requires a data literate workforce, i.e., employees that are able to read, work, analyze, and argue with data (D'Ignazio & Bhargava, 2015). For instance, business managers should be able autonomously to tackle basic tasks such as defining the requirements of a simple dashboard, while also accessing and analyzing data on it (Lennerholt et al., 2021). However, most companies still lack a workforce with the requisite data skills that would allow them to draw meaningful business insights out of data (Grover et al., 2018). Besides, managers tend to overestimate their workforce's capabilities and their readiness to work with data (Vohra & Morrow, 2020). Also, they struggle to establish working relationships between business and data experts (Redman, 2022).

Data literacy has mostly been studied as an educational theme that equips students with a list of core skills to prepare them for the job market (Carlson et al., 2013; Koltay, 2017). In the enterprise context, data literacy is often embedded in digital literacy research (Cordes & Weber, 2021; Goel et al., 2021). Thereby, it does not account for the distinctive nature of data (Paparova, 2023), nor for the fact that data use is highly situated (Alaimo & Kallinikos, 2022). Since data mainly gains value when it is put to use as a part of local actors' sense-making processes (Aaltonen et al., 2021), data literacy must be taught in context (Jones, 2019; Micheli et al., 2020). Although some studies have focused on listing skills for certain roles such as for data scientists (Demchenko et al., 2016; Saltz et al., 2018), they overlook the majority of employees who are less technically skilled but should still have a key role in creating value by making business sense out of the data they work with.

In contrast to IT literature that has investigated upskilling to address the surge in demand for IT workers at the beginning of the millennium (Ho & Frampton, 2010), very little research has looked into today's need to train a larger emerging community of employees to fulfil data roles in context. Considering this lacuna, we ask the following research question:

RQ: *How do companies develop data literacy programs to upscale their data practices*?

We opted for multiple case studies to capture rich and diverse insights directly from practitioners' working contexts (Paré, 2004). Based on Bennet et al.'s (1999) curriculum model, we analyzed data literacy programs with different scopes and target audiences from five multinational companies. This prism helped us to examine the cases using a common framework, to compare them, and to identify recurring patterns (Miles et al., 2014). As a contribution to data literacy research, we propose a theory-inspired and situated curriculum for large-scale data literacy programs that comprises five learning blocks covering *generic skills*, *disciplinary content*, *disciplinary skills*, *workplace awareness*, and *workplace experience*. For each block, we indicate the target audience, the scope, and the delivery mode. Besides contributing to data literacy research, our cases and findings inform practitioners on how to build their data literacy curriculum.

In this paper, we first review the literature on data literacy and competence development, and we identify the research gap. Second, we explain our case study methodology and the research process. Third, we present our findings related to the five learning blocks. Finally, we discuss our findings and provide an outlook on future research.

2 Background

2.1 Data literacy

Data literacy refers to the ability to read, work, analyze, and argue with data (D'Ignazio & Bhargava, 2015). Interestingly, the existing body of knowledge on data literacy has mainly built upon concepts from educational research (e.g., high school, university) and library studies to define and investigate data literacy as a bundle of skills (Calzada Prado & Marzal, 2013; Carlson et al., 2011; Ridsdale et al., 2015). To identify them, researchers have mainly analyzed data experts' profiles and derived a set of generalizable and context-independent skills. Across these studies, data literacy is traditionally associated with a generic set of key skills such as data analysis, data curation, data visualization, data ethics and data security (see Table 26). More recent studies have extended these sets of data literacy skills for work and society (Schüller, 2020; Sternkopf & Mueller, 2018; Wolff et al., 2016). The resulting set of skills also reflects their application in a more specific context e.g., in developing hypotheses, identifying related sources of data that could support an investigation, accessing data, analyzing and creating explanations from data, or communicating with data. Additionally, a recent understanding of data literacy not only encompasses skills, but also includes behavioral dimensions such as attitude and values toward data (e.g., act data driven, data ethics).

Research has also emphasized the role of situated learning as pivotal for employees participating in data and analytics activities (Dubey & Gunasekaran, 2015; Lefebvre & Legner, 2022). Hence, data literacy cannot be characterized as a passively ingested skills set, which is detached from actual work applications (Zhu et al., 2019). Thus, we distinguish between skills as static uncontextualized properties and competences, i.e., abilities to apply a job's requisite skills (Bartram, 2005). Competences, then, refer to the ability to put the developed generic (i.e., cross-discipline) and situated (i.e., specific to workplace) skills into practice. Applied to the enterprise context, the goal is to develop employees' competences so that they are able to use and make sense of given data in a way that supports their daily work (Aaltonen et al., 2021). Finally, applying the behavioral competence approach (McClelland, 1973) to data literacy suggests that competences are not innate, and can be taught through programs that combine generic upskilling and workplace-relatable content. Data literacy then becomes a personal trait or set of habits that can lead to better job performance.

Research field and sources	Examples of data literacy skills
Library Studies / Education (Carlson et al., 2011) (Calzada Prado & Marzal, 2013) (Ridsdale et al., 2015)	• Data discovery and acquisition • Data management • Data visualization • Data curation • Data processing • Data analysis • Data ethics and security • Data culture
Work and society (Wolff et al., 2016) (Sternkopf & Mueller, 2018) (Schüller, 2020)	In addition to the above: • Act data driven • Solve a problem with data • Identify data use cases • Coordinate data use cases • Evaluate impact of data • Trace back data transformations

Table 26. Generic data literacy skills in the literature

2.2 Competence development and curriculum

Firms have acknowledged that developing talent and learning is vital for sustaining their business. Thus, they seek to equip their employees with the necessary competences to sustain such a new and competitive environment (Ho & Frampton, 2010; Merchel et al., 2021). Specifically, competence development has become a critical factor in preparing employees for a more technology-driven future (Li, 2022). Such development is often associated with a set of learning outcomes based on expected job qualifications. These learning outcomes support the mapping of learning content into a curriculum (Walker, 2003). A curriculum is defined as a collection of documents and learning activities aiming to deliver a structured series of learning experiences. It includes theoretical and practical content to equip learners with predefined competencies (Prifti, 2019). Clarifying learning outcomes ensures dedication to advanced proficiency levels and defined learning paths (von Konsky et al., 2016). A larger group of employees can share a subset of competences; yet, individuals' competences that are associated with personalized learning outcomes indicate that most competences are expected to address situated practices (Le Deist & Winterton, 2005). Despite the evidence of recent training success, many training programs still neglect the role of workplace experience, disregarding different learning formats such as secondment or projects (Zhu et al., 2019). Further, learning materials are key elements in training and its success. Companies should, therefore, select and organize learning materials to meet both generic and situated learning expectations (Wang et al., 2014).

As a general framework, Bennet et al.'s (1999) view on learning suits the development of data literacy well, as it is a highly situated competence developed through collective understanding and workplace-like experience. Their curriculum model helps to bridge the gap between classroom training practices and workplace expectations. The model displays five blocks representing components to be enabled for learning success, identified as generic skills, disciplinary content, disciplinary skills, workplace awareness, and workplace experience (see Figure 8).

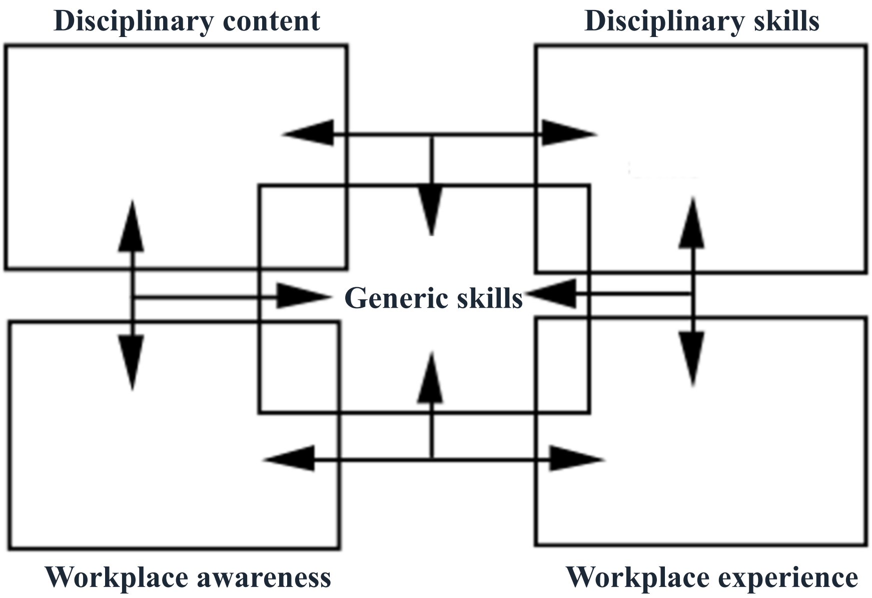


Figure 8. Curriculum model by Bennett, Dunne and Carré (1999)

Despite variations across disciplines, the generic skills in the middle of the model support all the other blocks by providing the necessary skills to engage in situated learning. Disciplinary content refers to conceptual knowledge corresponding to a trainee's own discipline. Trainees can develop a wider set of disciplinary skills relevant to disciplinary content and generic skills, which they can leverage and apply in a simulated (workplace awareness) or real (workplace experience) environment. The connections (arrows) between the blocks indicate the directionality of learning i.e., the options companies have in sequencing the learning blocks.

3 Methodology

Considering our research goal, we chose a qualitative research design using multiple case studies to investigate how companies develop their data literacy learning journeys (Paré, 2004). Case studies, as "*well-suited to capturing the knowledge of practitioners and developing theories from it*" (Benbasat et al., 1987, p. 370), are commonly used for answering "how" questions and multiple cases support better analytical generalization (Miles et al., 2014).

Data collection happened in two phases. The first entailed a focus group in June 2021 to exchange knowledge of best practices for developing data literacy competences. The participants were 12 experts from 9 companies representing different industries, with differences in scope (e.g., data analytics, data management) and maturity for data literacy. Since combining focus groups and surveys is generally recognized as suitable for sampling cases (Morgan, 1993), we used a survey to capture examples of data literacy initiatives, their target audiences, and their development phase. Following the survey results, we identified a subset of five mature data literacy training programs at five different companies. They differ in scope, audience, and industry (see Table 27). The second phase entailed semi-structured one-hour interviews with each of the five companies between July 2021 and November 2022. Preparing for the interview, we asked key informants (e.g., project manager, director analytics) to provide an overview of their data literacy curriculum. Our interview questions covered the theoretical framework's five areas to ensure results would be compatible, and we sent interview notes to the interviewees for validation. To enrich the case database and triangulate primary data, we searched for secondary data (e.g., press reports, presentations, company documentation). In this way, we also ensured reliability of the evidence. The five cases allowed us to reach theoretical saturation as we noticed redundance in incremental learning, for instance in patterns against our theoretical framework.

Data literacy program	**Industry**	**Audience (~# employees)**
*R&D Academy – AI & Data Analytics (**A**)*	Manufacturing, automotive	R&D community (20,000)
*Enterprise Data Literacy (**B**)*	Packaging, food processing	All employees (20,000)
*Roadmap for data handling & understanding (**C**)*	Manufacturing, automotive	IT & Digitalization (15,000)
*Digital Analytics Academy (**D**)*	Fashion and retail	Digital unit in Sales (400)
Data Literacy Journey (**E**)	FMCG	Operations & Sales (5,000)

Table 27. Cases overview

First, we ensured a thorough understanding of the context for each case (e.g., target groups, scope). For the analyses, we leveraged the theoretical insights on workforce development (see section 2.2), using Bennet et al.'s (1999) model as framework for individual analysis of the cases (within-case analysis). One researcher coded the case base against the framework dimensions (generic skills, disciplinary content, disciplinary skills, workplace awareness, workplace experience) and a second researcher reviewed the codes. The two researchers cleared the coding during a meeting in June 2023. Table 28 illustrates the coding process for case B.

The comparative analysis is particularly relevant to this study as it supports the aggregation, simplification, and generalization of complex cases (Miles et al., 2014). Moreover, natural variation between cases generally strengthens theory building (Dubé & Paré, 2003). For the cross-case analysis, we performed "pattern-matching," thereby identifying differences and commonalities on the learning blocks level to determine similar ways of developing both generic and situated learning. We iteratively searched for similarities between codes (e.g., all "workplace awareness" codes) and then created and grouped types of codes to examine cases for shared configurations. These we summarized in a curriculum based on the identified five blocks (see results in section 5).

Case description	**Coding**	**Explanation**
Data ethics class in the form of an e-learning for all based on a LinkedIn playlist.	Generic skills	Data ethics is currently a typical skill in all data and analytics roles.
"Data Playground" as a new data experimentation platform where trainees are assigned data experts as mentors.	Workplace awareness	Application of the skill in the form of simulation fosters situated learning.
70% of learning journey should happen in the workplace (e.g., projects and job rotations).	Workplace experience	Skills development primarily happens on-the-job.

Table 28. Within-case coding examples for Case B

4 Cases

Below, we describe each case in details, also showing how every case maps onto Bennet et al.'s (1999) five framework enablers for workplace upskilling. While directionality of learning is briefly addressed in each case narrative, our analysis focuses on each curriculum's learning blocks rather than their sequencing. We provide a figure that summarizes the case analysis and highlights the key blocks enabled by the data literacy curriculum in dark grey (*Major*), and the blocks enabled but at a lower intensity in light grey (*Minor*), with unactivated blocks in white.

4.1 Case A: R&D Academy – AI & Data Analytics Landscape

Company A is a large automotive supplier ($1B–$50B revenue/~150,000 employees) that invests considerably in next generation mobility ($2.5B in 2022), e.g., in automated driving. The firm released a "data enablement strategy" in 2020 in planning for data-driven innovation. Accordingly, they set up a data enablement team to break down data silos and to stimulate collaboration on various data use cases between data and business experts. As a first step toward their data-driven business model, the firm decided to focus on upskilling more than 15 000 employees in the R&D department. The company started developing a data literacy program, the *R&D Academy,* dedicated to the entire R&D community. Before this, only a few data experts had benefited from comprehensive data literacy training programs. It chose to personalize the program centered on three role families: *employees/managers*, *domain developers/subject matter experts*, and *AI experts*. Training is optional and the content is structured in one of the following proficiency levels: I-*Create Awareness* aims to raise awareness of the company's business and data strategies and their impact on R&D, and introduces selected foundational topics to employees/managers and domain developers/subject matter experts. At proficiency level I, role families can benefit from an introduction session on AI, Data Science, and Machine Learning leveraging LinkedIn Leaning Playlists. Level II - *Gain Deeper Understanding* focuses on R&D role families' specific technological and technical competences by offering one-day to three-day qualification courses. Level III-*Achieve Enablement* enables selected R&D engineers to fulfil the requirements of their specific technology domains through longer qualification programs (>10 days). At level III, they offer an expert program that enrolls 40 engineers per semester. So far, the program relies largely on virtual content such as sourced e-learning (e.g., LinkedIn, Udacity) and knowledge sharing via the analytics communities; however, the firm anticipates bringing in other learning formats such as conferences (more than 1000 participants from eight divisions during the 2022 edition). Further, an "AI adventure" program is being planned to raise non-experts' awareness of AI through a collaborative game presenting mini problems to solve with data. Figure 9 maps case A onto Bennet et al.'s (1999)'s building blocks.

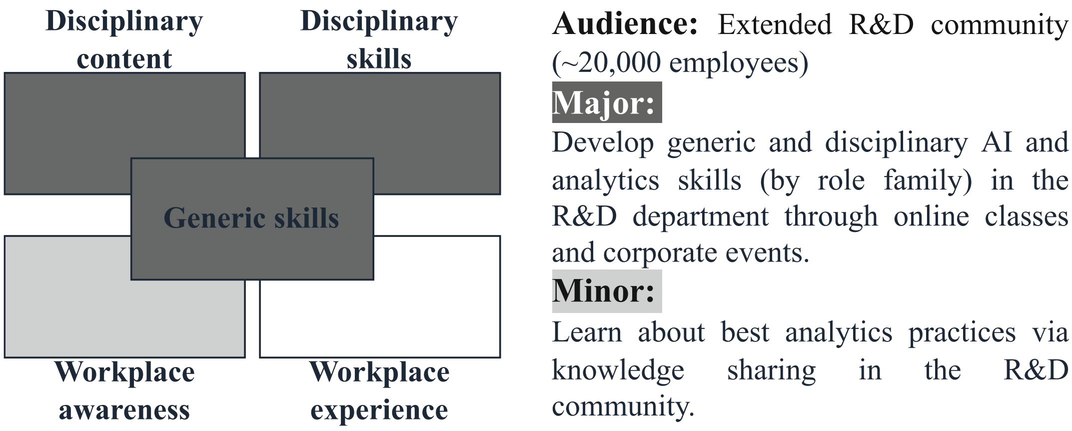


Figure 9. Case A mapped onto Bennet et al. (1999)'s framework

4.2 Case B: Enterprise Data Literacy

Company B is a large multinational ($1B–$50B revenue/~25,000 employees) operating in packaging and food processing. It has identified operational excellence as a key enabler in its business strategy, "Company 2030." They described the required data capabilities in their 2019 data and analytics strategy. The firm has been developing a corporate-wide data literacy initiative called *Enterprise Data Literacy* (EDL), a recommended but not obligatory program, by which they aim to upskill 20 000 employees. The company decided to design EDL for three participant groups: *Data citizens* (all employees) who should understand why data is important and how it is used for the firm's business; *business analysts* (e.g., a marketing analyst) who should have strong domain knowledge and be able to talk comfortably with the third group called *citizen data scientists*. The learning outcomes for each participant type are divided into to three proficiency levels, i.e., the *Conceptual*, *Core* and *Advanced* levels. EDL is implemented on an EdCast platform, mostly offering classes sourced from LinkedIn and addressing all proficiency levels. The classes are bundled into introductory "learning journeys" to inspire all role players. For more advanced players it deep dives into defined areas allowing them to "pick-and-choose" what is most relevant for them. However, this "structured learning" represents only 10% of EDL's learning design framework. The next 20% is about "learning from others," which aims to sustain the learning momentum through social and collaborative knowledge sharing. Thereby, employees can benefit from coaching and mentoring opportunities with experts or join communities of practice (e.g., a Business Intelligence (BI) community). Currently under construction, is a "Data Playground" that will offer a safe space for employees to practice data analytics skills. The last 70% is about "learning from experience" and integrating learning with work. This longer-term part of the program expects freshly trained employees to develop sustained autonomy in taking action and solving problems with data. Depending on the specific project or assignment, placements, secondments, and job rotations can be involved. Figure 10 summarizes our analysis based on the building blocks suggested by Bennet et al. (1999).

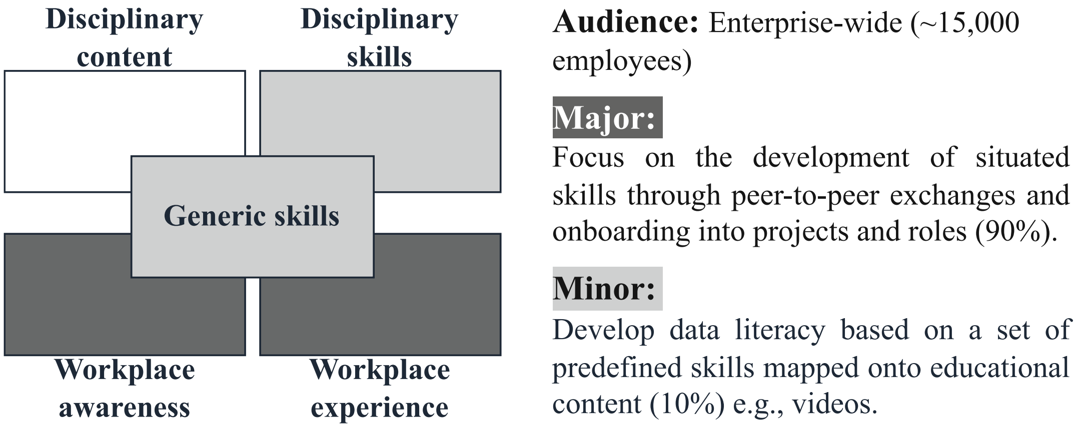
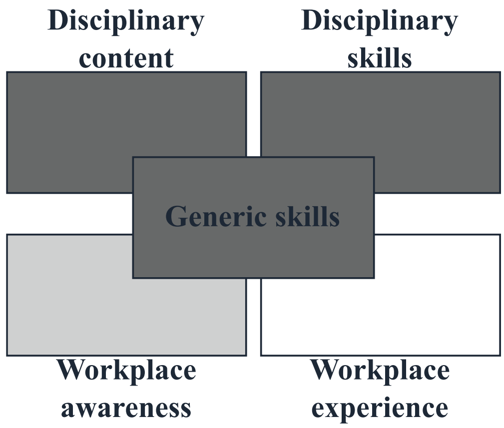


Figure 10. Case B mapped onto Bennet et al. (1999)'s framework

4.3 Case C: Roadmap for data handling and understanding

Company C is a large multinational ($1B–$50B revenue/~90,000 employees) involved in the automotive and manufacturing business. The firm seeks to develop a data culture to accompany its recently released data and analytics strategy (2021) focusing on industry 4.0 and AI in business processes. After releasing a new data organization, the firm needs to provide improved data access and to develop data and analytics skills for new roles. The company is developing a project, *Roadmap for data handling and understanding,* to increase awareness and to upskill various roles in several digitalization areas through a structured learning program. The program is designed for three proficiency levels, i.e., *Basic Knowledge –Interested and Affected by Digitalization*, *Advanced – Participate in Digitalization*, and *Experienced – Realization of Digitalization Projects*. While all employees in the IT and digitalization department are expected to know the foundations for data handling, most of the training is role-specific. For instance, the basic level includes generic and role-specific courses: generic courses intended for all participants offer e-learning content (e.g., What is BI? What is a digital twin? Introduction to data management), while role-specific courses (e.g., Data management basics, Data-driven decision making; Self-service BI basics) address different kinds of data expertise. The advanced and experienced levels are fully role-specific. They cover different specialized topics depending on the trainee's role: data analysis, data science, digital twinning, semantic models, and data management. To illustrate, the advanced level includes classes such as Consuming Analysis for Office, Consuming SAP Analytics Cloud, Digital Twin API hands-on, Semantic Modelling Fundamentals, or Data Modelling & Data Catalogue. Eventually, the experienced level aims to upskill "data-citizen" roles and IT roles by offering classes focused on creation and innovation, such as Design Principles for Self-Service BI, Data Science Workbench, Semantic Modelling – Projects, Data Management Processes, How to think like a Data Scientist. Also, the company is exploring alternative formats such as mini projects, while investigating the synergies with several existing communities of practice. Overall, Figure 11 maps case C onto Bennet et al. (1999)'s framework.



Audience: IT & Digitalization (~15,000 employees)

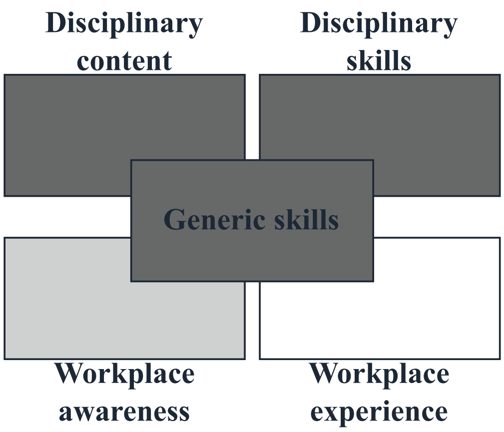
Major:
Develop data and analytics in IT and Digitalization. Courses are mainly taught by peers and are applied according to the degree of participation into digitalization (passive vs active).

Minor:
Use of mini-projects as classes for most advanced profiles.

