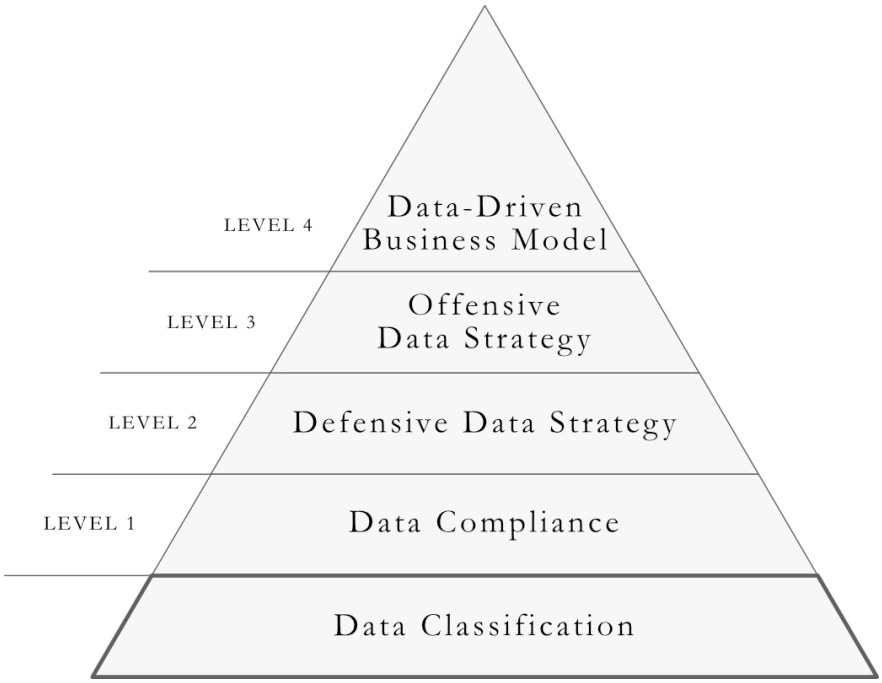
CHALMERS
UNIVERSITY OF TECHNOLOGY



From Strategy to Execution

Bridging the Gap between Data Strategy and Data Governance

Master's thesis in Intellectual Capital Management

ADELINE FREDRIKSSON
JESPER HAGBERG

DEPARTMENT OF TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF ENTREPRENEURSHIP AND STRATEGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2023
www.chalmers.se
Report No: E2023:074

Report number: E2023:074

From Strategy to Execution

Bridging the Gap between
Data Strategy and Data Governance

Adeline Fredriksson
Jesper Hagberg

Department of Technology Management and Economics
Division of Entrepreneurship and Strategy
Chalmers University of Technology
Gothenburg, Sweden 2023

From Strategy to Execution
Bridging the Gap between Data Strategy and Data Governance
Adeline Fredriksson
Jesper Hagberg

Supervisor: Bowman Heiden
Examiner: Ulf Petrusson

Master's Thesis 2023
Department of Technology Management and Economics
Division of Entrepreneurship and Strategy
Chalmers University of Technology
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Cover: Levels of strategic maturity in regard to data asset management.

Typeset in LaTeX
Printed by Chalmers Reproservice
Gothenburg, Sweden 2023

Abstract

Almost a decade ago, the term 'big data' emerged to describe a new and advanced stage of digitalization. Fast forward nearly ten years, and it is widely acknowledged that companies should seek to utilize data as a key mechanism for creating competitive advantage. Doing so is, however, easier said than done. Practical knowledge of data management remains limited, even as organizations generate, acquire, and process unprecedented amounts of data. While companies skilled at managing data outperform competitors by wider and wider margins, the legal landscape relating to data is growing increasingly complex by the minute. This makes data strategy, governance, and management vital for companies across all sectors - whether traditional industry giants or cutting-edge digital platforms.

Through a comparative multiple case study, this thesis explores how the gap between data strategy and data governance can be bridged, ultimately facilitating the execution of a firm's data strategy. The thesis examines three cases to uncover the governance mechanisms firms employ to manage their data assets, the varying levels of strategic maturity companies exhibit concerning their data assets, and how data governance supports the implementation of data strategies. The cases represent different levels based on identified strategic maturity, enabling a comparison of practices. Additionally, by adopting an interdisciplinary approach, the thesis explores how regulatory requirements can be integrated into strategy and governance structures to ensure compliant data management while facilitating data-driven digital innovation.

The findings show that strategic data maturity can be categorized into four distinct levels underpinned by the foundational process of data classification. Reviewing each individual level from a governance perspective has shown key characteristics of the individual levels. From this, specific governance mechanisms crucial for implementing each level of strategic maturity emerged. The study also demonstrates that perceptions of regulation vary across the different strategic maturity levels. This highlights that strategically mature firms possess the expertise to incorporate compliance into their strategy and governance structures, allowing them to maintain both compliance and innovation in managing their data assets.

Keywords: data assets, data governance, data strategy, big data, data management, compliance, artificial intelligence, digital innovation

Acknowledgements

This thesis represents the result of a master thesis internship conducted in the spring of 2023. After countless iterations, several late nights of work, and lots of laughter, we can finally say: Ladies and gentlemen, we have a master thesis! We want to thank many people, but some deserve a special mention.

First and foremost, thank you to our supervisors at the company, Katarina Wendin, and Marta Sadriu, for your exceptional support throughout the past six months. Your guidance and expertise have been invaluable, and we are truly grateful for your commitment to our thesis project. We will never again forget about stakeholder management!

We would also like to sincerely thank all the interviewees who generously shared their insights, experiences and valuable time with us. Without your contributions, this thesis would not have been possible. Also, special thanks to Suzanne S. Harrison and Patrick H. Sullivan, whose book *Edison in the Boardroom* inspired our approach to the chosen topic and whose hierarchical model of intellectual asset management provided valuable ideas for the design of our final model.

Thank you to our ICM class for the very short periods of mañana mañana and the otherwise great mix of nomenclature and fun. You are some of the brightest people we have ever met, and we look forward to starting a company together in five years!

Lastly, thank you to the ICM faculty for demonstrating an unprecedented commitment to our education! Thank you, Bowman Heiden, for the experience of 'learning by trauma.' Also, thank you for your support in supervising this thesis, for granting us access to your extensive network, and for generously providing us with opportunities all over the world. Thank you, Anna Holmberg Borkmann, for your unrivaled energy and for teaching us the value of post-its. Thank you, Christoffer Hermansson, for teaching us to claim the shit out of everything. Lastly, thank you, Ulf Petrusson, for founding the ICM education and perfecting it over the years – at last, we do follow you on this.

Adeline Fredriksson & Jesper Hagberg
Gothenburg, May 2023

List of Abbreviations

Below is the list of abbreviations that have been used throughout this thesis listed in alphabetical order:

AI Act	Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (COM/2021/206 final)
BIPA	Biometric Information Privacy Act, 740 Illinois Compiled Statutes (ILCS), §§ 14/1 - 14/99 (2008)
CCPA	California Consumer Privacy Act, CA Civ Code § 1798.192 (2018)
Data Act	Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) (COM/2022/68 final)
Database Directive	Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases
Digital Markets Act	