Figure 11. Case C mapped onto Bennet et al. (1999)'s framework

4.4 Case D: Digital Analytics Academy

Company D is a large fashion company ($1B–$50B revenues /~60,000 employees) experiencing a digital transformation of its sales channels, notably triggered by a surge in digital sales during the Covid-19 period. In this context, data literacy is mentioned as a strategic enabler for their large digital sales unit which is responsible for e-commerce and digital activities, including sales growth and advertising. The department expects the employees to be able to generate and leverage data-driven insights that help digital sales growth (e.g., using metrics to track net sales or to monitor product demand within and across e-commerce channels). This includes business roles, such as product category managers or digital marketing specialists, as well as data experts who currently lack integrating the business context when developing analytical products. For instance, product category managers currently struggle to find the information they need for decision-making or do not act on the analytical insights provided by the analytics team. To design its data literacy program (focused on analytics and data-driven insights), the company first reviewed the organization and the different analytical roles and value drivers. Then, they created a list of 13 job families (e.g., digital activation, digital marketing, product ownership, decision science). Through a comprehensive analysis of skills and job descriptions (internal and external), the firm derived 25 analytical skills groups (and more than 400 skills) to map onto the 13 job families. Company D then created a skill finder tool which is fed by raw data from the skill mapping project to support skills discovery for each employee. This enabled the development of individual learning paths for the different job families across three core areas: tools/dashboards, KPIs and data, technique and skills. Each of these areas is linked to six learning outcomes representing the different cognitive steps of learning: awareness, meaning, adoption, interpretation, communication, creativity. Accordingly, by Q1 2023, the company expects 100% of the employees in the digital sales unit (e.g., campaign managers, data scientists) to be data aware (i.e., achieved the *awareness* learning outcome) and by Q2 2023 they expect 60% of the digital sales unit to use a set of key dashboards in self-service at least on a quarterly basis. As a first step, the company offered several awareness sessions on MS Teams with 100+ digital sales employees to emphasize the importance of data (e.g., how metrics can provide business insights). The program has already shown progress: within a year, the number of consumers on the academy's SharePoint has tripled, as have the visits on the key dashboards. Figure 12 maps Case D onto Bennet et al. (1999)'s framework.



Audience: Digital unit in Sales department (~400 employees)

Major:
Develop essential analytics skills relevant to sales department mainly with videos externally sourced.

Minor:
Learning context is augmented with internal materials, e.g., link to sales dashboards.

Figure 12. Case D mapped onto Bennet et al. (1999)'s framework

4.5 Case E: Data Literacy Journey

Company E is a large FMCG company ($1B–50B revenue/~60,000 employees) halfway through a large business transformation started in 2017. The radical shift toward an electronic device product line required that the firm invest considerably in its digital capabilities. In this context, the firms embarked on a large data and analytics journey which started with a data and analytics organization of eight people tasked with setting up the data foundation (e.g., data governance, management, and quality control, and a business glossary) while drafting the initial analytical roadmap and engagement. Within six years, the firm managed to roll-out a 100+ FTE data organization with strong data management and analytics capabilities. For instance, in 2020, they had started developing 28 data science use cases (400+ million USD). In developing data roadmaps for various business functions, the firm realized soon that data literacy is a central skill in decentralized data enablement. Starting in the department with the highest needs, in this case the Operations and Sales department, company E identified four role families to be trained, i.e., *senior executives* (C-Suite roles and their direct reports (e.g., CEO, SVPs, VPs), *business leaders*, (e.g., directors and managers, data owners and stewards, subject matter experts, analytics product owner), *specialists in data-and-analytics functions* (e.g., business analysts, visualization experts), and *technical experts* (e.g., data architects, data engineers, data scientists, source systems specialists). After performing a skill gap analysis, the firm developed a pilot data literacy program called *Data Literacy Journey* focusing on a cohort of 400+ business leaders. They were considered the primary consumers of data-driven insights (e.g., in defining and using metrics, identifying opportunities for data use cases, taking responsibility for local data collection and quality). Figure 13 maps case E onto the Bennet et al. (1999) framework.

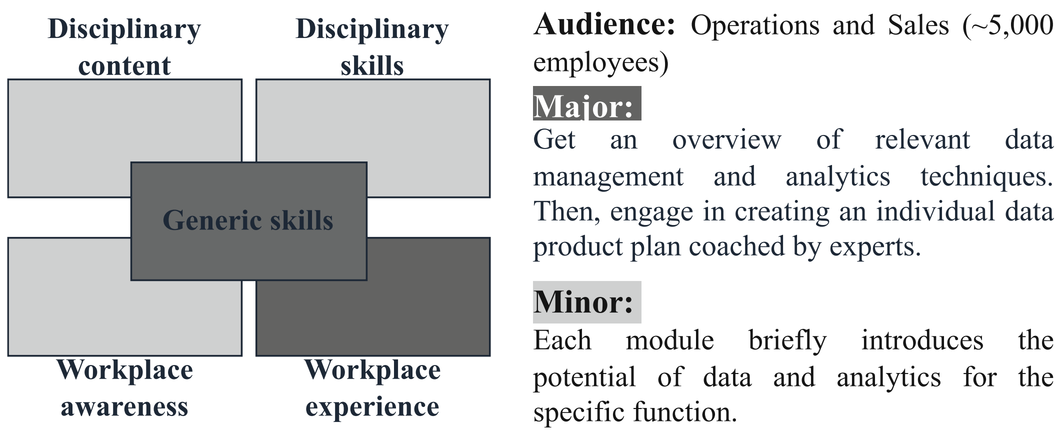


Figure 13. Case E mapped onto Bennet et al. (1999)'s framework

The training program offers 1) a three-hour pre-work self-paced awareness course); 2) 20 hours of virtual instructor lead training (VILT) fostering engagement and interaction in two modules, 9 hours of introduction to data and data products, articulating a business problem, metrics, data management basics, visualization, and storytelling, and 11 hours of a course introducing ML, data governance, and digital platforms.; 3) ongoing engagement after training through self-paced assignments (creating an individual data product plan with expert coaching). Upskilling materials are sourced from externally available programs and platforms (e.g., IMD business school, Coursera) and augmented with relevant company-specific content such as real-life use case examples. After being piloted, the program was scaled up to train 5000+ business leaders within 18 months, including 3000+ in the consumer and commercial department.

5 A data literacy curriculum built on five blocks

We generalize our findings in the form of five building blocks for a data literacy curriculum (see Figure 14). For each, we highlight the key findings on audience, scope, and delivery mode. Our model is generic enough to offer flexibility in customizing the learning outcomes and direction of learning.

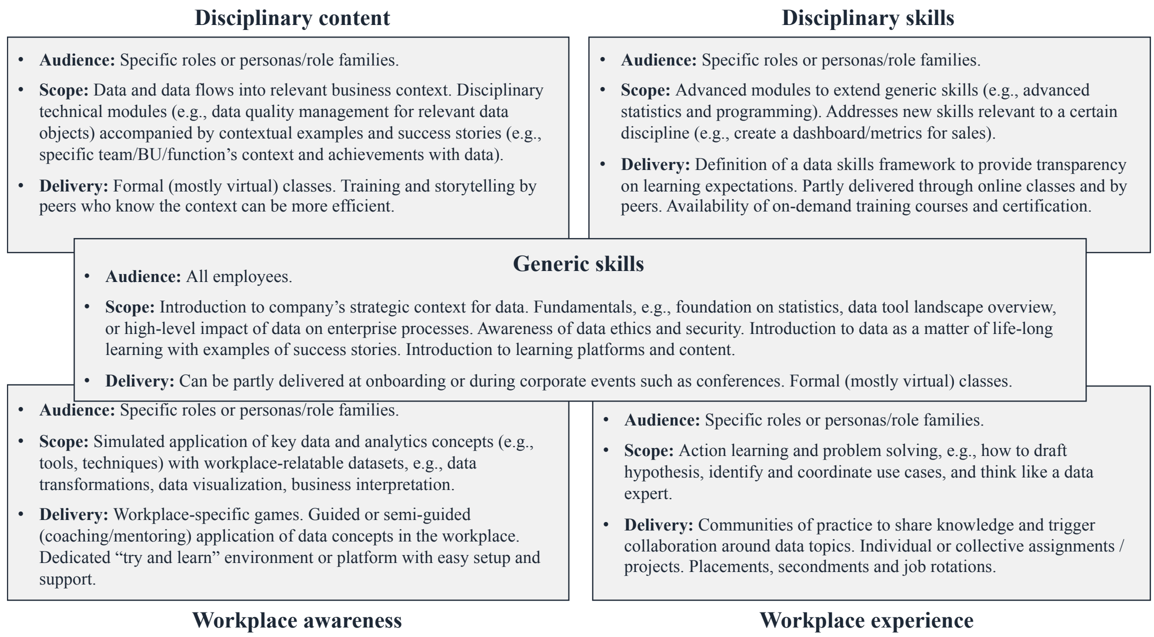


Figure 14. Data literacy curriculum

A key motivation for developing data literacy is to enable employees with different data backgrounds to work, collaborate, and communicate with others about data or in projects. We observed that data literacy programs have a common baseline or include foundational skills for the entire audience, which we characterize as **generic skills** since they should be transferrable to any work environment. In all cases, we found that common learning outcomes encompass motivational topics on the value of data in the context of the company's strategy. As highlighted in case D, these topics could also support the development of certain essential skills such as foundations of statistics, data tool landscape overview, or high-level impact of data on enterprise processes. Trainees are mostly expected to ingest basic concepts and be able to apply them in a meaningful way. It is also important for all role players to understand the impact of data literacy on their career progression, also in using success stories. Curiosity to engage in upskilling should be fostered at this stage.

Beyond generic skills, employees further need to engage in additional modules that show how data can be used in their specific (business) context, i.e., they need **disciplinary content**.

For instance, after having understood what tools (e.g., BI tools) and techniques (e.g., checking duplicates) are available to analyze data, one needs to understand how these relate to their disciplines i.e., their specific working environment. Here, both data specialists and experts should know about data's impact on specific business processes and other possibilities of value creation from data. As highlighted in case C, disciplinary content can be taught by peers, for instance data experts or subject matter experts.

As a participant of the focus group mentioned: "*Not everyone needs to be a data scientist.*" **Disciplinary skills** aim to transform disciplinary content into situated data activities, i.e., they are the skills necessary to use data in daily work. They are typically aligned with job descriptions. Hence, firms should communicate skills expectations for different roles and job levels, for instance in the form of a skills framework. In all cases, we found the development of disciplinary skills should be stimulated with advanced modules either cultivating basic concepts taught as part of generic skills, or with new learning materials specific to the working context. Firms should also offer trainees the possibility of requesting additional training and certification to sustain engagement.

Courses providing **workplace awareness** aim to support the application of theoretical knowledge in a simulated environment, as authentically as possible. To do so, these courses are organized for specific personas or role families. As in cases A, B, and C, firms can set up playful activities (e.g., gamification, workshops, or mini-projects) to immerse trainees (any persona or role families) in workplace-relatable problems and challenges. They can also offer dedicated "data sandbox" environments (e.g., an analytics platform) which approximate the workplace activities. Trainees can then benefit from ongoing support from their peers and especially from data experts to learn about their use of data. Workplace awareness is critical for trainees' sustained engagement and satisfaction since it becomes a first bridge between theory and practice.

Workplace experience is about inviting trainees to take on data responsibilities and commit to a continuous learning journey. Trainees are part of a data users' community from whose experience they can benefit. Employees benefit from knowledgeable community members by deriving mental frameworks to address typical data-related challenges or to work on solutions. As in cases B and E, employees can then be paired with experts on projects so that together they can contribute to visualizing data use cases. To gain expertise trainees can also be seconded, placed in temporary positions, or in a job rotation.

6 Discussion and implications

Overall, our results resonate with the ongoing discourse on data as a matter of practices (Aaltonen et al., 2021). Users interest in training offers and their desire to develop the required workplace competences depend on a proper fit between the curriculum and realistic workplace expectations. This is highlighted in Case B that offers a new and highly situated pattern of curriculum provision adding to the six patterns already identified by Bennet et al. (1999). Our results show that data literacy curricula should offer personalized learning paths that address specific audience needs, including those of existing data roles and of data experts who have often been neglected in existing data literacy literature. We derive and propose three typical persona requiring data literacy training: data amateurs (e.g., casual data consumers with no data responsibility), data specialists (data consumers or creators for whom data is a part of their work routine, e.g., business managers, data owners), and data experts (data professionals who can act as coach e.g., data scientists, data quality manager).

Our cases also show that learning outcomes vary considerably across persona. Data literacy encompasses more than a simple set of generic skills (such as the ones in Table 26). The context-specific nature of data literacy also requires situated enablement by means of disciplinary content, disciplinary skills, workplace awareness, and workplace experience. In other words, apprenticeship will hopefully lead employees from novice levels to mastery (Gherardi, 2000). Further, we find that many data literacy skills (e.g., communicating with data, presenting with data) can be interpreted as generic and transferrable to various work environments. These skills become disciplinary depending on the associated level of proficiency. In fact, a single skill can be observed at various cognitive levels, i.e., ingested rather passively or by enacting it in practice. This marks the distinction between "*knowing that*" and "*knowing how*" (J. R. Anderson, 1983). For instance, the seminal Bloom's taxonomy suggests six progressive levels of cognitive learning identified as remember, understand, apply, analyse, evaluate, and create (L. Anderson et al., 2001). Hence, firms should clarify learning outcomes in terms of the level of cognition applied to the skills and should ask themselves questions such as: *When I conceptualize data analysis as a skill, what do I expect concretely from a given target group?* We believe this crucial point unlocks opportunities for further research on cognitive expectations for different data and analytics roles and on the pre-requisites and skills at the boundaries between roles.

Further, our cases add to existing evidence that shows how a diverse learning toolbox is a success criterion for skill transformation in enterprises (Billing et al., 2021). Companies should complement their own business-specific materials addressing disciplinary content with content from third-party providers or external mainstream sources. Such learning design is essential to trigger behavior change toward establishing a data culture. For instance, several cases in our study by default relied on external learning platforms, such as the prominent LinkedIn learning. A data manager in our focus group said about the latter: "*We are pragmatically using what learning opportunities are already available to us, and ideally they should be free.*" Additionally, researchers could do a more detailed study of what makes a successful data literacy learning environment.

To conclude, we contribute to data literacy research on various levels. First, we offer a theory-inspired and situated curriculum concept that relies on successfully enabling learning blocks to develop data literacy in enterprise. Second, we provide detailed descriptions of five data literacy programs with different scopes and target groups, and we highlight data literacy curriculum patterns. Third, by offering a blueprint for developing data literacy curricula, this research will also inform the practitioner community.

Our study comes with certain limitations. Our sample includes only large multinational companies with a certain level of experience in data management and analytics, and with access to human and financial resources. Therefore, our findings may not be generalizable to smaller companies and their specific challenges (e.g., a smaller audience for data literacy, lack of data awareness and organization).

7 References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. *Journal of Management Information Systems*, *38*(2), 401–429.

Alaimo, C., & Kallinikos, J. (2022). Organizations Decentered: Data Objects, Technology and Knowledge. *Organization Science*, *33*(1), 19–37.

Anderson, J. R. (1983). *The Architecture of Cognition* (Cambridge, MA: Harvard University Press).

Anderson, L., Krathwohl, D., Airasian, P., Cruikshank, K., Mayer, R., Pintrich, P., Raths, J., & Wittrock, M. (2001). *A Taxonomy for Learning, Teaching, and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York : Longman.

Bartram, D. (2005). The Great Eight Competencies: A Criterion-Centric Approach to Validation. *Journal of Applied Psychology*, *90*(6), 1185–1203.

Benbasat, I., Goldstein, D. K., & Mead, M. (1987). The Case Research Strategy in Studies of Information Systems. *MIS Quarterly*, *11*(3), 369–386.

Bennett, N., Dunne, E., & Carré, C. (1999). *Patterns of core and generic skill provision in higher education*. *Higher Education*(37), 71–93.

Billing, F., De Smet, A., Reich, A., & Schaninger, B. (2021). *Building workforce skills at scale to thrive during—And after—The COVID-19 crisis*. McKinsey. https://www.mckinsey.com/capabilities/people-and-organizational-performance/our-insights/building-workforce-skills-at-scale-to-thrive-during-and-after-the-covid-19-crisis

Calzada Prado, J., & Marzal, M. Á. (2013). Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents. *Libri*, *63*(2).

Carlson, J., Fosmire, M., Miller, C., & Sapp Nelson, M. R. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Portal: Libraries and the Academy*, *11*(2), 629–657.

Carlson, J., Johnston, L., Westra, B., & Nichols, M. (2013). Developing an Approach for Data Management Education: A Report from the Data Information Literacy Project. *International Journal of Digital Curation*, *8*(1), 204–217.

Cordes, A.-K., & Weber, J. (2021). Design Principles for Digital Upskilling in Organizations. In F. Ahlemann, R. Schütte, & S. Stieglitz (Eds.), *Innovation Through Information Systems* (pp. 509–525). Springer International Publishing.

Demchenko, Y., Belloum, A., Los, W., Wiktorski, T., Manieri, A., Brocks, H., Becker, J., Heutelbeck, D., Hemmje, M., & Brewer, S. (2016). EDISON Data Science Framework: A Foundation for Building Data Science Profession for Research and Industry. *Proceedings of the 8th IEEE International Conference on Cloud Computing Technology and Science (CloudCom)*, 620–626.

D'Ignazio, C., & Bhargava, R. (2015). Approaches to Building Big Data Literacy. *Proceedings of the Bloomberg Data for Good Exchange Conference*, 6.

Dubé, L., & Paré, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597–636.

Dubey, R., & Gunasekaran, A. (2015). Education and training for successful career in Big Data and Business Analytics. *Industrial and Commercial Training*, *47*(4), 174–181.

Gherardi, S. (2000). Practice-based Theorizing on Learning and Knowing in Organizations. *Organization*, *7*(2), 211–223.

Goel, K., Sindhgatta, R., Ilavarasan, V., & Kar, A. K. (2021). Building a knowledge society: Learnings from a digital literacy program. *Proceedings of the 29th European Conference on Information Systems (ECIS)*.

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.
Ho, S. Y., & Frampton, K. (2010). A Competency Model for the Information Technology Workforce: Implications for Training and Selection. *Communications of the Association for Information Systems*, *27*.
Jones, M. (2019). What We Talk about When We Talk about (Big) Data. *The Journal of Strategic Information Systems*, *28*(1), 3–16.
Koltay, T. (2017). Data literacy for researchers and data librarians. *Journal of Librarianship and Information Science*, *49*(1), 3–14.
Le Deist, F. D., & Winterton, J. (2005). What Is Competence? *Human Resource Development International*, *8*(1), 27–46.
Lefebvre, H., & Legner, C. (2022). How Communities of Practice Enable Data Democratization Inside the Enterprise. *Proceedings of the 30th European Conference on Information Systems (ECIS)*, 16.
Lennerholt, C., Van Laere, J., & Söderström, E. (2021). User-Related Challenges of Self-Service Business Intelligence. *Information Systems Management*, *38*(4), 309–323.
Li, L. (2022). Reskilling and Upskilling the Future-ready Workforce for Industry 4.0 and Beyond. *Information Systems Frontiers*.
McClelland, D. C. (1973). Testing for competence rather than for "intelligence." *American Psychologist*, *28*(1), 1–14.
Merchel, R., Süße, T., Iqbal, T., & Knop, S. (2021). Digital Competencies and IT Skills as Employees' Resources to Cope with Digitalization Demands. *Proceedings of the 42nd International Conference on Information Systems (ICIS)*.
Micheli, M., Ponti, M., Craglia, M., & Berti Suman, A. (2020). Emerging models of data governance in the age of datafication. *Big Data & Society*, *7*(2).
Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.
Morgan, D. L. (1993). *Successful Focus Groups: Advancing the State of the Art*. SAGE Publications.
Paparova, D. (2023). Exploring the ontological status of data: A process-oriented approach. *Proceedings of the 31st European Conference on Information Systems (ECIS)*.
Paré, G. (2004). Investigating Information Systems with Positivist Case Research. *Communications of the Association for Information Systems*, *13*(1), Paper 18.
Prifti, L. (2019). *Professional Qualification in "Industrie 4.0": Building a Competency Model and Competency-Based Curriculum*. Universitätsbibliothek der TU München.
Redman, T. C. (2022). Your Data Initiatives Can't Just Be for Data Scientists. *Harvard Business Review*. https://hbr.org/2022/03/your-data-initiatives-cant-just-be-for-data-scientists
Ridsdale, C., Rothwell, J., Smit, M., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., Wuetherick, B., & Ali-Hassan, H. (2015). *Strategies and Best Practices for Data Literacy Education Knowledge Synthesis Report*.
Saltz, J., Armour, F., & Sharda, R. (2018). Data Science Roles and the Types of Data Science Programs. *Communications of the Association for Information Systems*, *43*, 615–624.
Schüller, K. (2020). *Future Skills: A Framework for Data Literacy* (53). https://hochschulforumdigitalisierung.de
Sternkopf, H., & Mueller, R. M. (2018). Doing Good with Data: Development of a Maturity Model for Data Literacy in Non-governmental Organizations. *Proceedings of the 51st Hawaii International Conference on System Sciences*.
Vohra, S., & Morrow, J. (2020). *The Human Impact of Data Literacy*. Accenture. https://www.accenture.com/us-en/insights/technology/human-impact-data-literacy
von Konsky, B. R., Miller, C., & Jones, A. (2016). The Skills Framework for the Information Age: Engaging Stakeholders in Curriculum Design. *Journal of Information Systems Education*, *27*, 37–50.

Walker, D. F. (2003). *Fundamentals of curriculum: Passion and professionalism* (2nd ed). L. Erlbaum Associates.
Wang, Y., Gray, P. H., & Meister, D. B. (2014). Task-driven learning: The antecedents and outcomes of internal and external knowledge sourcing. *Information & Management*, *51*(8), 939–951.
Wolff, A., Gooch, D., Cavero Montaner, J. J., Rashid, U., & Kortuem, G. (2016). Creating an Understanding of Data Literacy for a Data-driven Society. *The Journal of Community Informatics*, *12*(3).
Zhu, S., Gupta, A., Paradice, D., & Cegielski, C. (2019). Understanding the Impact of Immersion and Authenticity on Satisfaction Behavior in Learning Analytics Tasks. *Information Systems Frontiers*, *21*(4), 791–814.

Essay 4

Data Monetization as a New Frontier for Data Governance

Martin Fadler, Hippolyte Lefebvre, Christine Legner

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

Published in the proceedings of the 29th European Conference on Information Systems (2021)

Extended version based on feedback received from first round review from Information & Organization journal

Abstract: *Today, most firms – digital born as well as incumbents – recognize the strategic potential that data have and search for new ways to monetize their data. Data governance is considered a success factor in the value creation process from data, as it enables the alignment between the management of data assets and business objectives. While the importance of data governance is increasingly recognized, research still does not explain how it addresses the changing role of data: Data governance has initially been seen as integral part of IT governance, and existing studies on data governance focus more on gaining control over data than on enabling value creation across the organization. Against this backdrop, a more thorough understanding is needed how data governance practices evolve in response to data's increasing business criticality and strategic importance. Based on nine case studies from multinational companies, we analyze data how governance practices evolve when companies move from defensive to offensive data strategies, resulting in three archetypes: (1) Improve master data quality, (2) Establish enterprise-wide data transparency, and (3) Enable data monetization. We further show that moving towards data monetization requires more sophisticated procedural and relational data governance practices that facilitate data-driven innovation across the organization. For practitioners, our research provides insights into the priorities of data governance initiatives and outlines pathways to manage data as a strategic asset.*

Keywords: IT Governance Mechanisms, Data Governance, Data Monetization, Data Quality, Data Value, Data-driven Innovation.

Table of contents

List of tables

1 Introduction

In the last decade, data has transformed from a byproduct of economic activity into a strategic asset, significantly influencing firm performance more than ever before (Akter et al., 2016; Wamba et al., 2017). The strategic value of data is recognized not only by digital-native enterprises but also by established firms ("incumbents"), which are now actively exploring innovative approaches to leverage data for tangible economic benefits – a trend referred to as data monetization (Jones, 2019; Wixom & Ross, 2017). However, innovation with and monetizing data remains to be a significant challenge for many companies: To achieve value from data, companies have not only to embed data into their everyday work practices, but also develop more sophisticated data management capabilities (Aaltonen et al., 2021; Legner et al., 2020) and specifically data governance. Grover et al. (2018) even argue 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*" (p. 417). Today, data governance is acknowledged as a success factor to manage the value creation process from data (Grover et al., 2018) and to use data at scale (Mikalef et al., 2018).

Despite the growing body of literature, data governance is still in need of more research (Vial, 2023). Initially, data was viewed as a fundamental component of IT systems, and by extension, IT governance, which has led to it being overlooked in research (Kohli & Grover, 2008; Tiwana et al., 2013). Beginning in the 1990s, a nascent body of literature began to advocate that analytical systems, such as data warehouses (Rifaie et al., 2009; Watson et al., 2004), as well as operational systems need dedicated data governance, thereby initiating research on master data and data quality management (Khatri and Brown, 2010; Otto, 2011a, 2011b; Weber et al., 2009). Consistent with the established paradigms in corporate governance and IT governance literature, researchers (Abraham et al., 2019; Tallon et al., 2013; Vial, 2023) have conceptualized data governance as set of structural, procedural, and relational governance mechanisms. While the foundations of data governance are increasingly clear, the existing body of knowledge still emphasizes the control of data assets through the formalization of standards and data policies that enforce data quality and compliance (Abraham et al., 2019; Chua et al., 2022), and through clarifying the accountabilities for data along the data lifecycle (Tallon et al., 2013). This narrow perspective overlooks the dual mandate of data governance, which aims not just at ensuring control but also at fostering innovation (Vial, 2023). As data is increasingly used to innovate, data governance has to evolve to support day-to-day data production, use, and reuse by a growing number of employees (Aaltonen et al., 2021; Alaimo & Kallinikos, 2022; Benfeldt et al., 2020; Legner et al., 2020).

Against this backdrop, our research aims to provide a more thorough understanding on how data governance unfolds when data's role changes in enterprises and companies move towards data monetization. Hence, we ask the following research question:

> *How do companies develop their data governance practices to address the changing role of data?*

To address this research questions, we employ multiple case studies (Dubé & Paré, 2003) which are frequently used in data governance research (Otto, 2011c; Parmiggiani & Grisot, 2020; Tallon et al., 2013). They facilitate an intricate examination of data governance and the nuances of data governance practices in a broader organizational context. In our study, we examined data governance setups and practices across a diverse set of nine multinational companies which vary in terms of their industry, data strategy, data scope, and experience with data governance. First, our analysis reveals that as the role of data evolves and organizations increasingly focus on data monetization, there is a concurrent expansion of their data governance frameworks and associated data governance practices. Whereas data governance has traditionally focused on defining accountabilities, data standards and policies, we find that data monetization involves building relational governance practices that foster data literacy and data sharing in an extended network as well as extending structural practices for strategic decision-making and investment planning. Second, we also derive three archetypes, each mirroring the evolving role of data within organizations and characterized by a set of specific data governance practices: (1) *Improve master data quality,* (2) *Establish enterprise-wide data transparency* and (3) *Enable data monetization.*

From an academic standpoint, our findings extend the existing data governance research by providing empirical insights how the structural, procedural, and relational mechanisms are enacted by specific data governance practices, and implement through sub-practices. Thereby, our study provides empirical evidence reaffirm the emancipation of data governance from IT governance to achieve data monetization. For instance, while previous literature has recommended situating data governance within the IT organization as the "preferable" approach (Tallon et al., 2013, p.169), our research reveals that enterprises are adopting structural data governance practices distinct from IT. We further show how data governance practices and sub-practices evolve in the form of archetypes as data role changes in enterprise, thereby responding to recent calls for research for data governance as a dynamic configuration of practices (Vial, 2023). Specifically, our findings advocate extending procedural data governance practices supporting prioritization, investments and implementation of data use cases required for data monetization at scale. These are accompanied with enhanced relational practices, which are crucial for ensuring effective coordination and alignment within the enlarging data network that is charged with delivering innovation.

For practitioners, we provide insights into the priorities established by data governance initiatives, the governance mechanisms employed, and the concrete practices to be defined and implemented. Consequently, our research delineates clear pathways that assist in managing data as a strategic asset, guiding organizations in leveraging their data for competitive advantage.

The remainder of this paper is structured as follows. Firstly, we analyze the changing role of data and implications for data governance and review IT and data governance research along structural, procedural and relational governance mechanisms. Secondly, we describe our multiple-case study research approach. Thirdly, we present the findings from our cross-case analysis and identify typical archetypes for data governance designs that accommodate the changing role of data. Lastly, we summarize the contributions and discuss the implications of our research.

2 Background

Data's increasing business criticality and strategic importance implies that data governance approaches need to evolve to support the overall organization's goals (Mikalef et al., 2018; Vial, 2023). In this section, we review how the role of data in enterprises has changed over time, resulting in an increasing awareness and need for managing data at the enterprise-wide level. So far, research has conceptualized data governance as a combination of structural, procedural and relational governance mechanisms that are instantiated by governance practices, but the suggested data governance practices focus mostly on controlling critical data resources along their life-cycle rather than on maximizing value creation from data assets.

2.1 Data's Changing Role and Implications for Data Governance

Data and information are an integral component of IT artifacts and serve as the foundational elements that drive system functionality, automate business processes, and enable the delivery of value to users (Chua et al., 2022). As a result, data management has been naturally embedded within IT management and governance (Kohli & Grover, 2008; Tiwana et al., 2013). Stimulated by technological innovations and changing business requirements, data's role in enterprises has significantly evolved over the past decades (Alaimo & Kallinikos, 2022; Chua et al., 2022; Legner et al., 2020), leading first to the emergence of data governance as sub-discipline of IT governance, then to a dedicated discipline and now casts for a rethinking to account for the strategic role of data. Table 29 displays data's evolving role in enterprises along three phases with implications for data governance.

In the first phase (1980s), data were used primarily for automated data processing in specific business functions, e.g., financial accounting or inventory management. Data resided in isolated databases, and responsibilities for data were limited to database design and administration. Goodhue et al. (1988) were among the first to address the problem of "*unmanaged*" data and drew the attention to data model quality and data reuse beyond a single database.

In the second phase (1990s-2000s), integrated operational and analytical systems started to appear, resulting in an imperative to enhance data availability and integrate data across the entire enterprise. This transition prompted companies to consider data as "*subsumed under organizational resources*" (Chua et al., 2022, p. 5) that enable and improve enterprise-wide business processes and decision-making. Consequently, data governance emerged within IT governance' scope and primarily revolved around the specific contexts of Enterprise Resource

Planning (ERPs) and data warehouses (Rifaie et al., 2009; Watson et al., 2004). Similar to other firm's resources, the quality of data resources, specifically master data , became pivotal for value creation, particularly highlighted by Wang (1998)'s seminal work on Total Data Quality Management. To improve data quality across the enterprise (Ballou et al., 1998), academic studies draw attention to relevant organizational (Khatri & Brown, 2010) and technical capabilities (e.g., enterprise-wide data integration and architecture). The common denominator of these studies is a management-oriented perspective on data, with data governance defining accountabilities, standards and policies in order to control and manage critical data resources.

	Phase 1 (1980s): Database administration	**Phase 2 (1990s-2000s): Data as enterprise resource**	**Phase 3 (since 2010s): Data as strategic asset**
Roles of data	Data as integral part of IT systems and as an enabler of automation in business functions	Data as valuable enterprise resource that enables enterprise-wide business processes and decision-making	Data as strategic asset that can be monetized directly and indirectly
Focus of data governance	Not existing (increasing awareness for data model quality and data reuse in the systems development process)	Control of critical data resources (master data), their quality and compliance	Control and coordinate data value creation in a growing portfolio of operational and analytical use cases
Relevant literature	(Goodhue et al., 1988; Grover & Teng, 1991; Jain et al., 1998)	(Khatri & Brown, 2010; Otto, 2011c; Tallon et al., 2013)	(Alaimo & Kallinikos, 2022; Benfeldt et al., 2020; Grover et al., 2018; Gupta & George, 2016; Mikalef et al., 2018)

Table 29. Data's evolving role in enterprises and implications for data governance (based on Legner et al. (2020))

In today's third phase (since 2010s) that experienced the surge of big data and analytics, data is considered a strategic enterprise resource that can help companies gain a long-term competitive advantage (Grover et al. 2018; Gupta and George 2016; Mikalef et al. 2018). Accordingly, organizations are creating, collecting and curating increasingly large amounts of data from internal and external sources (for instance, user-generated content in social media) with the hope to monetize them. The term data monetization (Mehta et al., 2021; Wixom & Ross, 2017) has been coined to describe the different ways to create quantifiable benefits from data, either directly or indirectly, through improved and automated decision making (Wixom & Ross, 2017), by developing data-driven business models, and by selling data to third parties. Value creation, and thereby data monetization, is achieved by *making sense* of data to support business outcomes (Aaltonen et al., 2021; Lycett, 2013) and requires companies to develop data practices in all parts of the organization and beyond just data experts (Benfeldt et al., 2020; Lycett, 2013). Therefore, in line with the view of data as company assets (Benfeldt et al., 2020; Otto, 2011c), data governance must broaden its perspective beyond control and compliance and support data-driven innovation (Mikalef et al., 2020; Vial, 2023). This implies a shift in data governance from operational aspects to governance as strategic instrument emancipated from IT, aimed at ensuring the continuous relevance and alignment of firms' activities toward better performance (Hoetker & Mellewigt, 2009; Lavie et al., 2012).

2.2 Governance Mechanisms and Practices

IT and data governance literature rely on a set of generally applicable governance mechanisms from corporate governance literature (Lavie et al., 2012; Poppo & Zenger, 2002; Tihanyi et al., 2014) which describe them as universally applicable for enterprise assets: structural mechanisms define the hierarchical structure and assign responsibilities; procedural mechanisms define and structure decision-making processes; and relational mechanisms describe communication, knowledge sharing, alignment, and collaboration. Accordingly, the prevailing understanding sees IT governance *"as the decision rights and accountability framework deployed through a mix of structural, processual, and relational mechanisms and used to ensure the alignment of IT-related activities with the organization's strategy and objectives"* (Gregory et al., 2018, p. 1227). The distinction of structural, procedural and relational governance mechanisms has also been picked up by data governance research, specifically by two studies (Abraham et al., 2019; Tallon et al., 2013) that have made an attempt to develop a holistic data governance framework and a recent study on analytics governance (Baijens et al., 2021). Structural and procedural governance mechanisms are often tangible and implemented in a top-down manner, whereas relational governance mechanisms are usually intangible and tacit as they are "voluntary" actions and cannot be programmed (Vial, 2023).

While governance mechanisms are universally applicable, they have to be enacted by specific practices (Alhassan et al., 2016). Thus, to synthesize data governance research comprehensively, Table 30 correlates the three governance mechanisms with specific practices that operationalize them. The comparison of governance mechanisms and practices found for IT artifacts with those identified for data/information and analytics reveals distinct disparities, highlighting the narrow scope of many data governance studies which are anchored in the perspective of "data as enterprise resource", with strong focus on data quality and compliance.

Related literature	Governance mechanisms and related practices		
	Structural	Procedural	Relational
IT GOVERNANCE			
(Sambamurthy and Zmud, 1999)	Assignment of decision rights, contingency factors		
(Karimi et al., 2000)	Decision making in steering committees and boards		
(Sambamurthy and Zmud, 2000)		Documentation of IT activities	
(Weill and Ross, 2004)	IT decision domains and assignment		
(De Haes and Van Grembergen, 2004)	Committees and council	Monitoring	Learning
(Peterson, 2004)	Creation of formal positions and roles	Strategic IT decision making	Business-IT partnerships
(Xue et al., 2008)		IT investment management	
(Huang et al., 2010)			Communication policy and steering committee
(Wu et al., 2015)			Strategic alignment
DATA GOVERNANCE			
(Weber et al., 2009)	Roles and assignment of decision rights for data quality management	Strategic tasks for data quality management	
(Khatri and Brown, 2010)	Data governance decision domains and assignment		
(Otto, 2011c)	Organizational dimensions: goals, form, and transformation		
(Velu et al., 2013)	Allocation of decision rights for data management		
(Tallon et al., 2013)	Data ownership, rights and responsibilities; User involvement in policy setting; Shared oversight of policy setting, monitoring, and revision	Access monitoring; Backup practices; Retention policies; Information protection; Costs monitoring and chargebacks; Data migration	User education; Communications of needs and results
(Abraham et al., 2019)	Roles and responsibilities; Location of decision-making authority	Data strategy; Policies, standards, processes, procedures; Contractual agreements; Performance agreements; Compliance monitoring; Issue management	Communication; Training; Coordination of decision-making
(Fadler & Legner, 2021b)	Definition and assignment of data and analytics roles and responsibilities		
(Fadler & Legner, 2020, 2021a)	Data ownership types		
(Vial, 2023)	Policy-setting procedures; Oversight mechanisms; Data ownership responsibilities	Enforce retention/archiving; Apply backups practices; Establish and monitor access; Classify information by value; Service levels for data protection; Monitor costs via chargebacks; Migration between storage tiers	User education; Communications/idea exchange
ANALYTICS GOVERNANCE			
(Baijens et al., 2020, 2021)	Organizational structure, roles and responsibilities, coordination and alignment	Process model, monitoring and evaluation of analytics projects, development roadmap	Shared perceptions, collaboration, transfer of know-how

Table 30. Governance objects, mechanisms and practices in prior IS literature

2.2.1 Structural Mechanisms

Structural governance mechanisms take *"the shape of formal positions and (integrator) roles, and/or formal groups and (management) team arrangements"* (Peterson, 2004, p.14). They specify the organization's hierarchy, positions and roles and define their responsibilities for decision making. As part of structural governance, a company first defines which decisions have to be made, before assigning the responsibilities for decision-making.

For IT and information, Weill and Ross (2004) define IT principles, IT architecture, IT infrastructure, business application needs, IT investment, and prioritization as decision domains. When it comes to data governance, Khatri and Brown (2010)'s seminal paper outlines decision domains that have been picked up by many follow-up studies and include data principles, data quality, metadata, data access, and data lifecycle. Interestingly, neither business needs, nor investment and prioritization that are prominent decision domains in IT governance literature, are present in this list. While decision domains are typically quite abstract, several scholars have investigated decision rights for data and their assignment on a more granular level. For instance, Winkler and Wessel (2018) analyze different decision right classes and distinguish between decision right input, control, and management rights.

Once the decision domains are identified, an enterprise must define who is responsible for decision making. For IT governance, Sambamurthy and Zmud (1999) distinguish central, decentral, and federated decision making according to the location of the decision authority. Weill and Ross (2004) go one step further and derive typical archetypes for this assignment, for instance, "business monarchy" and "IT monarchy" for central decision making, and "feudal system " for business-lead decision making. Centralized IS decision making allows for company-wide control, efficiency and reliability in the utilization of IT assets, but decreases the local units' flexibility, agility, and innovation potency (Gregory et al., 2018; Huang et al., 2010). On the other hand, a complete decentralization of IT decision making has the opposite effect and comes with risks of misalignment with corporate strategy and redundant efforts as well as inefficiencies due to lack of standardization. Recent studies underline that data must be governed in a different way than IT because business organizations are data creators and consumers; therefore, accountability for data should never be centralized to ensure value creation (Vial, 2023) and the allocation is a function of the uncertainty in and similarity between business units (Velu et al., 2013).

Concrete roles and responsibilities have been a focus topic of data governance research for more than a decade (Otto, 2011c). For instance, Weber et al. (2009) define the typical data roles needed for managing data quality. Besides the strategic roles, such as the executive sponsor or chief data steward, they also include operational roles, such as the business data steward or technical data steward. As the overarching authority, a data quality board *"defines the data governance framework for the whole enterprise and controls its implementation"* (Weber et al., 2009). With big data and analytics becoming strategic value drivers, however, companies must incorporate additional roles and responsibilities, especially for the analytical use context (Fadler and Legner, 2021; Grover et al., 2018). Notably, companies adapt their fundamental decision control rights and distinguish three data ownership types: the data owner, the data platform owner, and the analytics product owner. Lee et al. (2014) also argument for a Chief Data Officer role that fosters alignment with business and IT stakeholders on a strategic level and provides the overarching direction for organizing, analyzing, and deploying an organization's data assets (Dallemulle & Davenport, 2017).

In additional to roles, steering and operational committees are commonly seen as an effective governance mechanism in IT governance (Huang et al., 2010; Karimi et al., 2000). These committees *"alig[n] IT-related decisions and actions with an organization's strategic and operational* priorities" (Huang et al., 2010). This view is supported for data governance: Weber et al. (2009) suggest an enterprise-wide data quality board which defines the data governance framework and controls its implementation, and Tallon et al. (2013) emphasize the need of shared oversight for information governance policy setting, monitoring, and revision. However, none of them takes a more strategic stance and mentions a steering committee's role for aligning data activities with the organization's strategic and operational priorities.

2.2.2 Procedural Mechanisms

Focusing only on structural mechanisms would ignore the activities and processes taking place inside an organization's established structures (Sambamurthy and Zmud, 2000). Consequently, procedural mechanisms complement organizational structures and roles in defining how decisions are made. In IT governance literature, procedural or process governance mechanisms are defined as *"the formalization and institutionalization of strategic IT decision making or IT monitoring procedures"* (Peterson, 2004, p. 15) and ensure that the IT policies meet business requirements (De Haes and Van Grembergen, 2004). Peterson (2004) synthesizes three essential IT governance processes that align strategic IT investment decisions with company goals: *"(a) the identification and formulation of the business case and/or business rationale for IT decisions; (b) the prioritization, justification, and authorization of IT investment decisions; and (c) the monitoring and evaluation of IT decision implementation and IT performance"* (p. 15). These processes ensure the (administrative, sequential, reciprocal, or full) integration of business and

IT decisions (Peterson, 2004) and strike the balance between centralization and decentralization (Gregory et al., 2018).

Existing data and analytics governance literature mentions procedural mechanisms, but mainly relates them to operational aspects along the data life-cycle rather than strategic aspects. For instance, one of the CIOs who participated in the study conducted by Tallon et al. (2013) argued that *"procedural practices permit a greater understanding of the changing value of information and how this value needs to be matched with the characteristics of different storage systems that will maximize and protect that value"* (p. 163). For analytics, the procedural mechanisms typically comprise methods that guide analytics experts to successfully execute analytics projects, such as methodologies like CRISP, or an agile development framework (Baijens et al., 2020, 2021).

2.2.3 Relational Mechanisms

While structural and procedural mechanisms define which, by whom, and how decisions should be made, relational mechanisms facilitate communication, coordination, and a shared understanding between business and IT stakeholders (Gregory et al., 2018). Thus, relational governance mechanisms are *"the active participation of, and collaborative relationships among, corporate executives, IT management, and business management"* (Peterson, 2004, p.15). In IT governance literature, these mechanisms focus on the specific horizontal link between IT and business departments. IT units must establish their communication channels to disseminate IT governance policies, roles, guidelines, and procedures (Huang et al., 2010; Wu et al., 2015). Having the appropriate communication channels in place helps companies create shared mental models, facilitate collaboration, and enhance alignment. Collaboration and alignment are achieved through direct stakeholder participation, business–IS partnerships, or colocation (De Haes and Van Grembergen, 2004). Prasad et al. (2012) emphasizes that collaborative structures can also be built by using tools (e.g., Wiki). Relational mechanisms also include communication and shared learning (Wu et al., 2015). To put this mechanism in place, an IS organization should provide training to educate professionals and establish a shared language (De Haes and Van Grembergen, 2004).

We find that data- and analytics-related research has not investigated relational mechanism in detail, although alignment and collaboration on strategic and operational levels have been emphasized as important drivers of value generated by investing in big data and analytics (Grover et al., 2018). Tallon et al. (2013) highlighted that relational practices lead users to re-orient their perception of storage as a cheap and infinite resource and, instead, regard it as a finite and costly resource (p. 165).

2.2 Research Gap

The existing body of data governance research has for a long time viewed data and information as supporting enterprise resource, with a strong focus on operational concerns. Even the comprehensive study by Tallon et al. (2013) concentrates on data governance practices *"that span all the stages of the information life cycle from the point of data creation through data destruction"* (p. 162) and considers them a responsibility of IT organizations. Moreover, recent studies that examine data (information) and analytics as dedicated objects of governance (Baijens et al., 2021; Grover et al., 2018; Tiwana et al., 2013) focus mostly on controlling critical data resources. However, data's increasing business criticality and strategic importance implies that data governance approaches need to evolve to support the overall organization's goals (Mikalef et al., 2018; Vial, 2023). This extension of data governance mandate has long been discussed in the context of value creation from investments into big data and analytics (Abbasi et al., 2016; Goes, 2014; Grover et al., 2018; Hassan, 2019; Phillips-Wren et al., 2015; Tihanyi et al., 2014). In order to enable companies monetizing their data, data governance research must thus broaden its perspective beyond the existing focus on control and compliance and extend its mandate to also support data-driven innovation in an extended network of data creators and users (Vial, 2023).

3 Methodology

To answer our research questions (i.e., *How do companies develop their data governance practices to address the changing role of data?)*, we follow a case study approach that allows us to investigate the particular phenomenon in a natural context (Paré, 2004). Case studies are widely used in data governance research as they provide insights into data governance as a complex enterprise endeavor (Parmiggiani & Grisot, 2020; Tallon et al., 2013). In our study, they allow us to study the concrete data governance practices that have been implemented and relate them to mechanisms, while also analyzing how the strategic context that shapes them. As we seek for understanding how data governance practices evolve in response to the evolving role of data, we opted for multiple case studies which are more likely than single case studies to lead to robust theories and generalizable results (Miles et al., 2014).

3.1 Case Selection

Our study is integrated with a multi-year research program which follows the collaborative practice research tradition (Mathiassen, 2002) and aims to enhance data and analytics management approaches at large corporations. Thanks to this program, we have trusted relationships with data experts from more than 20 companies and privileged access to historical and present documentation and regularly exchange on their data strategies and data governance initiatives. For this study, we used theoretical sampling (Eisenhardt and Graebner, 2007) to identify and select companies that have diverse characteristics in terms of their industry and strategic contexts, as well as the selected data scope and experience with data governance. The resulting case base comprises nine enterprises that have different levels of maturity in data governance, as illustrated by the number of roles and data domains that they focus on (see Table 31). Through the variation in our sample, we can analyze differences and commonalities in data governance practices and extract generalizable patterns (Dubé and Paré, 2003).

3.2 Data Collection

For gaining a deep understanding of the strategic context and motivation for data governance as well as the chosen setup and practices, we collected primary data through two rounds of semi-structured interviews. At each firm, we selected key informants who have a mandate for enterprise-wide data governance at their organization, typically as head of enterprise data & analytics or data management. In addition, we made sure that these key informants have worked for a longer period in the company and know the history of data governance initiatives and the issues and challenges that come with implementing data governance (e.g., the challenges that come with involving business stakeholders or assigning roles and responsibilities). With each key informant, we conducted a first semi-structured interview of 1.5 hours between September and October 2020 to discuss the strategic context and scope of data and analytics activities as well as the current status and target state for data governance. We followed an interview protocol that covers the three generic data governance mechanisms, as analytical framework, and collected information about the concrete governance practices that implement them (see Appendix, Table 34). During the interview, we also discussed whether specific practices from literature and developed in previous research activities in the research program were relevant for the company. To review the progress and improve our insights into the governance practices, we conducted a second interview with the key informants between August and November 2021. Whenever necessary, we scheduled follow-up calls for further clarifications.

We used Microsoft Teams to conduct and record the interviews. We complemented the interviews with an analysis of additional documents that we had gathered during our research activities (e.g., on the company's business and data strategy, data roles and responsibilities, or relevant processes) and publicly available information (i.e., news articles, financial reports and presentations at conferences). Through this combination of primary and secondary sources, we could triangulate the gathered information and ensure construct validity (Yin, 2003). After the interview, a write-up comprising key statements and links to company material were sent to the interviewees to confirm the statements' correctness, clarify misunderstandings, and answer open questions. An overview with details on each company can be found in Table 31

Case	Industry	Revenue / Employees	Key informants	Data strategy focus	Data scope	Data governance approach
A	Public transportation	\$1B–\$50B / ~35 000	Product Owner Data Strategy; Enterprise Architect for Data & Analytics; Responsible for Analytics Strategy	DM strategy focused on data governance in place since 2017. A new version planned for 2021 will focus on data quality, roles, and decentralized responsibilities. Analytics strategy for 2021 focuses on further decentralizing analytics initiatives. D&A skillset development is also a core part of the company's digital transformation.	Five data domains with varying maturity: Assets, Business Partners, Production Plans, Product, and Material. Complex datasets spanning business processes/divisions assigned to data managers. Focus on master and transactional data. SAP transformation: R/3 to S/4HANA.	Central DM is a support function with four FTEs and three non-FTEs and is responsible for strategy, methods, and governance. Decentralized teams have 17 FTEs and are organized by "clusters" in business units with data owners. The central analytics team is part of IT with 40 FTEs. Analytics is also decentralized in the IT of business units (100+ FTEs).
B	Manufacturing, chemicals	\$1B–\$50B / ~5 000	Head of Corporate Data Management	DM strategy drafted in 2020 with a focus on data foundation and aligned with the group digital strategy. To be spread over all areas of the organization. 2021 integrated data management and analytics strategy addressing value creation with use cases, target architecture (e.g., establisbh a data mesh, and a data "lakehouse"), new central and decentral organizational model.	Four data domains: Product, Customer/Account, Material, and Vendor/Supplier. Historically focused on master and reference data, now also including external data. Business Partner data (90% are customer data) are emerging. SAP MDG-S implemented for 150 users globally. SAP MDG being implemented.	Central DM is a support function with seven FTEs (Head of DM, data quality manager, data engineer, three data stewards, data architect in IT). It defines methods and guidelines, data models, oversees DQ initiatives, and supports business/IT projects with data know-how. Eight decentralized data experts (e.g., data coordinator) in business functions (non-FTEs).
C	Packaging, food processing	\$1B–\$50B / ~25 000	Director of Global Master Data Strategy; Director for Business Information Management	The company's 2030 strategy will drive D&A initiatives with the goal of monetizing data. The firm's strategic program integrates all data-related strategies since 2019: MDM, BI, Marketing, and Engineering. The first MDM strategy dates to 2005 and the BI strategy to 2009. Developing a corporate data culture is core to the data strategy.	Six data domains: People, Customer, Supplier, Finance, Products/Material, Brand/Category. Self-service exists in BI and AI, with SAP BW and Alteryx. Currently engaging SAP transformation from R/3 to S/4HANA. Strategic program data scope: master, transactional, purchase, machine.	Central data governance team (six FTEs) with decentralized leadership (22 non-FTEs) and business experts (100+ non-FTEs). Two central services for MDM and material data maintenance (total of 32 FTEs). Central BI team operates in IT. BI coordinators and the network of BI experts are decentralized in regions and by process. Domains are assigned ownership, standards, and a model.
D	Manufacturing, automotive	\$1B–\$50B / ~90 000	Vice-President Data and Analytics Governance; Data and Analytics Governance Manager	DM strategy since 2018 and MDM since 2016. DM and analytics will be integrated into the IT and digitalization strategy in 2021. The current analytics strategy is focused more on IT.	Forty-seven data domains with all data types, either established or emerging. Data domains are structured by data objects. D&A is spread across business functions, divisions, and regions.	Central D&A governance agile team (10 FTEs), no role model. Decentralized D&A in domains (FTEs: 40 data domain managers, eight KPI managers, 15 advanced analytics managers; non-FTEs: 200 data coordinators)

E	Consumer goods	\$50B–\$100B / ~350 000	Master Data Lead; Product Group Manager Data and Analytics Products & Services	MDM strategy revised in 2015 to expand the scope of the data. An integrated data management and analytics strategy has been released in 2021. It focuses on the funding and details the commitments with regards to expected value created by data and analytics.	More than 10 functional, business and master data domains e.g., Customer, Vendor, Product, Material, Financial, Sales and marketing Employee, Procurement. Master data is well established, and other data types (e.g., internal) are emerging.	Central data governance and methods (15 FTEs) in IT. Central analytics in IT without a role model. A network of data standard owners in business functions (200 non-FTEs). Seven shared services for master data operations (100 FTEs).
F	Manufacturing, automotive	\$1B–\$50B / ~150 000	Head of Master Data Management; Corporate Head of Market Master Data Management; Head of Advanced Analytics, Self-Service Analytics; Member of the data enablement team	The first draft of "Data enablement strategy" presented to management in March 2020. It will focus on operational excellence and digital transformation: creating data capabilities and analytics capabilities from business capabilities/use cases. Current focus is operational excellence and establish a data catalog.	Nine data domains: HR, Market, Finance, Quality, Purchasing, Supply Chain, Development, Production, Business Partners. Group data classes in domains with high governance. SAP family tree for MD domains. Most data types are already on the data lake except media data that is emerging.	The central team in IT called "data and insights analytics" has four pillars: MDM (17 FTEs), classical BI, finance reporting and advanced analytics. Decentralized data stewardship in domains for DQ and demand. Decentralized reporting in other IT departments. Dedicated data enablement team (7 FTEs).
G	Pharmaceutical	\$1B–\$50B / ~70 000	Global Data Lead-Enterprise Solution; Associate Director Supply Chain Management; Enterprise Solutions Architect Analytics Lead	Data strategy is not defined, but data and analytics are separate pillars of the overall digital transformation initiative to be launched in 2021 and are addressed as two separate enablers.	Two data domains: Material and Account. Governance is established only over Material master data. Secured sponsorship from a VP to extend the scope.	Central data team (10 FTEs) with data support and maintenance decentralized in the regions (55 FTEs, including a special team for DQ). Central analytics team (six FTEs) attached to supply chain.
H	Consumer goods, retail	\$1B–\$50B / ~30 000	Vice-President: Data and Analytics	Data governance strategy since 2019. BI strategy since 2015. Integrated enterprise-wide D&A strategy in progress, with a release planned for 2021. A data governance framework is currently being rolled out.	Twenty-six data domains and 100 sub-data domains defined by business objects (and functions). All the data-related terms have a glossary. Master, transactional and reference data are established.	Central D&A team of more than 20 FTEs (three for governance) reporting to controlling, while data science reports to strategy. The decentralized data organization in business functions has 15 data stewards (equivalent three FTEs).
I	Consumer goods, retail	\$100B–\$150B / ~450 000	Head of Enterprise Data Management	Data scope extended from MDM to DM through the "Enterprise Architecture and data strategy" (released in 2020) is synchronized with IT strategy and is an enabler of the enterprise-wide digitization strategy.	Six data domains: Article, Vendor Customer, Material, Financial, and Employee. Master data are well established. Transactional, behavioral, and classic analytical data are not fully covered by DM.	Central data management organization (60+ FTEs) working mainly on master data. Decentralized data organization in the branches/divisions (about 100 FTEs) and also by retail countries with country managers (30 FTEs). Shared services for article master data.

Table 31. Case companies

3.3 Within and Cross-Case Analysis

Our within-analysis focused on developing an understanding of data governance practices and how they are instantiated at each company. To conduct this theory extension endeavor, we applied abductive reasoning because it allows for embedding empirical findings into an existing theoretical model (Ketokivi & Mantere, 2010). This approach facilitated theorization through a detailed examination of the data by employing deductive coding with a set of deductive theoretically-derived categories (data governance mechanisms) and sub-categories (data governance practices), as displayed in Table 30). We subsequently generated inductive *codes* from the interview data to reveal empirical findings that were not sufficiently explained by the theoretical categories. To this end, we labeled all the interview statements with first-order codes representing instantiations of data governance practices (e.g., educational programs, data quality monitoring), and aggregated them in second order codes aligned with the level of abstraction applied to data governance practices in literature. This step involved a critical review by two researchers to assess whether the emerging codes echoed established IT governance practices or unveiled novel practices particular to data governance. As a result, new inductive sub-categories (data governance practices) were identified, such as for instance "investment management" which was not part of the theoretical framework. Our coding, highlighting the extension of the theoretical framework through inductive coding, resulted in a thematic relationship between data governance mechanisms, data governance practices, and sub-practices.

In the second step, we conducted a cross-case analysis in the form of a comparative analysis of the five cases. Performing a cross-case analysis which is particularly relevant to this study as it supports the aggregation, simplification and generalization from complex cases (Miles et al., 2014). We searched for differences and commonalities between cases by iteratively searching for similarities between codes. We were able to generalize a set of 7 data governance practices and 15 sub-practices by reviewing common codes necessary to describe each of the three data governance mechanisms (see Table 32). Moreover, to better understand the evolution of data governance within varying strategic contexts and data scopes, we searched for patterns in the implementation of structural, procedural, and relational governance practices. We also analyzed the strategic context to accurately categorize the cases.

From this analysis, we could generalize a more comprehensive understanding of data governance as configuration of practices that support the evolving significance of data. To validate our findings, we discussed the governance practices and archetypes that we identified –

firstly, in a focus group meeting with all interviewees, and secondly, in a focus group including data governance experts other than the interviewees. In both focus groups, the participants confirmed the archetypes – in other words, they could position and relate their data governance approach to one of the identified data governance archetypes. In addition, they found the data governance practices and archetypes very helpful to articulate their organization's strategy and governance requirements.

4 Data Governance Mechanisms and Practices

Our sample includes companies that employ varied data strategies, highlighting not only a shift in the scope of data but also in its application. This strategic evolution, with data monetization at the forefront, significantly influences the methods companies employ to implement structural, procedural, and relational governance mechanisms. In fact, our findings confirm that companies develop additional data governance practices to accommodate data's changing role, thereby transcending beyond the operational focus that is put forward in prior literature. They complement their structural data governance practices with procedural and relational practices to better control and coordinate the strategic value creation process from their data. Table 32 catalogs the data governance practices that not only consolidate established research on data governance but also augment it with specific practices previously unexplored in the literature.

4.1 Structural data governance practices

First, and in accordance with existing literature, all companies make a fundamental data governance model decision by choosing between a centralized, decentralized, or federated data organization, and spread decision domains accordingly. This choice is reflected by the practice *Shape the operating model*, and we can identify different sub-practices. With data taking a more significant in their enterprises, all case companies grow their data teams and adapt their operating model. They move from a more centralized to a federated data organization, where the central data team works with decentralized roles or teams in business functions. Only the companies that mainly focus on master data – here B, G, H – have a pre-dominantly central data organization and case company G has only decentral data teams in their business functions.

Assign data roles and responsibilities complements this first data governance practice. In line with previous studies, all case companies have assigned data roles on different organizational levels (i.e., data stewards, data architects, data quality managers, data documentation managers, data editors) and, when chosen a federated organizational structure, also to business functions (i.e., data owners). The cases show that companies increasingly manage data and analytics in an integrated fashion. Therefore, case companies A, C, D, H complement data management roles with analytics roles (i.e., analytics product owners, analytics product lifecycle owners, data scientists, data engineers, analytics product architects, data platform owners, analytics experts) to steer their analytics initiatives enterprise wide. Roles and responsibilities are also assigned to board and committees which play a key role into decision-making, for instance when approving enterprise-wide guidelines.

4.2 Procedural data governance practices

Companies set up specific processes to make strategic, governance, and operational decisions in a structured way. For this, they need to implement three procedural data governance practices: *Define and monitor data strategy and investments; Define and enforce data principles;* and *Define and enforce data principles*.

Although recognized in IT governance literature, *Define and monitor data strategy and investments*. has been newly identified in our analysis as an overlooked aspect of data governance research. All enterprises in our case set have a structured means to manage investments in data-related aspects (Investment management*)*, which was also highlighted as an IT governance practice in literature. However, processes to plan and control the implementation of the data strategy *(*Planning and control*)* and proactively identify business cases *(*Business case identification*)* are new sub-practices that reflect the growth of data use for strategic value creation.

The second one is *Define and enforce data principles*. All case companies establish processes to make decisions about *Data standards and guidelines* and *Data models and architecture*. While the former process handles the standards, rules, and formal procedures, the latter process comprises activities to technically implement the business requirements into the databases. These processes have been identified in prior literature. Weber et al. (2009) outline activities for data quality and master data management on an organizational and information systems level in accordance to the two distinct process mechanisms found in our case set.

The third one is *Manage data operations*. All case companies establish processes for *Data quality monitoring and support* and *Data lifecycle management*. On the one side, they proactively manage the lifecycle steps from data creation towards deletion to create transparency and control data flow across systems. On the other side, they monitor data quality and provide support to ensure correctness of data across the enterprise.

4.3 Relational data governance practices

All case companies have implemented relational governance mechanisms to align with key stakeholders and coordinate data activities across the growing data network. We distinguish two new data governance practices: *Align and collaborate with business and IT stakeholders;* and *Develop and Share data knowledge*. Although they have been briefly mentioned into literature (e.g., business IT partnership, educational programs), our findings significantly extend their scope to account for the data network's growth.

The data governance practice *Alignment and collaboration with business and IT stakeholders* reflect how pervasive the data organization becomes as the role of data changes. In fact, our findings show that the network of data roles extend beyond the core data teams and reach into business (e.g., data owners) and IT (e.g., data architects). Therefore, further coordination (e.g., through data councils) and communication (e.g., newsletters, project updates) are needed to strengthen collaboration toward strategy execution.

In addition, companies seek to *Develop and Share data knowledge*. As described in the literature, they typically use educational programs to increase the data literacy of the business professionals. They also rely on communities of practice (Wenger, 2000) that stimulate practice-based learning though practice exchange between data roles (e.g., challenges and best practices, shared understanding of solutions, peer coaching, workplace relatable data applications). Interestingly, while forming Communities of Practice seems to be a common practice among the case companies, it is notably absent from both IT and data governance literature, having only recently been acknowledged by contemporary research on data democratization within enterprises (Lefebvre & Legner, 2022).

.

Data Governance Mechanisms & Practices		Sub-practices	Examples of Supporting Statements (Case)
Structural	Shape the operating model	Data organization (central, decentral, federated)	*"The Central data team has 11 FTE working on governance. We then have data standard owner spread across domains, located in HQ, and mainly in functions."* **(E)** *"Our organization is located in IT digitalization department and includes data governance, IT governance, enterprise architecture."* **(D)**
		Allocation of decision rights	*"The global team ensures governance and data quality."* **(G)** *"While the strategy is developed centrally, the data lifecycle is a combination of central and decentral activities."* **(C)**
	Assign data roles & responsibilities	Steering committees and boards	*"The MDM board includes several decision makers from the division, on a 4 to 6 weeks, more like 8 weeks in reality."* **(I)** *"The data governance board is discontinued. Now it is a data and analytics board taking the sponsorship & ownership. For data domains decisions board, the central team meets with data definition owners. Both boards happen quarterly or biannually."* **(H)**
		Data roles	*"We have corporate data owners dealing with data content and structure per domain."* **(F)** *"The decentral data team is comprised mostly of global master data drivers who set business rules on a field level, for the various master data objects. There are also various leaders from a few groups who carry out extensive data stewardship activities."* **(C)**
Procedural	Define and monitor data strategy and investments*	Planning & control*	*"Domain managers should have a strategic plan for their domain. Only few domains with strategy (e.g., finance)."* **(D)** "For the domains that are part of the global process team's scope, we make sure they are aligned with the global process strategic roadmap." **(C)**
		Investment management*	*"We work on Smart rail 4.0 with a sister project in Personal Train Division to replace static planning of trains and persons. We will enable the Infrastructure Unit to create dynamic new slots for trains (10+ years perspective)."* **(A)** *"We need to buy marketing data from multiple research companies to execute certain use cases for instance."* **(C)**
		Business case identification*	*"The company 2030 strategy drives D&A initiatives. We want to get to the point where we monetize our data to customers. We have a lot of data that we could sell to our customers actually."* **(C)** *"We are heavily working on data strategy what we call rather a data enablement strategy started in March 2020. The focus more on business process optimization or operational excellence and digital transformation."* **(F)**
	Define & enforce data principles	Data standards & guidelines	*"Strategy and governance processes are done by central team: strategy, communication, standards & methods, quality measurements, maturity check, status reports."* **(A)** *"The responsibilities of the central data organization are the following: methods, guidelines, data definition including data models, DQ initiatives, support of business/IT projects with data know how."* **(B)**
		Data modeling & architecture	*"Data architecture and data modeling is done in the central data and analytics governance team. Still, we work closely with enterprise architecture management and data domain managers."* **(D)**
	Manage data operations	Data quality monitoring & support	*"Business data stewards take care of the governance structure, data quality monitoring, and support data owners."* **(F)** *"We have a corporate data quality index to measure it in 22 domains and we publish the results bi-yearly. It is signed off by the CFO and CIO."* **(D)**
		Data lifecycle management	*"We are not very mature with regards to the end of the data lifecycle. We often make data inactive, not getting rid of them."* **(C)** *"Domains have their own procedures to manage data lifecycle"* **(F)**
Relational	Align and collaborate with business & IT stakeholders	Communication	*"We have a monthly communication of practice globally. Each of region has only the monthly communication."* **(G)** *"We send monthly newsletter to inform our main stakeholders."* **(A)**
		Coordination* (formal, informal)	*"We have a service-provider relationship with IT or partnering."* **(G)** *"The data council meets twice a year and bring together all our digitalization heads to discuss projects, performance management, enterprise quality targets, architecture of data, KPI & Analytics"* **(D)**
	Develop & share data knowledge*	Communities of practice*	*"25 people are involved in a community of experts. Some of the data experts are informally linked to the data management and are also working on projects."* **(B)** *"We have multiple communities, for instance around BI tools. Also for master data management, we a larger only community on Yammer with hundreds followers"* **(C)**
		Educational programs	*"We have trainings for ownership but not anymore with covid, they will come back."* **(E)** *"This quarter we have worked with the business to deliver powerBI trainings (6-8 weeks program in projects with business) with one-to-one peer learning 1 to 1. We also had a conference organized by the AI enablement team and an AI rally where colleagues had to solve some task using AI concepts."* **(F)**

***Data governance practices that emerged from the cross-case analysis in response to the changing role of data.**

Table 32. Identified data governance mechanisms and practices

5 Data Governance Archetypes

From our analysis, we identify three data governance archetypes that elucidate the role of data governance as the role of data changes in organization: *Improve master data quality*, *Establish enterprise-wide data transparency*, and *Enable data monetization* (see Table 33). Each archetype is characterized by a set of distinct data governance practices and sub-practices that implement structural, procedural, and relational governance mechanisms.

Each archetype is driven by a particular *Data Strategy* focus, which mirrors the strategic orientation of the organizations' data-related initiatives. This strategic direction can be classified as either more defensive or offensive in nature. While all companies incorporate defensive strategies primarily aimed at enhancing data quality, some also pursue offensive objectives, seeking to monetize data in various ways to generate both direct and indirect business value. This often involves extending the underlying *data scope i.e.* the data types and domains that are needed to execute strategic use cases and which have been prioritized for data governance. In our sample, a narrow scope is associated with a strong emphasis on master data, typically less than five data domains.

In the following, we start with a brief overview of the data governance archetypes and then illustrate each of them based on our empirical insights from the nine cases (for detailed background information on each case, see Table 31).

5.1 Overview

Companies (here: B and G) belonging to the first governance archetype have a narrow scope and focus on improving data quality for master data in a few data domains, like customers, products and finance. We characterize this archetype as *Improve master data quality*. Companies use this initial structuring to focus on the most relevant data objects, typically supplier, customer, product, material or product master data, and define distinct areas of responsibility. While this approach remains the same for the other data governance archetypes, Archetype I has distinct characteristics: A central data team is granted operational responsibilities for collecting business requirements, setting up data quality measures, monitoring data quality, and supporting projects that involve data quality issues. Hence, the responsibilities are mainly centralized, although the data content is created in business units.

Companies (here: E, F, H, I) belonging to the second data governance archetype have a broader scope and comprise a diverse set of data domains and more data types than just master data.

Hence, we describe this archetype as *Enable enterprise-wide data management*. With this extended scope, the central data team has a wider array of responsibilities and starts defining a data strategy. While data quality remains a key central responsibility to ensure that data stays *fit for purpose* (Wang and Strong, 1996), data strategy and data access/availability to a broader number of employees are added to the central data team's responsibilities. While Archetype I nominates only a few decentralized roles that support data lifecycle activities, Archetype II decentralizes responsibilities for collecting business requirements and maintaining data according to domain-specific standards and guidelines. Therefore, relational mechanisms are more intensively established than in the first archetype. For instance, roles and responsibilities are communicated, and regular meetings and steering committees foster collaboration and alignment between data and business professionals.

Companies (here: A, C, D) belonging to the third data governance archetype recognize data as a strategic asset and a major driver of their digital transformation. Therefore, we characterize this archetype as *Coordinate the network to enable data monetization*. Building on their extensive experience in data management, these companies put specific emphasis on finding and enabling new ways to monetize data and establish a coordinated network of data roles that are not centrally organized. As data is considered a major value driver, these companies have an integrated view of data and analytics through which they foster synergies and seamlessly manage data quality and usage. The remaining central data team mostly undertakes strategic responsibility and is closely aligned with C-level executives. Hence, companies establish the role of the Chief Data Officer to foster alignment and steer data monetization activities enterprise wide.

5.2 Archetype I: Improve Master Data Quality

Strategic context and scope: Companies B and G are representative of the data governance–oriented Archetype I as both put in place data governance mechanisms to enable business processes/reporting, with a focus on master data quality. Company B has been facing numerous data quality issues in its operational processes, primarily in the financial domain (e.g., incorrect invoices). Hence, achieving high financial data quality for reporting and controlling is the company's major driver in its digital initiative, which debuted in 2020. Company G faces operational challenges regarding its supply chain, which is typical for the pharmaceutical industry (Desai and Peer, 2018). High-quality data is a major pillar of Company G's digital transformation journey, which the company embarked on in 2019 to optimize operations, anticipate business risks and enhance information transparency along the supply chain. The value of the data unfolds *"by bringing more information together, harmoniz[ing] data from different locations and us[ing] analytics to support product development"* (Head of Corporate Data Management, Company B).

	DATA GOVERNANCE ARCHETYPES		
	Archetype I ***Improve master data quality***	**Archetype II** ***Establish enterprise-wide data transparency***	**Archetype III** ***Enable data monetization***
STRATEGIC DATA OBJECTIVES AND DATA SCOPE			
Data strategy	Improve data quality to enable business processes/reporting	Improve data quality to enable business processes/reporting, broaden data access/availability to enable value creation	Improve data quality, broaden data access/availability, monetize data
Data scope	Narrow focus on master and reference data and few data domains (e.g., Supplier, Customer, Product, Material)	Broad focus on any data type and increasing number of data domains (e.g., Finance, HR, Controlling)	Broad focus on any data type including analytical data and stable number of data domains (any relevant)
STRUCTURAL MECHANISMS			
Shape the data organization	Small central data organization aligned with business and IT through projects or master data boards	Growing central data organization with decentral staff allocation or role assignment to business stakeholders, and emerging boards to decide on data governance principles with business	Large, federated data organization relying on divisional, functional and regional data governance hubs. Boards and councils to connect within and across the network
Assign data roles and responsibilities	Only essential data roles (head of data management, data steward, data architect)	Additional central oversight completed with the expansion of data steward and data owner roles into the business	Complete role model addressing strategic (Chief Data Officer), governance (e.g., data quality manager, data documentation manager) and operational roles (e.g., data citizen, data editor)
PROCEDURAL MECHANISMS			
Define and monitor data strategy and investments	Isolated strategy planning activities, investments in data quality improvements and infrastructure	Emerging data strategy planning process, investments in data quality improvements and infrastructure, business case analysis for new data domains	Data strategy planning and control process, pro-active identification, and management of data monetization opportunities
Define and enforce data principles	Creation of standards and data models for master data	Data governance framework and process for data modeling and architecture design	Data and analytics data governance framework, unified data architecture
Manage data operation	Data quality monitoring and support	Data quality monitoring and support, coordinated data lifecycle management	Data quality and use monitoring and support, and data lifecycle management in functions
RELATIONAL MECHANISMS			
Align and collaborate with business & IT stakeholders	Mostly through procedures or extended boards. Collocation with 1-2 data roles in IT functions	Collocation with an extended array of responsibilities for data-related aspects in IT function.	Collocation or even combined with a focus on delivering data and analytics products
Develop and share data knowledge	Few communities for master data. Few training options for non-specialists besides about compliant access and use	Regular updates. Emerging community management. Training in data quality methods and data literacy.	Enterprise-wide promotion of data. Personalized data literacy learning paths with peer coaching.

Table 33. Data governance archetypes

Structural mechanisms: Both companies have formed a small central data team that comprises fewer than 10 full-time equivalent employees (FTEs) and operates with a narrow scope on a few data domains relevant for their operations. Company B manages four (material, product, customer/account, vendor/supplier) and Company G two data domains (customer/account, material). For each data domain, the data teams have defined the company's core business objects (master data). This is a typical approach for this particular data governance archetype. The central data team includes data stewards who take over responsibility for the master data quality in a data domain. They work on developing the methods, standards, and guidelines to create and maintain master data and improve data quality in their data domain. In Company B, the central data team extends the scope to include managing reference data (e.g., product colors) as well. Besides the data stewards, data quality manager and data engineer, a dedicated data architect has been nominated as part of the IT organization to support data modelling purposes. Company G has a similar role: a data integration expert. While there are no accountabilities for data on a strategic level, the accountability for data's content lies within the business where they are created.

Procedural mechanisms: As yet, neither of the two companies have defined a comprehensive data strategy, but data is either formulated and embedded in their overarching digital strategy (Company B) or *"data and analytics were identified as core pillars of the overall digital transformation initiative"* (Global Data Lead-Enterprise Solution, Company G). In both companies, the central data teams are responsible for most of the data management processes in the organization (data quality monitoring, data standards), while the data lifecycle is mostly decentralized (in regions or business functions). Company G does have independent decentralized data teams that help to monitor data quality, maintain data, and support the central data team on projects. Company G also relies on a shared service center that supports data maintenance activities. Investment flows into data quality management and is driven either by the IT budget or by the budgets of business stakeholders. Hence, procedural mechanisms mainly focus on operational aspects and on deciding about the data's lifecycles.

Relational mechanisms: Data teams in both companies closely collaborate with business and IT. Company G characterizes the relationship with IT as a *"service-provider relationship,"* with IT providing solutions for the central team. In Company B, the data architect is collocated with IT, and the central data team participates in biweekly meetings related to IT enterprise architecture to align with the data requirements. Alignment and collaboration with business stakeholders happen through projects or collocation with process stewards (Company G). In Company G, monthly global and regional communication ensures knowledge sharing regarding common practices in using data. Company B facilitates active communities (e.g., material master data community) and a governance body for projects, which invites subject matter experts to contribute to data management projects.

5.3 Archetype II: Establish Enterprise-Wide Data Transparency

Strategic context and scope: Companies E, F, H, and I represent the data governance–oriented Archetype II, which is characterized by a strategic will to stimulate data provision and use in the entire organization. Company E sees digitalization as vital to its evolution in a connected world. It considers customers as business partners and seek to make customer experience a core dimension of its digital transformation. As a manufacturer in the automotive sector, Company F aims to enhance products and processes by becoming data driven. Company H is a large retailer and active in an industry that faces serious challenges because of digital competition and highly informed customers. It heavily relies on data to improve customer satisfaction, conversion rates, and customer reach. Owing to its growth through mergers and acquisitions, Company I relies on data for operational excellence and IT system landscape consolidation. Thus, the data architecture is key to establishing data governance.

Structural mechanisms: All four companies have a larger central data team (more than 15 FTEs) and a much broader scope than those in the first data governance, both in terms of data domains and in terms of data types. For instance, Company F has nine data domains (HR, market, purchasing, finance/controlling, supply chain, production, quality, development/engineering, business partners), and Company I has six (customer/consumer, vendor/supplier, product/article, material financial, employee). Company H follows a slightly different approach to define its areas of responsibility and has 26 domains (e.g., accounting/controlling, data assets, sellables/services) defined by *"going through all processes and business objects that we know to create a holistic view"* (VP Head of Data and Analytics, Company H). Besides master data, which is well established for all four companies, other data types are gaining momentum. These new data types include metadata, which is of the utmost importance to document data for different user groups, and transactional data, which is essential for analytics use cases. Beyond managing data quality, data availability and access are among the major concerns and responsibilities of the central data team. Accordingly, roles other than data steward are required across the data domains. These include dedicated roles for data quality (e.g., for creating metrics and monitoring), data standards and methods, and metadata management. Data management also contributes to analytics projects with the provided data and support for data architecture. Besides the centrally organized roles, the data team aims to decentralize responsibilities for managing the data lifecycle to business departments. This includes assigning accountabilities to business stakeholders in the core data domains, who proactively formulate their business requirements for data. Decentralized teams are organized either by region (Company I) or by business function (Company F). They include nominated roles such as data editors and data owners (for content, domain, or data definition), which are accountable for the data lifecycle or assigned governance responsibilities (Company H). Company E relies on a wide network of data standard owners (about 200 non-FTEs) who are spread across domains and nominated by the central team.

Procedural mechanisms: As data has greater strategic importance than Archetype I, the procedural mechanisms focus not only on operational aspects but also on strategic ones. Decisions are continuously made to review and update the data strategy, which is closely aligned with the IT strategy. In 2015, Company E released its master data strategy to integrate common elements across multiple flows and functions. To define the requirements and have an impact on business, Company E has an integrated strategy that extends the existing master data strategy to further data types and add analytics. Company F released its data enablement strategy in 2020, which focuses on business process optimization (operational excellence) and explores ways to turn business capabilities into data and analytics capabilities. Company H has had a data governance strategy since 2019 and will unveil its enterprise-wide data (and analytics) strategy in 2021. Since 2019, Company I has followed a "Data and Architecture strategy" synchronized with the "IT Strategy and Digitization Strategy" and aims to address *"how the organization can work on enterprise architecture with a greater leverage"* (Head of DM, Company I). All companies regularly monitor data quality through business stewardship and defined metrics (e.g., data quality KPIs at Company F). The budget for data management activities can be shared or is directly financed by the business. For instance, all master data–related activities are financed by the business at Company F as part of the MDM committee. The central data team ensures that domains have their own procedures to manage the data lifecycle. A roadmap of data management activities and a portfolio of data management projects help these central data teams to manage and monitor investments.

Relational mechanisms: Companies communicate regularly about data-related topics and projects through different channels. Company E uses newsletters and forums. As the data team aims to decentralize responsibilities, communication includes not only standards and compliant use but also roles, responsibilities, and methods that help to achieve the desired behavior. Boards and committees design the roadmap, nominate roles, and ensure the alignment of decision-making on data management activities between different stakeholders. They meet four to six times a year. Hence, the central data team aligns and collaborates more actively with business stakeholders. Collaboration with business can also happen through internal consulting services (Company I) or a network of support functions (Company H). Companies F and I use online collaboration platforms or chatbots to enable knowledge sharing and develop skillsets.

5.4 Archetype III: Enable Data Monetization

Strategic context and scope: This data governance archetype is represented by Companies A, C, and D, which are united in their strategic goal of regarding data as valuable products that can be both shared internally across teams, and commercialized. Company A is undergoing a digital transformation driven by increasing competitive pressure, changing customer needs, and new legal requirements. It aims to leverage data in order to improve customer satisfaction while reducing costs through automation. As a result, a transformation of the workforce is expected to address new skillset requirements and staff turnover in the coming years. By 2030, Company C aims to grow revenues by augmenting business with data and analytics insights and reducing costs through operational excellence. It has established a roadmap for enterprise data management by connecting data foundation, capabilities, and organization with business value as the outcome. Company D is active in the automotive industry, which is facing numerous challenges such as market changes toward e-mobility and automotive driving, customer requirements, and cost pressure (Koch, 2015). Company D has made major investments in implementing structured data management to support the company's business transformation. It has demonstrated results with regard to data excellence, innovation, and business value.

Structural mechanisms: As companies see data as a vital driver for the whole enterprise, the central data team sets priorities on formulating and rolling out the enterprise-wide data strategy by establishing the right set of data governance mechanisms. Companies in this data governance archetype establish the role of Chief Data Officer (or Head of Data and Analytics) to foster alignment and steer data monetization activities on a strategic level and across the firm. Business units are planning their data strategy and detailing standards for their respective areas of responsibility, having roles established on a strategic and operational level. Company C has a very small central data management team (six FTEs) setting priorities and designing data governance. This team also coordinates a wide, decentralized network of 100 business experts through an extended data leadership team of 22 business leaders. This structure is typical for the other companies as well. The decentralized data leadership team at Company A comprises 15 leading data managers. Company D has implemented data governance across 47 data domains and established enterprise-wide and data domain-specific standards, clarified data ownership, and assigned data management responsibilities. Its remaining central data and analytics governance team (10 FTEs) reports directly to the CEO and coordinates a decentralized network of 40 data domain managers (FTEs) in business functions, divisions, and regions, as well as 200 data coordinators (non-FTEs). A data council for project oversight and alignment focuses on prioritizing projects and data governance implementation concerns, among others.

Procedural mechanisms: Procedural mechanisms are established for managing investments in data, planning, and strategy. They are conducted in centralized (e.g., investment in data platform) and decentralized (e.g., staff supporting analytics projects) ways. Thus, data monetization opportunities are proactively identified, and business cases are formulated accordingly (e.g., by using analytics to predict machine outages). Company A renews its data strategy every four years – the current version dates to 2017 and is currently being renewed with a focus on having better data quality, developing roles and skillsets, establishing decentralized responsibilities, and increasing business data awareness. A dedicated strategy for analytics - separate from but coordinated with the data strategy – will also be unveiled to support the consolidation and decentralization of the analytics processes. Company C has integrated all data-related strategies (Master Data Management, BI, Marketing, Engineering) under the umbrella of an enterprise data strategy updated in 2019. BI governance is managed centrally while coordination is more spread out, following global processes and regions. At Company D, the data management strategy started with a focus on master data in 2016, and its scope was extended to all data areas in 2018, leading to a large, decentralized data management network. For the four companies, procedural mechanisms are established for data and analytics on strategic and operational levels.

Relational mechanisms: In this archetype, coordinating an increasing number of data communities and experts becomes a key concern. Alignment and collaboration occur on both an operational level (through communities) and a strategic level (through boards). Communication and knowledge sharing happen through data communities, which comprise key data users and are actively coordinated as virtual networks. For strategic alignment and collaboration, companies establish data steering committees in which key business stakeholders regularly assess and review the roll-out of the data strategy. Beyond formulating the company's vision related to data, quantified goals, and required operations, Company C's strategy encompasses topics related to enterprise culture transformation (e.g., training) and organizing (e.g., teams, principles). Establishing data teams in business is highlighted as a key milestone for the development of capabilities such as data literacy and data democratization. Company A also ensures alignment and collaboration through boards (e.g., data management board) and communities (e.g., AI network group, shared learning group for similar jobs). Company D is building its next-generation enterprise architecture, which will include alignment beyond IT collocation. All three companies ensure alignment and collaboration thanks to regular high-level DM and D&A board meetings.

6 Contribution and Discussion

Our results contribute to data governance research on two principal fronts. Firstly, our research substantiates the proposition that data governance emerges distinctly from IT governance. This is attributed to the unique nature of data as a governance subject, whose utilization is deeply rooted into specific work practices. Secondly, we conceptualize data governance as a repertoire of practices and identify configurations in the form of archetypes.

6.1 How Data Governance Emancipates from IT Governance

For a long time, data governance has been treated in research and practice as an inclusive responsibility of IT management (e.g., Tallon et al. (2013)). While the general governance mechanisms hold true for both data and IT, our study confirms that the type and nature of data governance practices differs significantly from those for IT. Data assets, by nature, tend to be highly decentralized and subject to fluctuations, with their value depreciating at a much quicker pace than that of IT assets. While IT infrastructure and applications are built once and are then operated for years, data in terms of data sets consisting of individual records are created, updated, and deleted on a continuous basis. Also, data must be shared across business units to expand the array of data repurposing possibilities which requires an overarching perspective independent of the underlying IT application landscape. Accordingly, structural data governance practices are more decentralized than IT governance practices. They comprise data roles and teams which operate mostly in a decentral manner where data are created and consumed, while establishing new roles and responsibilities to coordinate and facilitate this process. In this way, the value creation process from data can be best supported.

Therefore, our study provides evidence that data is governed independently from IT and that data governance should therefore be recognized as such. This finding goes somewhat counter to earlier studies, which advocate the placement of IT and data/information governance under the same structure as the "*preferable*" option (Tallon et al., 2013, p.169). In contrast to the prevailing view that data is an integral responsibility of IT organizations, our study demonstrates the importance of data governance as separate instrument to sustain a strategic competitive advantage. However, the collaboration between both organizations remains essential in all case companies, albeit the IT organization is more seen as a service-provider, especially when developing data augmented software applications, e.g., dashboards, or integrate machine learning models in workflows and enterprise applications. IT applications are often created by central teams, and the role of end-users is limited to specifying requirements. This is the opposite in the data domain view. Here, the data creators and consumers are the end-users. Hence, companies have an interest in levelling up their data literacy through situated curricula. This way, data roles can apply their data skills in a relevant working context (e.g., analyze data with self-service) and generally improve decision-making (D'Ignazio & Bhargava, 2015; Lefebvre & Legner, 2024).

6.2 How Data Governance Changes Towards Data Monetization

Our study provides evidence that data governance practices change according to the higher value contribution companies pay to data. Whereas traditional data governance practices are control-oriented to accommodate defensive strategies, offensive data strategies which emphasize on data monetization go hand in hand with expanding data practices to stimulate data-driven innovation. With data monetization, companies proactively search for and monitor their data use cases, which has implications on the strategic decision-making processes and procedural governance mechanisms. We also spotlight the pivotal yet previously understated role of relational data governance practices in fostering the enablement of new data roles, particularly through the sharing of practices and context-specific learning.

Our findings directly responds to a recent call for research to develop a conceptualization of data governance as repertoires of mechanisms that form configurations that contribute to the achievement of organizational outcomes (Vial, 2023), with data monetization as frontier.

The three archetypes illustrate how data governance design evolve beyond the focus on the data quality and operational control aspects shown in previous studies (e.g., Otto, 2011; Tallon, Ramirez and Short, 2013). The first archetype is widely acknowledged as a foundational model in current data governance literature, as extensively explored in current scholarly works. However, the other two archetypes, which are not strictly indicative of progressive stages of maturity, illustrate the strategic implementation of data governance by businesses to harness data as a strategic asset and leverage data's monetization opportunities.

The Archetype I is representative for the traditional data governance research, with strong focus on data quality and control of critical data resources. It reflects data's role as supporting resource, and indirect value creation from data through automated and integrated business processes or reporting. Data governance practices are mostly established centrally and aim at harmonizing enterprise-wide master data across business units to improve data availability in operational systems. Archetype II and III can be interpreted as the evolution of this first archetype in response to data's changing role. The data organizations belonging to the archetype

II are in an enterprise-wide consolidation and expansion phase of their data initiatives. The corresponding companies consolidate the different data teams, i.e., master data, business intelligence, data science, under one umbrella. To better align these teams and steer them enterprise-wide, an overarching data governance framework is required which helps not only in decentralizing the data lifecycle operations but also in broadening data access so that basically everyone in the company can contribute to the value creation from data. Assigning some accountabilities for data to authorities sitting directly in the business units is a common structural data governance practice. Data are created and consumed in business units and must be steered from there to avoid any bottlenecks in changing and delivering datasets. Organization in Archetype III are in their decentralization and growth phase. Data is a key part of the overall business strategy to foster the digital transformation and monetize it in multiple ways. Most of the accountabilities have been decentralized to the business units and only some remain as part of the dedicated central data governance team for coordination and steering purposes. Thus, the nature of data governance changes from ensuring compliance with standards towards enabling local, situated work practices of dispersed data users. This goes hand in hand with redefining the interplay between informal and formal governance arrangements. Accordingly, relational data governance practices, which are "*less formalized means of ensuring that data governance principles are understood and enforced by actors* "(Vial, 2023), gain in importance to coordinate and enable a broad network of data creators and users. Key relational data governance practices include data literacy upskilling, data culture development, and knowledge sharing within data roles, and between data and IT teams for delivering data augmented applications. Such evolution is mirrored by the ongoing discourse on the democratization of data in enterprise by putting forward practice exchange as pivotal for the development of situated data practices among non-specialists (Awasthi & George, 2020; Lefebvre & Legner, 2022; Zeng & Glaister, 2018).

6.3 Implications

Our study, which provides fundamental insights how companies adapt their data governance practices to address the emerging strategic role that data plays, have several implications for research.

First, our findings are of high relevance to explain federated data governance (Grover et al. 2018; King 1983). Despite numerous benefits such as greater local autonomy, faster issue resolution, and improved agility, federated data governance is often challenging to implement (Otto, 2011b) due to global firms' complex organizational structures (Otto, 2011b). As displayed in Archetype 3, which combines central and decentral data governance responsibilities, data governance is pivotal in coordinating data operations (e.g., data use, data curation) across an organization's core structure composed of functions, divisions, or regions. These insights further show that data governance must reach many different parts of an organization and shape the situated data practices through which data acquires its value. This extensive reach is often facilitated by the introduction of supplementary coordination mechanisms, which are implemented through relational data governance practices. This reinforces the innovative perspective of data governance as a coordinating function that recognizes decentral data ownership, rather than merely serving as a control mechanism (Vial, 2023). We encourage further research in that direction, for instance looking at how decentral data governance team are integrated into the organizational structure to foster data-driven innovation.

Second, our findings extends and complements prior research that proposed overarching frameworks for data governance (Abraham et al., 2019; Tallon et al., 2013). They also align with recent studies that suggested "*a shift from data governance as a matter of asset management to data governance as a matter of work practice*" (Parmiggiani & Grisot, 2020, p. 3). Researchers argue that data governance cannot simply focus on *what* data governance practices should be implemented while ignoring *how* to implement these practices (e.g., curating data following certain governance standards) (Benfeldt et al., 2020; Vial, 2023). As Alhassan (2016) highlighted, data roles and responsibilities should specify how data governance practices are defined, but also how they are implemented and monitored by various employees in different business units. By elucidating sub-practices that implement data governance practices, we therefore directly address these concerns.

Third, our findings hold significant implications for the emerging discourse on the dynamism of data governance (Vial, 2023). The identified archetypes provide empirically and theoretically-ground evidence of how data governance evolves in symbiosis with strategy and operations, thereby supporting prior arguments about data governance's contribution to firm performance (Mikalef et al., 2020). Obvious avenues for further research thus include a detailed investigation of how data governance continuously maintains its dual mandate of control and coordination and manages the possible tensions between them. Although our archetypes delineate distinct "states" of data governance, interesting research opportunities lie into the analysis of the transition between these states and the implementation of corresponding, potentially new, data governance practices.

For practitioners, our study provides insights into data governance initiatives in multinational corporations and identifies data governance mechanisms that can guide them to manage data as a strategic asset. As an implication, practitioners should not only focus on structural mechanisms, but concretize these roles by establishing data-related processes (procedural governance) and put particular emphasis on improving collaboration on data-related topics between business, IT and data and analytics groups (relational governance).

6.4 Limitations

Our study comes not without limitations. Firstly, our sample includes only large, multinational corporations that have complex organizational structures and are characterized by a high degree of specialization and division of labor. They also require more resources for alignment and collaboration. Therefore, our findings might not be applicable to smaller organizations. Secondly, we solely focus on understanding data governance mechanisms and comparing them between companies. We did not analyze the interplay between corporate, IT, and data governance, which presents an interesting avenue for future research.

7 References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. *Journal of Management Information Systems*, *38*(2), 401–429.

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).

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.

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113–131.

Alaimo, C., & Kallinikos, J. (2022). Organizations Decentered: Data Objects, Technology and Knowledge. *Organization Science*, *33*(1), 19–37.

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems*, *25*(sup1), 64–75.

Awasthi, P., & George, J. J. (2020). A Case for Data Democratization. *Proceedings of the 26th Americas Conference on Information Systems (AMCIS)*, 23.

Baijens, J., Helms, R. W., & Velstra, T. (2020). Towards a Framework for Data Analytics Governance Mechanisms. *Proceedings of the 28th European Conference on Information Systems (ECIS), An Online AIS Conference*.

Baijens, J., Huygh, T., & Helms, R. (2021). Establishing and theorising data analytics governance: A descriptive framework and a VSM-based view. *Journal of Business Analytics*, 1–22.

Ballou, D., Wang, R., Pazer, H., & Tayi, G. K. (1998). Modeling Information Manufacturing Systems to Determine Information Product Quality. *Management Science*, *44*(4), 462–484.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, *22*(2), 299–313.

Chua, C., Indulska, M., Lukyanenko, R., Maass, W., & Storey, V. C. (2022, February 14). *MISQ Research Curation on Data Management*. MIS Quarterly. https://www.misqresearchcurations.org/blog/2022/2/11/data-management

Dallemulle, L., & Davenport, T. (2017). What's your data strategy? *Harvard Business Review*, *May-June*, 112–121.

De Haes, S., & Van Grembergen, W. (2004). IT Governance and its Mechanisms. *Information Systems Control Journal*, *1*, 27–33.

Desai, S. S., & Peer, B. (2018). Big Pharma; Big Data—Big Deal? Yes; Really! Infosys.

D'Ignazio, C., & Bhargava, R. (2015). Approaches to Building Big Data Literacy. *Proceedings of the Bloomberg Data for Good Exchange Conference*, 6.

Dubé, L., & Paré, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597–636.

Eisenhardt, K., & Graebner, M. (2007). Theory Building from Cases: Opportunities and Challenges. *The Academy of Management Journal*, *50*(1), 25–32.

Fadler, M., & Legner, C. (2020). Who Owns Data in the Enterprise? Rethinking Data Ownership in Times of Big Data and Analytics. *Proceedings of the 28th European Conference on Information Systems*.

Fadler, M., & Legner, C. (2021a). Data ownership revisited: Clarifying data accountabilities in times of big data and analytics. *Journal of Business Analytics*, *5*(1), 123–139.

Fadler, M., & Legner, C. (2021b). Toward big data and analytics governance: Redefining structural governance mechanisms. *Proceedings of the 54th Hawaii International Conference on System Sciences*.

Goes, P. (2014). Big data and IS research. *MIS Quarterly*, *38*(3), iii–viii.

Goodhue, D. L., Quillard, J. A., & Rockart, J. F. (1988). Managing the Data Resource: A Contingency Perspective. *MIS Quarterly*, *12*(3), 373–392.

Gregory, R. W., Kaganer, E., Henfridsson, O., & Ruch, T. J. (2018). IT Consumerization And The Transformation Of It Governance. *MIS Quarterly*, *42*(4), 1225–1253.

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.

Grover, V., & Teng, J. T. C. (1991). HOW EFFECTIVE IS DATA RESOURCE MANAGEMENT?Reassessing Strategic Objectives. *Journal of Information Systems Management*, *8*(3), 16–20.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049–1064.

Hassan, N. R. (2019). Where are we headed in business analytics? A framework based on a paradigmatic analysis of the history of analytics. *Proceedings of the 39th International Conference on Information Systems (ICIS)*, *6*, 17.

Hoetker, G., & Mellewigt, T. (2009). Choice and performance of governance mechanisms: Matching alliance governance to asset type. *Strategic Management Journal*, *30*(10), 1025–1044.

Huang, R., Zmud, R. W., & Price, R. L. (2010). Influencing the effectiveness of IT governance practices through steering committees and communication policies. *European Journal of Information Systems*, *19*(3), 288–302.

Jain, H., Ramamurthy, K., Ryu, H.-S., & Yasai-Ardekani, M. (1998). Success of Data Resource Management in Distributed Environments: An Empirical Investigation. *MIS Quarterly*, *22*(1), 1–29.

Jones, M. (2019). What we talk about when we talk about (big) data. *The Journal of Strategic Information Systems*, *28*(1), 3–16.

Karimi, J., Bhattacherjee, A., Gupta, Y. P., & Somers, T. M. (2000). The Effects of MIS Steering Committees on Information Technology Management Sophistication. *Journal of Management Information Systems*, *17*(2), 207–230.

Ketokivi, M., & Mantere, S. (2010). Two strategies for inductive reasoning in organizational research. *Academy of Management Review*, *35*(2), 315–333.

Khatri, V., & Brown, C. V. (2010). Designing Data Governance. *Communication of the ACM*, *53*(1), 148–152.

Koch, M. (2015). Big data and analytics in the automotive industry. Automotive analytics thought piece. Deloitte.

Kohli, R., & Grover, V. (2008). Business Value of IT: An Essay on Expanding Research Directions to Keep up with the Times. *Journal of the Association for Information Systems*, *9*(1), 23–39.

Lavie, D., Haunschild, P. R., & Khanna, P. (2012). Organizational differences, relational mechanisms, and alliance performance. *Strategic Management Journal*, *33*(13), 1453–1479.

Lee, Y., Madnick, S., Wang, R., Wang, F., & Zhang, H. (2014). A Cubic Framework for the Chief Data Officer (CDO): Succeeding in a World of Big Data. *MIS Quarterly Executive*, *13*(1).

Lefebvre, H., & Legner, C. (2022). How Communities of Practice Enable Data Democratization Inside the Enterprise. *Proceedings of the 30th European Conference on Information Systems (ECIS)*, 16.

Lefebvre, H., & Legner, C. (2024). Toward a Curriculum for Data Literacy in Enterprises. *Proceedings of the 57th Hawaii International Conference on System Sciences (HICSS)*.

Legner, C., Pentek, T., & Otto, B. (2020). Accumulating Design Knowledge with Reference Models: Insights from 12 Years' Research into Data Management. *Journal of the Association for Information Systems*, *21*(3), 735–770.

Lycett, M. (2013). 'Datafication': Making sense of (big) data in a complex world. *European Journal of Information Systems*, *22*(4), 381–386.

Mathiassen, L. (2002). Collaborative Practice Research. *Information Technology & People*, *15*(4), 321–345.

Mehta, S., Dawande, M., Janakiraman, G., & Mookerjee, V. (2021). How to Sell a Data Set? Pricing Policies for Data Monetization. *Information Systems Research*, *32*(4), 1281–1297.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2020). The role of information governance in big data analytics driven innovation. *Information & Management*, *57*(7), 103361.

Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. N. (2018). Big data analytics capabilities: A systematic literature review and research agenda. *Information Systems and E-Business Management*, *16*(3), 547–578.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.

Otto, B. (2011a). A morphology of the organisation of data governance. Proceedings of the 19th European Conference on Information Systems (ECIS).

Otto, B. (2011b). Data Governance. *Business & Information Systems Engineering*, *3*(4), 241–244.

Otto, B. (2011c). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. *Communications of the Association for Information Systems, 29*.

Paré, G. (2004). Investigating Information Systems with Positivist Case Research. *Communications of the Association for Information Systems*, *13*(1), Paper 18.

Parmiggiani, E., & Grisot, M. (2020). Data curation as governance practice. *Scandinavian Journal of Information Systems*, *32*.

Peppard, J. (2018). Rethinking the concept of the IS organization. *Information Systems Journal*, *28*(1), 76–103.

Peterson, R. (2004). Crafting Information Technology Governance. *Information Systems Management*, *21*(4), 7–22.

Phillips-Wren, G., Iyer, L., Kulkarni, U., & Ariyachandra, T. (2015). Business Analytics in the Context of Big Data: A Roadmap for Research. *Communications of the Association for Information Systems*, *37*(1).

Poppo, L., & Zenger, T. (2002). Do formal contracts and relational governance function as substitutes or complements? *Strategic Management Journal*, *23*(8), 707–725.

Prasad, A., Green, P., & Heales, J. (2012). On IT governance structures and their effectiveness in collaborative organizational structures. *International Journal of Accounting Information Systems*, *13*(3), 199–220.

Rifaie, M., Alhajj, R., & Ridley, M. (2009). Data governance strategy: A key issue in building Enterprise Data Warehouse. *Proceedings of the 11th International Conference on Information Integration and Web-Based Applications & Services (iiWAS)*, 587–591.

Sambamurthy, V., & Zmud, R. W. (1999). Arrangement for Information Technology Governance: A Theory of Multiple Contingencies. *MIS Quarterly*, *23*(2), 261–290.

Sambamurthy, V., & Zmud, R. W. (2000). Research Commentary: The Organizing Logic for an Enterprise's IT Activities in the Digital Era—A Prognosis of Practice and a Call for Research. *Information Systems Research*, *11*(2), 105–114.

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.

Tihanyi, L., Graffin, S., & George, G. (2014). Rethinking Governance in Management Research. *Academy of Management Journal*, *57*(6), 1535–1543.

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.

Velu, C. K., Madnick, S. E., & van Alstyne, M. W. (2013). Centralizing Data Management with Considerations of Uncertainty and Information-Based Flexibility. *Journal of Management Information Systems*, *30*(3), 179–212.

Vial, G. (2023). Data governance and digital innovation: A translational account of practitioner issues for IS research. *Information and Organization*, *33*(1), 1–12.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S. J., Dubey, R., & Childe, S. J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356–365.

Wang, R. Y. (1998). A product perspective on total data quality management. *Communications of the ACM*, *41*(2), 58–65.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33.

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.

Weber, K., Otto, B., & Österle, H. (2009). One Size Does Not Fit All-A Contingency Approach to Data Governance. *J. Data and Information Quality*, *1*(1), 1–27.

Weill, P., & Ross, J. W. (2004). IT Governance: How Top Performers Manage IT Decision Rights for Superior Results. Harvard Business Press.

Wenger, E. (2000). Communities of Practice and Social Learning Systems. *Organization*, *7*(2), 225–246.

Winkler, T. J., & Wessel, M. (2018, December 13). A Primer on Decision Rights in Information Systems: Review and Recommendations. *Proceedings of the 39th International Conference on Information Systems (ICIS)*.

Wixom, B. H., & Ross, J. W. (2017). How to Monetize Your Data. *MIT Sloan Management Review*, *58*(3), 9–13.

Wu, S. P.-J., Straub, D., Liang, T.-P., Temple University, Korea University Business School, Liang, T.-P., & National Chengchi University. (2015). How Information Technology Governance Mechanisms and Strategic Alignment Influence Organizational Performance: Insights from a Matched Survey of Business and IT Managers. *MIS Quarterly*, *39*(2), 497–518.

Xue, Y., Liang, H., & Boulton, W. R. (2008). Information Technology Governance in Information Technology Investment Decision Processes: The Impact of Investment Characteristics, External Environment, and Internal Context. *MIS Quarterly*, *32*(1), 67–96.

Yin, R. K. (2003). *Case Study Research—Design and Methods* (3rd ed.). Sage Publications.

Zeng, J., & Glaister, K. W. (2018). Value creation from big data: Looking inside the black box. *Strategic Organization*, *16*(2), 105–140.

8 Appendix

Section	Sample questions
1. Drivers and strategy	What are the drivers for data and analytics in the company? Do you have a data and/or analytics strategy? If yes, since when and what is its focus? What is the business value and benefit created by data and analytics?
2. Scope	Which data domains do you distinguish? How do you define them? Which data types are established or emerging? Wich data and analytics products do you deliver?
3. Data and analytics organization (structural governance practices)	What organizational form has been chosen (line function, shared service etc.)? Is the central team/department part of the primary organization and - if so - where is it located in the organizational structure? What are the responsibilities, headcount, structure an composition of data and analytics teams? Are there any boards and committees for data and analytics? What is their role?
4. Processes (procedural governance practices)	Which data management processes have you established? Which steps / tasks are taken over by the central / decentral data organization? Which analytics processes have you established? Which steps / tasks are taken over by the central / decentral data organization?
5. Alignment and collaboration (relational governance practices)	How do you align and collaborate with business stakeholders? How do you align and collaborate with IT stakeholders? How align and collaborate between data and analytics? Which data / analytics communities exist? How do you engage with them?

Table 34. Interview protocol

Essay 5

Rethinking Data Governance: A Viable System Model

Hippolyte Lefebvre, Christine Legner

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

Published in the proceedings of the 32nd European Conference on Information Systems (2024)

Extension added to this manuscript

Abstract: *Data governance is a prerequisite for organizations wanting to harness the strategic potential of data. Although the conceptual foundations of data governance have reached a sound level of clarity, research still does not explain how data governance unfolds in large and complex organizations. To address this gap, we introduce the Viable System Model as theoretical lens and examine data governance at five multinational companies with varied organizational structures. We find that data governance orchestrates data practices on multiple, interconnected levels, through sub-systems. The interplay between these sub-systems facilitates the establishment of a dynamic balance, enabling (1) the delineation of responsibilities, distinguishing between global and local data governance that orchestrates data practices, and (2) the implementation of data practices at the operational level that simultaneously emphasize control and foster innovation. Our research contributes to rethinking data governance and addresses previous calls for research that accounts for its dynamic nature in practice.*

Keywords: Data governance, Data practice, Viable System Model, Systems thinking.

Table of contents

List of figures

List of tables

1 Introduction

Successful organizations recognize the strategic potential of data for sustainable competitive advantage (Jones, 2019) and its vital role in creating business value, such as cost efficiency or better market positioning (Günther et al., 2022). A prerequisite for unlocking the potential of data is data governance, i.e., the specification of "*a cross-functional framework for managing data as a strategic enterprise asset*" (Abraham et al., 2019, p. 425). Grover *et al.* (2018) even argue 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*" (p. 417). Data governance has long been concerned with the quality and protection of data assets and the adherence to regulatory requirements (Weber et al., 2009; Otto, 2011). Today, data is at the heart of value creation in enterprises, resulting in data governance having the dual purpose of simultaneously balancing control and innovation (Vial, 2023).

Data governance research has mainly focused on clarifying the basic understanding and defining the scope and overall framework of data governance (Abraham et al., 2019; Khatri & Brown, 2010). Building on IT governance literature, it conceptualizes data governance as an ensemble of mechanisms (Abraham et al., 2019; Tallon et al., 2013; Vial, 2023) encompassing structural mechanisms (e.g., roles, responsibilities, locus of decision making), procedural mechanisms (e.g., processes, monitoring), and relational mechanisms (e.g., communication, training). While the foundations of data governance are increasingly clear, criticism has emerged from practice claiming that data governance cannot be viewed only "*as series of mechanisms implemented in organizations, at the expense of understanding the process of governing data*" (Vial, 2023, p. 6). Concretely, research still mainly lists *what to do* and does not explain *how to do* data governance, i.e., data governance in practice (Aaltonen et al., 2021; Alhassan et al., 2016; Vial, 2023). Moreover, given global firms' complex organizational structures, establishing data governance for them remains a challenge (Otto, 2011). In order to be effective, data governance must reach many different parts of an organization and shape the situated data practices through which data acquires its value (Parmiggiani & Grisot, 2020). Federated data governance models, which combine global and local data governance responsibilities, have been proposed as a solution in rolling out data governance in accordance with the primary organizational structure (Grover et al. 2018; King 1983). However, so far, no link has been established for understanding how data governance mechanisms materialize at local and global levels. Further, the rather static view of data governance mechanisms does not properly explain the dynamic nature of data governance which must evolve in symbiosis with strategy and operations (Benfeldt et al., 2020). As markets, regulations, and organizational culture are continuously evolving, data governance obviously has to adapt (Abraham et al., 2019; Tallon et al., 2013).

In such a context, we ask the following research question (RQ):

RQ: *How does data governance unfold in multinational companies?*

In our study, we apply system thinking to data governance and use the Viable System Model (VSM) as theoretical lens. The VSM explains a system's viability, i.e., its ability to maintain its existence in a changing environment (Beer, 1985), and it has been used to explain IT governance setups (Huygh & De Haes, 2019; Peppard, 2005). Our study is embedded in a collaborative practice research (Mathiassen, 2002), with 17 multinational companies. It is informed by insights from nine focus groups, as well as in-depth case studies. To understand how governance mechanisms are implemented in large and complex organizations, we analyzed the cases of five companies that have developed global and local data governance responsibilities. Our findings reveal that data governance orchestrates data practices on multiple, interconnected levels, through sub-systems. The interactions between data practices happening at operational, governance, and strategic levels make it possible to establish an appropriate balance that mediates (1) between global and local data governance, and (2) between data governance activities that seek control on the one hand and innovation on the other. Overall, closing this research gap advances the academic understanding of federated governance, paving the way for a new angle in investigating data practices at strategic, governance, and operational levels. Our research offers practitioners guidelines on how to set up a data governance framework that aligns with their overall strategy and organizational structure.

In the remainder of the paper, we first give information on prior data governance literature and highlight the research gap. Second, we motivate the relevance of system thinking and the applicability of VSM as a theoretical lens. Next, we present our methodology, and finally, we summarize and discuss our findings, and also provide an outlook on future research.

2 Background

2.1 Data governance

Data governance is seen as a framework describing cross-functional efforts for maximizing the value of data as strategic enterprise assets and ensuring the compliant and strategic use of data (Abraham et al., 2019; Tallon et al., 2013). It thus fosters the contribution data makes to achieving organization goals and generally aims to improve firm performance (Mikalef et al., 2020). Data governance is shaped by both external environmental antecedents, such as legal and regulatory, industry, or regional conditions, and internal ones, such as business strategy, corporate culture, or organizational structure (Baijens et al., 2021; Tallon et al., 2013).

To set up data governance, firms should clearly identify its scope along three dimensions (see Figure 15). First, organizational scope refers to "*expansiveness of data governance*"(Abraham et al., 2019, p. 430), which can be intra-organizational or inter-organizational. Second, firms define the data scope and identify the relevant data objects, data types, and data domains to prioritize for data governance. For instance, master or transactional data objects are usually governed first, but other big data-related types such as media data and sensors can come in scope later to support new data applications (Abraham et al., 2019; Fadler et al., 2021). Third, the "depth" of the data governance program is defined by its domain scope, i.e., the different data decision domains, such as data quality, data security, data architecture, data lifecycle, metadata, data storage, and infrastructure (Abraham et al., 2019; Khatri & Brown, 2010).

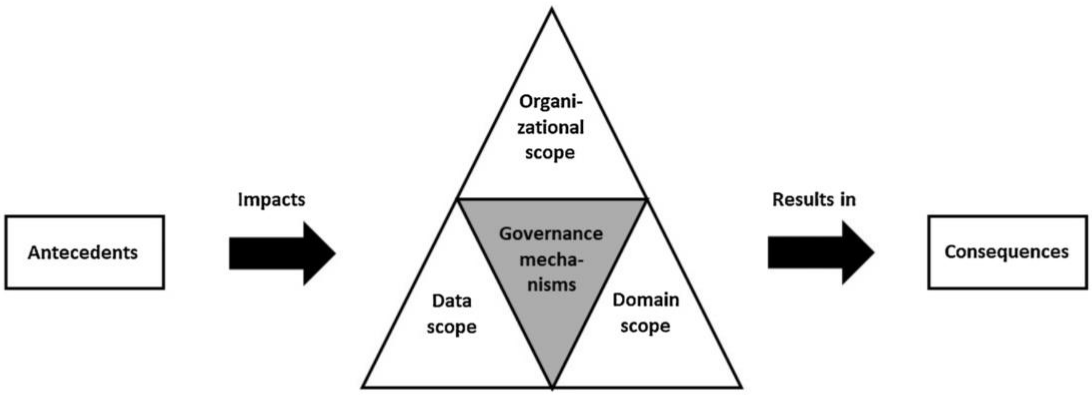


Figure 15. Conceptual framework for data governance by Abraham et al. (2019)

Three mechanisms—structural, procedural, and relational—constitute the core of data governance, drawing from established IT governance frameworks (Abraham et al., 2019; Tallon et al., 2013; Vial, 2023). These mechanisms should be combined and not addressed separately for maximum efficiency (Tallon et al., 2013), and can typically be bundled into archetypes aligned on the strategic context and scope for data (Fadler et al., 2021).

Structural mechanisms focus on specifying roles (e.g., data owner, data steward) and responsibilities in line with the organizational structure, and on allocating decision-making authority. This entails defining where the different data teams are positioned and their reporting lines (Otto, 2011). The literature differentiates between centralized, decentralized, and federated data governance designs (Brown, 1999; Sambamurthy & Zmud, 1999; Weber et al., 2009). A strict centralized data governance model implies that a central data unit has global authority and responsibility regarding data. Such a model is convenient for company-wide control, efficiency, and reliability in the (re)utilization of data assets because it leverages lateral organizational capabilities between units. However, it decreases local units' flexibility and capacity to innovate (Grover et al., 2018; Velu et al., 2013). Conversely, in a fully decentralized model, business units hold local responsibility for their data, each with their respective governance principles which enable rapid adaptation to changing requirements (Velu et al., 2013). In this model, the lacking standardization leads to coordination challenges, compliance concerns, data quality issues, limited collaboration, and complex data access management. Federated (also called hybrid, or Hub-Spoke) models combine the two forms in a global hub responsible for enterprise-wide standards, policies, methods, and tools, with business units as spokes taking care of responsibilities closer to the relevant data operations (e.g., data creation, data quality, data maintenance) (Grover et al., 2018; King, 1983). While offering numerous benefits such as greater local autonomy, faster issue resolution, and improved agility, a federated model generally requires better coordination mechanisms and acknowledged data ownership by respective business units (Velu et al., 2013).

The procedural and relational mechanisms instantiate the structural mechanisms. *Procedural mechanisms* describe decision-making related to data activities and processes, and thereby "*emphasize the operational means that are put in place to ensure compliance with governance principles"(Vial, 2023, p. 4)*. These include data strategy; policies, standards, and procedures; contractual agreements; performance measurement; compliance monitoring; and issue management (Abraham et al., 2019). *Relational mechanisms* ensure alignment, collaboration, and knowledge sharing between stakeholders. To expand the reach and understanding of data governance principles, these mechanisms usually comprise both formal (e.g., working groups, collaboration platform, training events) and informal (e.g., job rotation, corporate events, communities) means of coordination (Abraham et al., 2019). For instance, communities of practice foster knowledge sharing and data literacy among both data experts and non-experts (Lefebvre & Legner, 2022).

The above view on data governance has attracted criticism because the governance mechanisms do not explain data governance in practice (Aaltonen et al., 2021; Alhassan et al., 2016; Vial, 2023). Recent research suggests "*a shift from data governance as a matter of asset management to data governance as a matter of work practice*" because data governance is enacted as part of local actors' sense-making processes, such as during data curation tasks (Parmiggiani & Grisot, 2020, p. 3). Therefore, firms naturally evolve toward federated data governance that accommodates both global and local needs (Benfeldt et al., 2020), thus pragmatically reflecting the organizational complexity of the organization, specifically in multinational companies (Velu et al., 2013; Khatri & Brown, 2010). This shift is also reflected in the emerging data mesh paradigm which emphasizes data management responsibilities close to data creators because they know the context the best (Machado et al., 2021). Further, data governance should be addressed as a "*dynamic element that is implemented and should evolve in conjunction with strategy and operations*" to maintain its dual purpose of balancing control and data-driven innovation (Vial, 2023, p. 9). However, the literature neither explains how data governance responds to growing operational needs (e.g., data requests in business) nor clarifies data governance's role in assimilating strategic decisions. This gap calls for further investigation of how data governance unfolds in practice.

2.2 A systems thinking approach to address data governance in practice

We argue that systems thinking, and especially the VSM, offers a promising lens to study data governance as a system dynamically shaped by antecedents and composed of a set of interrelated sub-systems. The VSM introduces the concept of viability, suggesting that a system is able to remain functional despite a dynamic and fluctuating environment (Beer, 1985). It provides a framework for describing organizations and how they process information between different entities, including internal departments, external partners, and the broader environment which represents surrounding external factors that could influence the system (see Figure 16). This framework emphasizes the continuous interactions and information exchanges (symbolized by the arrows between each element), both critical aspects of organizational decision-making, adaptation, and innovation.

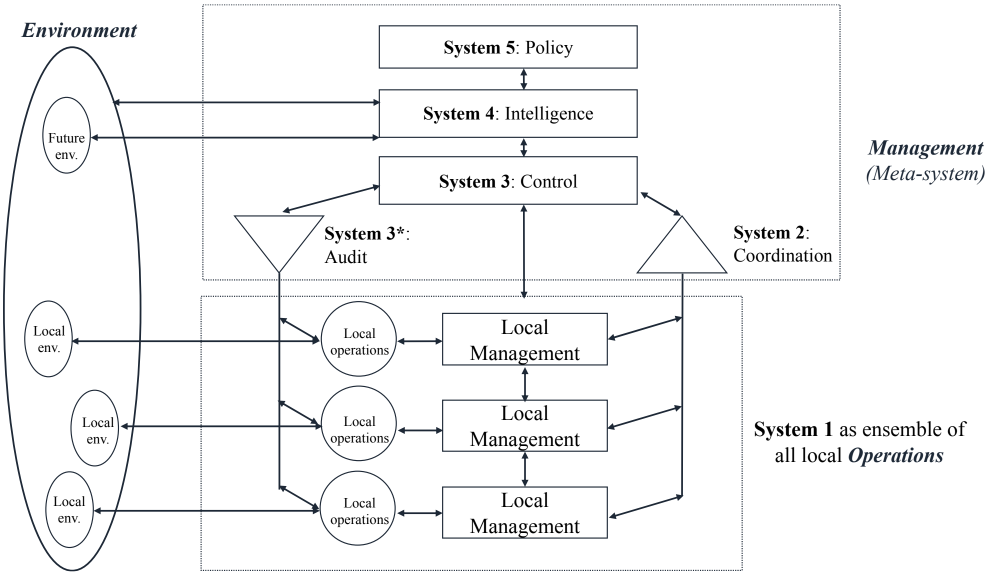


Figure 16. Structure and relationships in the Viable System Model (simplified representation based on Beer (1985))

The VSM posits self-organizing systems as composed of five sufficient interconnected sub-systems (Systems 1 to 5) that each have a role in maintaining the viability of the system (Beer, 1985), i.e., all sub-systems must be active and continuously exchange information:

- System 1 represents the *Operations* element of the VSM. As system-in-focus, it describes the different local operative units that execute the necessary tasks (i.e., work practices) that maintain the entire system's purpose. These operative units are typically embedded in the organization's primary structure and have their own local management. They can communicate with one another.
- Systems 2 to 5 – coordination, control, intelligence, policy – together form the *Management* element of the VSM, which acts as meta-system determining System 1. Thereby, they ensure smooth operation delivery (e.g., scheduling, strategic planning).

By applying the VSM as theoretical lens we can gain a thorough understanding of how data governance practices are arranged to assimilate and accommodate changes (e.g., in data scope). This lens also illustrates how data governance is embedded in the organizational structure. This approach has been employed to investigate IT governance (e.g., Peppard (2005), Huygh & De Haes (2019)) and, more recently, to examine analytics governance, which emphasizes the contextualized output of data utilization (Baijens et al., 2021). The latter authors notably argue that analytics governance is part of a meta-system for the totality of data analytics activities (e.g., data analytics projects). However, data use depends on input data and, consequently, on data governance practices (Aaltonen et al., 2021; Legner et al., 2020). Thus, we argue for data governance – as a separate VSM – because "*the actual work tasks carried out by individuals to curate and set up the data are typically downplayed*" (Parmiggiani et al., 2022, p. 139).

3 Methodology

3.1 Research design

Considering our research question (*How does data governance unfold in multinational companies?*) and our theoretical proposition (that data governance in multinational companies can be observed through the VSM lens), we follow a qualitative research design (Dubé & Paré, 2003). Our study spanned the period from September 2020 to November 2023. It was embedded in a collaborative practice study (Mathiassen, 2002) and informed by insights from focus groups of 17 multinational companies, as shown in Figure 17. To further deepen our analysis, we conducted five in-depth case studies (Yin, 2018).

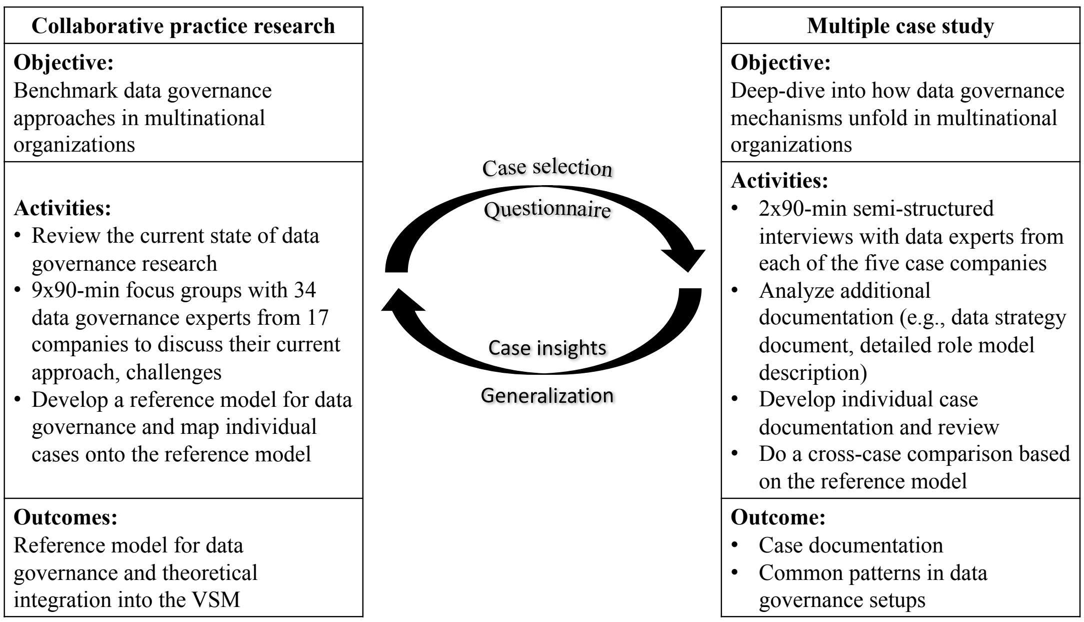


Figure 17. Overview of the research design

3.2 Collaborative practice research

In our collaborative practice research, we partnered with 17 companies seeking to benchmark their data governance approaches. We organized nine 90-min focus groups with 34 high-profile data experts, where participants provided an overview of their data governance approach, as well as describing its evolution over time, which gave all participants a first understanding of their data governance mechanisms. Besides the focus groups, we undertook research activities to review the literature on data governance and develop a reference model as basis for the benchmarking study. This study was used to map and compare individual companies' governance approaches. Using purposeful sampling (Patton, 1990), we identified five companies' data governance approaches for the subsequent case study analysis (see Table 35). Our interactions with the five case companies informed the subsequent focus groups iteratively. The final focus group consisted of 22 data executives from the 17 companies who discussed the findings, i.e., the reference model and the benchmarking study with illustrations from the cases.

4 Case studies

To be able to generalize a VSM, we opted for multiple cases as this supports better analytical generalization (Yin, 2018). We selected companies with diverse characteristics regarding their industry, the goal and scope of their data governance, and different organizational structures. The case companies had implemented federated data governance design decisions, e.g., they had complete role and process models at global and local levels.

Case, Industry	**Revenue/ Employees**	**Key informant**	**Data governance's goal and scope**	**Global data governance**	**Local data governance**
ManufCo *Automotive manufacturing*	$1B–$50B/ ~90,000	VP Data & Analytics Governance	Enterprise-wide data governance on 44 data domains to stimulate data use in all business units and address all strategic areas of digitalization	Data and analytics governance team (13 people) reporting to the Chief Digitalization Officer	Data and analytics coordination in each of the 12 organizational areas, i.e., functions, divisions, regions (100 people in total)
BeautyCo *Adhesives & Beauty products*	$1B–$50B/ ~20,000	Director Master Data & Product Lifecycle	Enterprise-wide master data governance on two domains (products, finance) to improve operational processes and to improve value generation from data	Master data team (35 people) split between business (supply chain, finance) and IT with respective reporting lines	Three regional data hubs close to the markets and overseeing data lifecycle in different countries (25 people)
PharmaCo *Pharma, Chemicals*	$1B–$50B/ ~100,000	Head of Data Framework & Stewardship	Enterprise-wide analytics-driven data governance supporting Analytics & AI innovation, enablement, and solutions	Data Framework and Stewardship (30 people) in Data & AI competence center reporting to Global Digital Services	20 divisional digital offices with about 200 data stewards
EnergyCo *Energy*	$100B-$500B/ ~100,000	Chief Data Officer	Enterprise-wide data governance on 16 data domains defined according to the business model to drive data use into operational processes	Small chief data office focusing on data foundation (5 people)	35 Chief Data Officers allocated to divisions with a small team each and 70 data architects
SoftCo *Software*	$1B-$50B/ ~110,000	VP - Head of Data Management	Enterprise-wide master data governance on two business-critical data domains (products and customers) to improve operational processes	Intelligent data management (IDM) team (98 people) in the Chief Data Office reporting to COO	Three regional hubs (20 people in Europe, APAC, South America), Outsourced (80 people in India)

Table 35. Cases overview

To gain in-depth insight on the five companies' federated data governance approaches, we conducted semi-structured interviews with key informants who had been mandated to oversee enterprise-wide data governance in the case companies. We selected only interviewees who had worked at the company for an extensive period (>3 years), who knew the history of data governance initiatives, and had experienced the issues and challenges associated with implementing data governance, such as involving business stakeholders across different regions and divisions or assigning roles and responsibilities. We designed our interview questionnaire to capture the strategic context and scope for data at the company, and we complemented it with questions that address the three generic data governance mechanisms (see Table 36).

Protocol areas		**Guiding questions**
Strategic context and scope	Strategic context	What are the drivers for data and analytics in the company? Do you have a data and/or analytics strategy? If yes, since when and what is its focus? What business value and benefits do data and analytics create? What are your top five data projects?
	Scope	Which data domains do you distinguish? How do you define them? Which data types are established or emerging? Which data and analytics products do you deliver?
Governance mechanisms	Structural	What organizational form has been chosen (e.g., line function, shared service)? Is the global team/department part of the primary organization and, if so, where is it located in the organizational structure? What are the responsibilities, headcount, structure, and composition of data and analytics teams? Are there any boards and committees for data and analytics? What is their role?
	Procedural	Which data management processes have you established? Which steps/tasks are taken over by the global/ local data organization? Which analytics processes have you established? Which steps/tasks are taken over by the global/local data organization? How do you monitor data governance progress and success? Which metrics do you use and how do you report them?
	Relational	How do you align and collaborate with business stakeholders? How do you align and collaborate with IT stakeholders? How do you align and collaborate between data and analytics? Which data/analytics communities exist? How do you engage with them?

Table 36. Semi-structured interview protocol

Two researchers conducted the interviews via MS Teams video conferencing. Each lasted, on average, 90 minutes as planned. The interviews were recorded and documented according to a pre-filled template structured around guiding questions. After the interviews, we asked the informants to review our interview reports and to confirm our documentation (e.g., key statements), and to address remaining questions. The continuous interaction within the focus groups raised additional data requests, which we addressed in follow-up discussions. After each interview, we performed an additional search for secondary materials that could add to our documentation (e.g., a data strategy document, a detailed role model, the structure of the primary organization), sometimes guided by the expert himself, e.g., to look something up on the company website. To ensure construct validity and reliability of our findings, we triangulated our interview data with further documentation (e.g., company presentations) that we had collected during our research program or from public sources (e.g. presentations at practitioner conferences, annual reports). The final set of data allowed us to obtain granular and complete details on each data governance approach covering all three governance mechanisms. Overall, we obtained a rich case study database built on a chain of evidence composed of primary and secondary data.

In analyzing our data, we applied abductive reasoning because it allows for embedding empirical findings into an existing theoretical model (Ketokivi & Mantere, 2010). This approach facilitated theorization through a detailed examination of the data by employing inductive coding for categorizing interview data and deductive coding for incorporating the VSM perspective. Figure 18 presents the coding process and illustrates the data analysis process with illustrative quotes from one of the cases. First, using inductive coding, the same two researchers labelled the statements following a bottom-up approach to derive open codes (Gioia et al., 2013). Next, they identified relationships, connections, and patterns between open codes, thus bringing a more comprehensive understanding of the underlying concepts. This led them to a set of axial codes reflecting data practices. Last, they used selective coding to derive core themes that describe clusters of these practices. They then used deductive coding to apply the VSM lens. They focused their analysis on assigning data practices for each of the five sub-systems so that they could clarify how the practices are distributed at various levels in the organization. Eventually, they obtained the grouping of the data practices into larger themes that map onto VSM sub-systems.

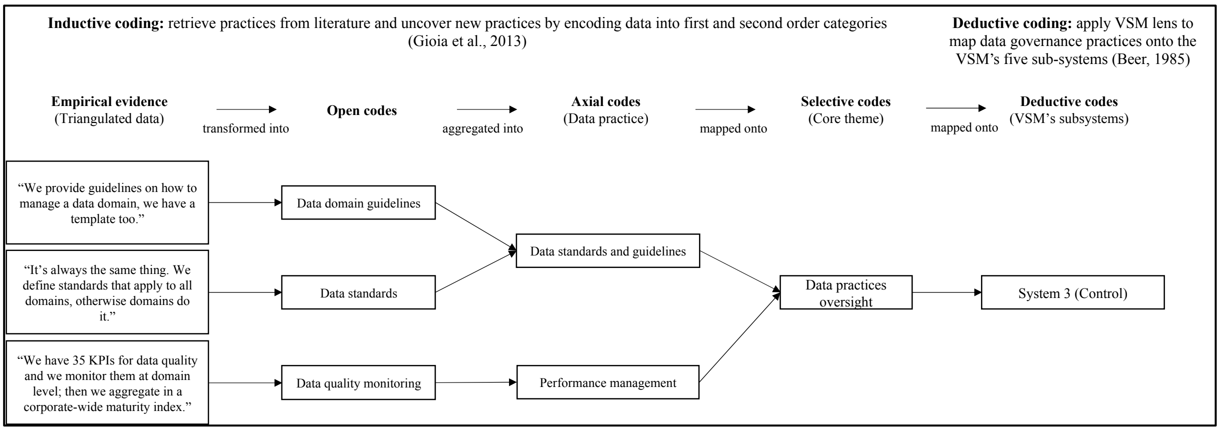


Figure 18. Data analysis process exemplified with quotes from ManufCo case

We did our cross-case analysis in the form of a comparative analysis of the five cases. A cross-case analysis is particularly relevant to this research as it supports the aggregation, simplification, and generalization of complex cases (Miles et al., 2014). For this, we leveraged a granular understanding of each data governance approach. We searched for differences and commonalities between cases by iteratively searching for similarities between codes. From the emerging patterns, we were able to generalize a VSM for data governance by reviewing common data practices (axial codes) and core themes (selective codes) necessary to describe each of the five sub-systems.

5 A viable system for data governance

From our cross-case analysis, we theorized a VSM for data governance that addresses both global and local data governance activities. We find that data governance should occur at multiple, interconnected levels, i.e., in sub-systems (see Table 37): S2, S3, S4, and S5 together form a metasystem of data practices performed in operational units (S1). While S2, S3, and S3* represent the data governance layer (i.e., the data governance teams) and orchestrate data practices, S4 and S5 form the strategy layer (through boards and committees) and shape data governance practices. In the following sections, in describing the different systems, we exemplify the VSM with examples and quotes from our cases (e.g., regional data governance at BeautyCo, divisional data offices at EnergyCo).

5.1 Operations – Perform data practices

System 1 has a set of **operative units** which are typically business functions that embed data in their work practices. These units provide data to their members and to other units, and consume data provided by their members or by other units, for instance in creating dashboards, reports, and increasingly also feeding advanced analytics/machine learning models. Two key data practices enact data provisioning, namely *data creation* and *data curation*. ***Data creation*** involves the intentional and systematic generation of data through various processes, for instance if the account manager in a regional sales team creates a customer record. ***Data curation*** involves the deliberate and systematic maintenance of data throughout its lifecycle to ensure that the data is processed in compliance with regulations and is fit for purpose (data quality). As EnergyCo stated, "*No-one owns the data lifecycle other than the data domains themselves.*" To support operative units, all five cases use shared service centers that handle a part of the data curation tasks, as shown at SoftCo: "*We have a team called 'data operations' that executes data processes. For this, we have a three-level classical shared services setup. We have a follow-the-sun approach with two regional teams in Prague and Manilla, it's about 20 people. We also consider a third offshore team in Brazil. We have a first level outsourced to a consulting company in India, which works with ticketing. There we have another 80 people. It seems really big but actually this is where we provide data maintenance services for all market units at the firm worldwide.*" Business functions are also responsible for addressing data consumption requests and should ensure that data quality follows both standards and consumer expectations. Hence, operative units take ownership of their data and manage data accessibility and data sharing in accordance with data access rights, as articulated by EnergyCo: *"We try to make data discoverable* *for possible usage through our data catalog, Collibra. We have started working with the business to define the key curated data products that we would like to see in place."* ***Data usage*** practice implies that business units use the data for operational and analytics purposes (e.g., in analyzing the data to create a sales forecast). Data can be consumed within the business functions or by outside units that need it to perform their own data analysis or to enrich their own data.

Systems		**Theory** (Beer, 1985)	**Description**	**Key data practices**	**Layer**
System-in-focus	S1	Describes the different operative units that execute the tasks expected to fulfil the system's purpose.	Represents all business units where data practices are embedded in work practices and performed by providers and consumers of data.	• Data creation • Data curation • Data usage	Operations: *Perform data practices*
Meta-system	S2	Handles coordination and communication across the different S1s, especially during disturbances affecting the VSM (e.g., environmental fluctuations).	Ensures coordination between data governance teams by assigning data roles and responsibilities and distributing the latest governance principles to the entire network. It also provides data management support, training, and data applications to data providers and consumers.	• Data roles and responsibilities • Data enablement • Data management support • Data documentation and architecture • Data applications specification	Governance: *Orchestrate data practices*
	S3	Oversees the activities of the system-in-focus (S1) through "day-to-day management" to ensure the smooth delivery of data operations against strategic goals.	Oversees all data practices in the system-in-focus (S1) and ensures that they are performed in line with strategic goals and according to standards and guidelines (e.g., for data collection, storage, use, documentation). Monitors the execution of the data strategy and provides periodic reporting.	• Definition of data standards and guidelines • Performance monitoring and improvement	
	S3 *	Complements System 3 act as a compliance system of operative unit (S1).	Performs data-related audits of operative units to ensure compliance with laws, regulations, and standards.	Data compliance auditing	
	S4	Senses data threats and opportunities to the system by scanning the environment.	Senses data opportunities (e.g., trends) and threats (e.g., compliance) that could impact the data organization.	Data threats and opportunities sensing	Strategy: *Shape data governance practices*
	S5	Maintains the system's identity by describing the system's norms and purpose.	Provides strategic direction for the entire data activities in alignment with company strategy.	Data strategy definition and monitoring	

Table 37. VSM sub-systems and their application to data governance

5.2 Governance – Orchestrate data practices

System 2, taking care of **coordination**, is managed by the data governance team, be it at global or local level. Its role is to communicate about data governance and to coordinate the network of data providers and consumers (S1). Thereby, it ensures alignment at enterprise-wide level, be it between data providers and data consumers within an operative unit (S1), or between several operative units (e.g., in data sharing between customer and sales data domains). We identify five key data practices enacted by S2, which are *definition of data roles and responsibilities*, *data enablement*, *data management support*, *data documentation and architecture,* and *data applications specification*. ***Definition of data roles and responsibilities*** is an established data governance practice that involves defining, assigning, and communicating data-related roles and responsibilities, such as those of data stewards or data editors. This practice also clarifies the role-players' interaction and the collaboration models. For BeautyCo, "*the role definition is central, but the execution happens in regions. For this, we set up the regional hubs and the roles have solid reporting lines to regional offices. But they also have a functional reporting line to me.*" ***Data enablement*** comprises an emerging set of data governance practices focused on empowering individuals and teams to harness the full potential of data by providing the necessary tools and skills. Typically, as the number of employees in data roles grows, increasing data literacy, for instance through training programs, is a priority. Firms also initiate global data culture initiatives, as EnergyCo explains: *"We have a company-wide initiative called 'The year of data,' which is about raising data awareness by showcasing three things: what you can do with data in general, where the company stands and what it struggles with, and what can be done. We also follow up with a data mood survey.*" Executives at ManufCo drive data-driven culture with axioms such as "*Data belongs to all employees, and all can benefit from knowledge of the data*", "*We acknowledge the value of data for the company", "We pay attention to error-free data and thereby guarantee a high level of customer satisfaction."* However, due to the growing business ownership of data, *data enablement* must also reach locally, as BeautyCo states: *"Data enablement is central and regional. In the future, we want most regional hub interactions to have local functions. For instance, our hub in Poland is quite active and does a lot in this instance. They have built their own way of communicating with newsletters and so on. They are very good. We are learning from them."* ***Data management support*** involves all data governance practices aimed at coordinating business and project support (e.g., compliance with data strategy, data needs), coordinating requirements with technical teams (e.g., data engineering), and generally ensuring functional communication across the different units. ***Data documentation and architecture*** practice involves systematically creating and updating comprehensive metadata documentation. Thereby the organization creates transparency regarding its data. This is achieved by designing and evolving the data architecture, and by how data is collected, stored, processed, documented, and used. ***Data applications specification*** aims to define the supporting applications for data provision and consumption. Applications with data governance in scope are typically related to master data management (e.g., SAP MDG), data quality, and data cataloging. As PharmaCo explains: "*I am responsible mostly for the content part. Our task is to translate the technical data into meaningful content. To make the data understandable and consumable for the entire organization, we manage the company-wide data catalog, and along with our divisional stakeholders we are filling it in. We also use a tool to build ontologies and knowledge graphs.*" Governance practices around data applications are performed in collaboration with IT (especially for the platform side). This involves defining the functional requirements, change management, workflows, and UI modelling.

System 3, taking care of **control**, monitors all data practices in S1 and ensures that they are performed in line with strategic goals and according to set standards and guidelines. At the interface of operations and strategy, System 3 plays a pivotal role in standardizing data practices, as well as in strategy delivery and reporting. It displays data governance practices identified as (i) *definition of data standards and guidelines,* and (ii) *performance monitoring and improvement.* The ***definition of data standards and guidelines*** involves creating a data governance framework, developing a local data ownership concept, data process documentation, and data access rights. Control is typically exercised by both global and local data governance teams, as ManufCo highlights: "*Standards and guidelines mainly come from us and are enriched in the specific domains. For instance, we do not give the guidelines for maintaining payment conditions; this is the task of the finance data domain.*" ***Performance monitoring and improvement*** is an emerging data governance practice that pertains to the structured methods and tools an organization employs to monitor, measure, and enhance data quality and data-related processes, through, for instance, maturity assessments. While firms traditionally monitor data quality, they now progressively integrate data consumption in their metrics framework (e.g., in the growing number of data objects available on the data catalog at ManufCo and BeautyCo). At BeautyCo, "*we measure the increasing number of data objects on the data catalog. For success, we measure time-to-market in regional hubs. We also monitor how the number of GTIN violations decreases.*"

System 3*, the **audit**, complements *System 3* by auditing data practices of operative units, thereby ensuring that they agree with legal requirements, industry standards, internal policies, and data standards and guidelines. It is mainly enacted through *data compliance auditing* practices which enforce adherence to rules, regulations, and standards that govern the collection, storage, processing, and sharing of data. For instance, at ManufCo, "*data management is a mandatory, auditable process in the quality management system. To support IT security and data protection, delicate data objects are flagged as sensitive in the data model. Data domains that contain intellectual property are also closely monitored to address potential risks and to initiate risk mitigation.*"

5.3 Strategy – Shape data governance practices

System 4, related to **intelligence**, ensures that the whole system can adapt to disturbances by scanning the environment to detect changes (e.g., new data trends, use cases) and by proposing mitigation plans. It is mainly performed through the strategic practice of ***data threats and opportunities sensing*** and involves actively monitoring, identifying, and responding to potential risks or beneficial situations in the organization's data landscape. This proactive approach enables timely mitigation of threats, such as to data security, and exploiting opportunities, such as new use cases for emerging technologies (e.g., Generative AI). A new local regulation can also impact the data activities, as raised by ManufCo: "*Let's say we want to react to the EU data governance act. It will be discussed in the data council but due to the effects on other enterprise areas we would also put it to the digital coordination council and to the board. We also take it to the global risk and compliance committee, and to several other committees that I am not going to list right now. So, it impacts way more than just data*." For this reason, and due to the critical role System 4 plays in the system's viability, companies might use ambassadors at executive level to help with communication and collective acknowledgement. As SoftCo observed: "*We have a super senior executive coms person in our team. This is one of my biggest assets. Yes, I sit in the organization at level three, but I communicate with everybody, including senior executives and the board. This is sometimes challenging, especially if you want to discuss data topics at a business and strategic level. The role is called 'executive communications lead' and helps us neutralize emotions and politics that come with data topics at strategic level.*"

System 5, dealing with **policy**, provides strategic direction for all data activities aligning with the corporate strategy and business priorities. Strategic data practices introduced here revolve around the enterprise-wide ***data strategy definition and monitoring*** customs and consist in planning, implementing, and optimizing systematic approaches to create value from data. It also involves identifying and assessing the data capabilities required to enable the business model. For instance, the opportunities Industry 4.0 offers and the C-level's recognition of data's strategic value led to ManufCo updating their integrated data and analytics strategy in 2022. In fact, all cases had recently updated their data strategies with a shift toward innovation and value creation from data. As PharmaCo explained: "*We are still working on our enterprise-wide data strategy. In such a big company, this is a long-term project. Our team manages it because it is not about technology; it is about communities, change, culture, this seamless data experience we want to bring. We also propose shifting to a kind of global data office in combination with larger domain responsibilities*."

6 Federated data governance as recursive system

Consistent with existing literature, we note that global firms adopt a federated governance model, albeit with various, sometimes subtle, distinctions. Using the VSM, we find that these federated data governance practices unfold through several systems-in-focus (i.e., multiple System 1). Consequently, global data governance practices can be distributed by being embedded, and often enriched, in local systems which mirror the primary organization's existing regional, divisional, and functional structure. This indicates a recursive logic with two (possibly more, depending on organizational structure) systems-in-focus: (1) at level "n", the totality of corporate data practices governed by global data governance practices, and (2) at level "n+1", local data practices governed by local data governance practices. As shown in Table 38,many data practices enacted at global level are replicated into local data practices, for instance by defining local data strategies, adapting global data guidelines into local ones, and by executing the assignment of local roles and responsibilities. This replication differs from a company to another or may be minimal like in the case of SoftCo. Above, in describing the different systems, we have already exemplified the recursion in the VSM, giving various examples and quotes from our cases (e.g., the regional data governance at BeautyCo, divisional data offices at EnergyCo).

Our analysis disclosed that ManufCo's approach is the most advanced of the cases in that its data operating model covers data governance practices at a global level and on a local level in data domains. ManufCo's VSM displays a patent example of recursion, showing that most data practices enacted in the five sub-systems are replicated into data domains. Data domains emerged from the project "Data Domain Management in all Data Areas," initiated in 2018, where the global data governance team seeked to remove bottlenecks in data use and to establish a network of data roles spread globally (across functions, divisions, and regions). This project triggered an extension of the data scope from key master data objects (such as suppliers and customers) to 44 data domains relevant for digitalization. Examples of these domains are "R&D Engineering", "Sales", "Manufacturing Planning", and "Finance Accounting".

Following here, we clarify the notion of recursion in the VSM by providing a vignette that illustrates ManufCo's federated data governance approach as a critical case (Yin, 2018).

System	Data practices	ManufCo		BeautyCo		PharmaCo		EnergyCo		SoftCo	
		Global	Functions, Divisions and Regions	Global	Regions	Global	Divisions	Global	Divisions	Global	Regional hubs
S2	Definition of data roles and responsibilities	X	X	X	X	X		X	X	X	
	Data enablement	X	X	X	X	X	X	X	X	X	
	Data management support	X		X	X	X	X	X	X	X	X
	Data documentation and architecture	X	X	X		X	X	X		X	
	Data applications specification	X		X		X		X	X	X	
S3	Definition of data standards and guidelines	X	X	X		X	X	X	X	X	
	Performance monitoring and improvement	X	X	X	X	X		X	X	X	
S3*	Data compliance auditing	X									
S4	Data threats and opportunities sensing	X	X	X		X		X	X	X	
S5	Data strategy definition and monitoring	X	X	X		X		X	X	X	

Table 38. Overview of metasystems' data practices for each case

6.1 Recursion of data governance at ManufCo

For ManufCo, we identified that global data governance is recursed into data domain governance, thereby observing two systems-in-focus with each their own metasystem: the totality of data practices and domains' data practices. Each data domain is itself a viable system that strategically self-organizes, independently from other domains. Data domain governance controls and coordinates domain data practices (e.g., data creation, usage, and maintenance in the data domain only), and interacts with the local environment (e.g., correlating function, division, region, and outside world). We show and compare in Table 39 ManufCo's data governance mechanisms corresponding to each meta-system i.e., systems 2-5.

S1		Level "n": totality of data practices	Level "n+1": local data practices
Meta-system	S2	• Network communication channels (e.g., info mails, newsletters, postings on collaboration platforms, corporate data conferences) **(R)** • Feedback loops with roles like Data Analysts, Data Scientists or Data Engineers **(R)** • Regular cross-domain meetings e.g., "Data Domain Manager Round Table", "Special Topics Call" **(R)** • Quarterly communities' coordinator meeting **(R)** • Data literacy training for the whole "IT and digitalization" department with shared modules and specific content for different target groups **(R)**	• Supply chain management (SCM) Data coordinator community **(R)** • "Data Quality Circle": meeting for plant data coordinators **(R)** • Colocation i.e., proximity into the same function or division **(R)** • Data domain specific trainings for data domain managers and their teams (e.g., data management methods, tasks, roles and responsibilities, and relevant processes) **(R)**
	S3	• Central data team works in agile environment with epic owners (no real role model, 10 employees) reporting to Strategic digitalization and IT. For data governance topics, reporting is to CEO **(S)** • Metrics (e.g., 35 KPIs for data quality) for data domain monitoring and support **(P)** • Data governance integrated into firm's digital maturity index with drill-down on function, division & region (e.g., data model definition ratio) **(P)** • Company-wide data modelling approach, incl. definition of (meta) data models and evaluation of requirements for relevant tools, incl. alignment with other operational models (e.g., semantic models) **(P)** • IT security and data protection: sensible data objects / data attributes are flagged in the data model to enable further analysis **(P)** • Intellectual Property: documentation indicates which data domains contain intellectual property to address potential risks and initiate risk mitigation of risks **(P)**	• Decentral data governance roles in 47 data domains across functions and division **(S)** • Data domain role model e.g., one data domain manager role for each data domain (about 40 in total due to cross-functional domains) **(S)** • Scaling is enabled thanks to decentral data coordinators (about 200) **(S)** • Domain specific processes e.g., Data request, Data Life Cycle, and Data Quality Assurance **(P)** • Data quality monitoring (e.g., data domain's data quality index) • Authorization management: Data domain manager implemented as the authority for changes in authorization concepts **(P)**
	S3*	• Data management is a mandatory, auditable process in the quality management system **(P)**	
	S4	• Data Council for project prioritization, oversight and alignment, and governance implementation issues **(S)** • Data strategy planning and control process **(P)** proactive identification, and management of data monetization opportunities **(P)** • Decision-making for data domain creation **(P)**	• Strategy for each data domain **(P)** • Data domain description, documentation and sharing **(P)**
	S5	• Digital Transformation Council (DTC) meets twice a year for topics such as strategy, vision, mission, purpose, claim, and core KPIs **(S)** • DTC Executives drive data driven culture with axioms e.g., "Data belongs to all employees, and all can benefit from its knowledge" **(R)**	• Data quality specific axioms e.g., "We acknowledge the value of data for the company. We pay attention to error-free data and thus guarantee a high level of customer satisfaction" **(R)**
Type of data governance practice: **(S)**= Structural, **(R)**=Relational, **(P)**=Procedural			

Table 39. Comparison of meta-systems for systems-in-focus "n" and "n+1" at ManufCo

6.2 Data governance at ManufCo: subsystems description

Figure 19 shows the role of each sub-system and highlights how the key structural mechanisms (e.g., boards, teams, roles) that enact the corresponding data practices can be mapped onto them. Next, we describe the five sub-systems, thereby showing the interplay between data strategy, data governance, and data operations.

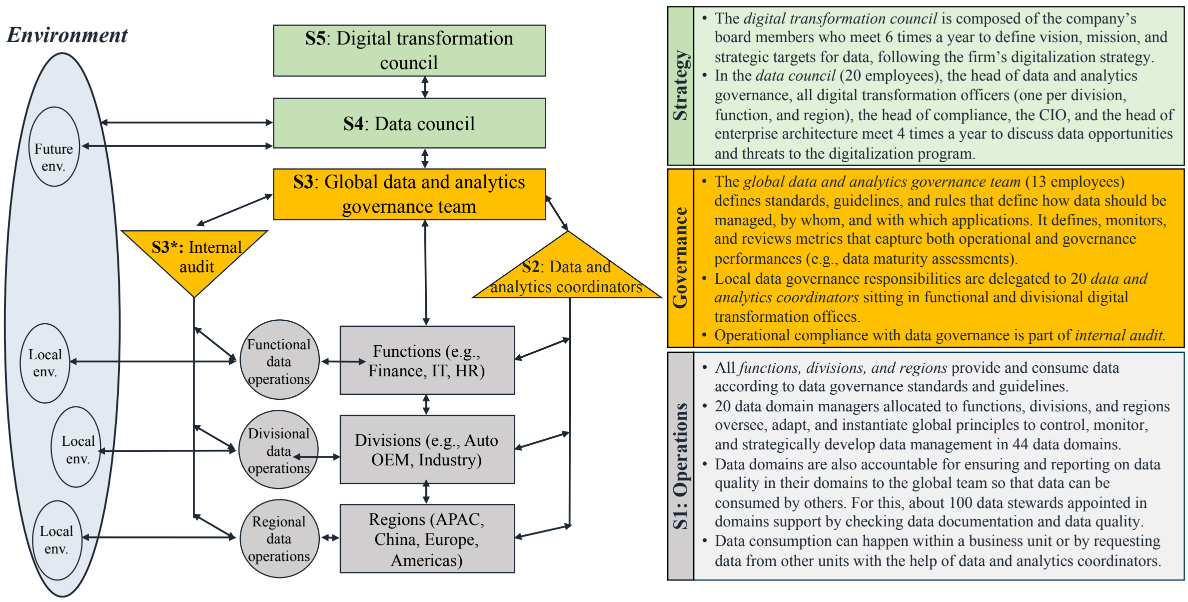


Figure 19. Viable System Model for data governance at ManufCo

System 5 is enacted through the Digital Transformation Council (DTC). It oversees the 2021 "digital agenda", defined by company board members, which aims to securing the company's long-term competitiveness, marking a paradigm shift in the role of data, which now forms a "core component of value creation." It is composed of six company board members who meet bi-annually to monitor the progress of the so-called "digital agenda," which is the digital arm of the company's strategic goal to be the technology leader in the "mobility of tomorrow." More specifically, the DTC aims to secure the company's long-term competitiveness through a paradigm shift in the role of data, which now forms a "core component of value creation." Concretely, having discussed data vision (e.g., data monetization), key drivers (e.g., data democratization, data economy), and associated KPIs, the DTC formulates the data strategy.

System 4 is enacted through the Data Council, the organizational body responsible for the underlying data-related activities, their prioritization, oversight and alignment, and possible implementation issues (S4). It is composed of 20 members that include the head of data and analytics governance, all digital transformation officers (one per division, function, and region), the head of compliance, the CIO, and the head of enterprise architecture. During quarterly meetings they discuss topics such as how to react to new regulations (e.g., the EU data act), or how data can support the different business processes in creating business value (e.g., where to find trustworthy data, what count as dependencies, as key vocabulary, as important security and privacy aspects, and as business processes' requirements). The data council also manages the data domains portfolio that currently includes 44 data domains. Eligibility questions for opening new data domains typically include: Which business processes or other principles would justify a new domain? Does the corresponding function or division generate its own data (e.g., specific data entries)? Would the domain be temporary or sustainable in the long term? Would data be useful in all departments? Would the domain be cross-functional? Are there synergies with other domains that could justify an integration/merger? Would setting up a regional satellite for this domain be wise? Based on lean templates, each data domain's profile is documented (e.g., in a description of its content and data objects, sensitive data, relevant business processes).

S*ystem 3* is enacted through the global data and analytics governance team. Composed of 13 experts who control all corporate data operations, the team provides the general data standards and guidelines applicable to all domains, and it monitors various metrics to demonstrate progress on the data strategy, such as data quality improvements, the data documentation rate, data tools use, and data literacy assessment. The team also gets support from internal audits to assess various data domains' compliance with global standards and guidelines.

System 2 is enacted by 20 data and analytics coordinators who act as counterpart in division, functions, and regions, who sit in the respective digital transformation offices. They communicate the global standards provided by *System 3* to all domains. This way, the entire network builds knowledge of the data strategy, data roles and responsibilities, data processes, data applications, data model, and data quality. Further, they provide data management support, for instance by coordinating data provisioning and data consumption requests across operational units. This is facilitated by the "Data Domain Manager Round Table" that enables cross-domain practice exchange.

System 1 represents all operational data practices across functional, divisional, and regional units. Each corresponding data domain takes ownership for creating, curating, and using their data or using other domains' data. Recursively, in each domain, *data domain managers* adapt global principles and define their own data domain principles, i.e., they control, monitor, and strategically develop data management in their data domain. These managers are also accountable to the global team for data quality in their domains, ensuring the quality and reporting on it. For instance, the Finance Accounting domain gets contributions from other domains, e.g., gaining inventory data that belongs to the Storage and Shipping domain or costing data that belongs to the Sales and Marketing domain. The *regional data domain managers* are responsible for coordinating the data domains in a given region, thus linking the data domain manager to the operative business units (e.g., helping to define the access authorization concept in compliance with local regulations). *Data stewards* support the data domain manager in documenting data (e.g., metadata) and maintain data quality in each domain by integrating business knowledge in data curation tasks. They are also responsible for responding to data users' data access requests, in- or outside the domain.

Since implementing their federated data governance model in 2021, ManufCo has observed substantial improvements in business performance. The duration of both the "Initial Order" and the "Request for Quotation" processes were significantly reduced thanks to data quality improvements. Further, the recorded cycle time of the business partner approval process was reduced by 30%, the cycle time of intercompany service requests by 45%, and ManufCo could achieve a striking 97,3% duration reduction in all processes within the 24-hour timeframe agreed in the service level agreement. These results show the critical role of data governance in enabling innovative local data practices.

7 Contribution, discussion, and implications

Although the foundations of data governance have reached a sound level of clarity, much of the research to date remains conceptual and proposes generic, static mechanisms. This study is among the first to focus on the implementation of data governance mechanisms and their adaptation in large and complex organizations. Our results explain how data governance unfolds in practice in multinational companies through a viable system composed of multiple, interconnected levels, i.e., sub-systems with their own sets of data practices. The application of the VSM in this study demonstrates that data governance cannot be seen only as a static framework that shapes structural, procedural, and relational mechanisms; rather, it needs a dynamic framework that supports the expansion of data practices in all areas of the organization. This is in line with Vial (2023, p. 6) who stresses that "*the instantiation of this design in practice is important to understand how an organization protects and leverages data for digital innovation.*" Overall, the use of the VSM supports a better understanding of such seemingly paradoxical activities by explicating both the dynamics of control (e.g., data protection) and the dynamics of value creation (e.g., from data use).

Our results confirm and extend prior research, arguing that global firms adopt federated (also called hybrid) models for data governance (Grover et al., 2018; King, 1983). Through the lens of the VSM, we show how companies thoughtfully merge and maintain global responsibilities, such as universal standards, protocols, and methodologies, with local responsibilities that are uniquely tailored to individual business units, including data quality monitoring and project execution. This model involves transferring certain data governance responsibilities from the global data governance unit and assigning new responsibilities to local roles in business (e.g., data steward). Data access is mainly managed by business experts (i.e., data managers) themselves, following corporate policies set by the global data team (*System 3*). This obliges the audit function (*System 3*)* to take on additional responsibilities that will mitigate data management risks. Overall, while global data governance fosters uniform enterprise-wide data management principles, standards, and methods, federated data governance practices favor local business expertise. *System 2* is then crucial for cross-functional projects and network enablement.

We find that data governance practices are enacted according to an organizational hierarchy, thus not at the same level. The recursion highlighted in our VSM demonstrates that federated data governance is enacted through a cascading system that assign data governance responsibilities across multiple hubs typically aligned on the organization's primary structure (e.g., corporate, functional, regional). This model further branches out through "spokes," representing the data creators and users within the business, ensuring that governance reaches all levels of operation. Hence, unlike a hub-spoke model that centralizes data governance responsibilities at a corporate level, hub-hub-spoke models, which can embed more than just one recursion, offer numerous benefits such as greater local autonomy, use of domain knowledge, faster issue resolution, and improved agility. For their respective sectors, each hub sets strategic data objectives, defines data standards and guidelines, creates transparency on data, and fosters data enablement. In return, a hub-hub-spoke model generally requires better coordination mechanisms (e.g., a data council, data communities, local hub monitoring). However, coordination mechanisms (*System 2*) generally "*do not arise prior to coordinating but are constituted through coordinating*" and they typically follow a system's disruption (Jarzabkowski et al., 2012, p. 907). This highlights the pivotal role of environment sensing on both a corporate and on local levels to continuously update data coordination mechanisms. For instance, the strategic need to develop new analytics use cases (e.g., Generative AI) might enlarge the scope of data governance (e.g., extending to new data types) and trigger an update on the role and board model. Future research could investigate hub-hub-spoke models in greater detail, and especially how they unfold into different organizational structures. In this regard, the study of global-local coupling in federated data governance systems could be an interesting starting point, for example, by examining the impact of external turbulences based on the responsiveness and specificity of the system in focus (Orton & Weick, 1990; Weick, 1976). This avenue could investigate how to build modularity, the right level of redundancies, adaptability, and resilience into federated models.

From an academic perspective, the VSM perspective paves the way for investigating data governance from a new angle. It contributes to the previously neglected dynamic nature of data governance and addresses the need to investigate data governance in practice (Vial, 2023). The insights developed in this study further provide valuable guidance on how to design the organizational counterpart to technical data mesh principles by showing, for instance, how different business units enact ownership of their data. Besides data creator and data user roles, our study shows that data steward and data (domain) owner roles, which are seldom clearly distinguished and are often misunderstood (Vial, 2023), are essential to the execution of domain-level data governance practices thanks to their knowledge of the business context. Future research could further investigate the interaction between the technical architecture and the operating model for data governance, especially considering the difficulty of knowledge integration and the data literacy gap between business and analytics teams (Kollwitz et al., 2018;

Someh et al., 2023). From a practical perspective, our findings support decision makers in global firms to define, adapt, and implement data governance. They can leverage the VSM to build their own federated data governance framework, that addresses both global and local levels.

Since this study takes a new, systems theory approach to examining data governance, it is inherently prone to first mover limitations, and we strongly encourage future research in this area. Beyond the potential future research activities mentioned above, the understanding of data governance as a self-organizing system could be further deepened. As this study focused mainly on elucidating the five sub-systems, our findings also open avenues for further research into how antecedents affect data governance as a system. For instance, researchers could investigate how different industries' strategies and operating environments impact the system's viability. In such a case, certain principles from VSM theory, like variety and transduction (Beer, 1985) with which this paper could not deal extensively, provide interesting possibilities for refining the model.

8 References

Aaltonen, A., Alaimo, C., & Kallinikos, J. (2021). The Making of Data Commodities: Data Analytics as an Embedded Process. *Journal of Management Information Systems*, *38*(2), 401–429.

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.

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: An analysis of the literature. *Journal of Decision Systems*, *25*(sup1), 64–75.

Baijens, J., Huygh, T., & Helms, R. (2021). Establishing and theorising data analytics governance: A descriptive framework and a VSM-based view. *Journal of Business Analytics*, 1–22.

Beer, S. (1985). *Diagnosing the system for organizations* (Repr). Wiley.

Benfeldt, O., Persson, J. S., & Madsen, S. (2020). Data Governance as a Collective Action Problem. *Information Systems Frontiers*, *22*(2), 299–313.

Brown, C. V. (1999). Horizontal Mechanisms under Differing IS Organization Contexts. *MIS Quarterly*, *23*(3), 421.

Dubé, L., & Paré, G. (2003). Rigor in Information Systems Positivist Case Research: Current Practices, Trends, and Recommendations. *MIS Quarterly*, *27*(4), 597–636.

Fadler, M., Lefebvre, H., & Legner, C. (2021, June 14). Data governance: From master data quality to data monetization. *Proceedings of the 29th European Conference on Information Systems (ECIS), An Online AIS Conference*.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods*, *16*(1), 15–31.

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.

Günther, W.A. *et al.* (2022) 'Resourcing with data: Unpacking the process of creating data-driven value propositions', *The Journal of Strategic Information Systems*, 31(4), p. 101744.

Huygh, T., & De Haes, S. (2019). Investigating IT Governance through the Viable System Model. *Information Systems Management*, *36*(2), 168–192.

Jarzabkowski, P. A., Lê, J. K., & Feldman, M. S. (2012). Toward a Theory of Coordinating: Creating Coordinating Mechanisms in Practice. *Organization Science*, *23*(4), 907–927.

Jones, M. (2019). What We Talk about When We Talk about (Big) Data. *The Journal of Strategic Information Systems*, *28*(1), 3–16.

Ketokivi, M., & Mantere, S. (2010). Two strategies for inductive reasoning in organizational research. *Academy of Management Review*, *35*(2), 315–333.

Khatri, V., & Brown, C. V. (2010). Designing Data Governance. *Communication of the ACM*, *53*(1), 148–152.

King, J. L. (1983). Centralized versus decentralized computing: Organizational considerations and management options. *ACM Computing Surveys*, *15*(4), 319–349.

Kollwitz, C., Mengual, P. M., & Dinter, B. (2018). Cross-Disciplinary Collaboration for Designing Data-Driven Products and Services. *In Proceedings of the Pre-ICIS SIGDSA Symposium on Decision Analytics Connecting People, Data & Things*, 16.

Lefebvre, H., & Legner, C. (2022). How Communities of Practice Enable Data Democratization Inside the Enterprise. *Proceedings of the 30th European Conference on Information Systems (ECIS)*, 16.

Legner, C., Pentek, T., & Otto, B. (2020). Accumulating Design Knowledge with Reference Models: Insights from 12 Years' Research into Data Management. *Journal of the Association for Information Systems*, *21*(3), 735–770.

Machado, I., Costa, C., & Santos, M. Y. (2021). Data-Driven Information Systems: The Data Mesh Paradigm Shift. *Proceedings of the 29th International Conference On Information Systems Development*, 1–6.

Mathiassen, L. (2002). Collaborative Practice Research. *Information Technology & People*, *15*(4), 321–345.

Mikalef, P., Boura, M., Lekakos, G., & Krogstie, J. (2020). The role of information governance in big data analytics driven innovation. *Information & Management*, *57*(7), 103361.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2014). *Qualitative Data Analysis: A Methods Sourcebook* (3rd ed.). SAGE Publications.

Orton, J. D., & Weick, K. E. (1990). Loosely Coupled Systems: A Reconceptualization. *Academy of Management Review*, *15*(2).

Otto, B. (2011). Data Governance. *Business & Information Systems Engineering*, *3*(4), 241–244.

Parmiggiani, E., & Grisot, M. (2020). Data curation as governance practice. *Scandinavian Journal of Information Systems*, *32*.

Parmiggiani, E., Østerlie, T., & Almklov, P. G. (2022). In the Backrooms of Data Science. *Journal of the Association for Information Systems*, *23*(1), 139–164.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. SAGE Publications, inc.

Peppard, J. (2005). The Application of the Viable Systems Model to Information Technology Governance. *In Proceedings of the 26th International Conference on Information Systems (ICIS)*.

Sambamurthy, V., & Zmud, R. W. (1999). Arrangement for Information Technology Governance: A Theory of Multiple Contingencies. *MIS Quarterly*, *23*(2), 261–290.

Someh, I., Wixom, B., Davern, M., & Shanks, G. (2023). Configuring Relationships between Analytics and Business Domain Groups for Knowledge Integration. *Journal of the Association for Information Systems*, *24*(2), 592–618.

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.

Velu, C. K., Madnick, S. E., & van Alstyne, M. W. (2013). Centralizing Data Management with Considerations of Uncertainty and Information-Based Flexibility. *Journal of Management Information Systems*, *30*(3), 179–212.

Vial, G. (2023). Data governance and digital innovation: A translational account of practitioner issues for IS research. *Information and Organization*, *33*(1), 100450.

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.

Weick, K. E. (1976). Educational Organizations as Loosely Coupled Systems. *Administrative Science Quarterly*, *21*(1), 1–19.

Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6th ed.). SAGE.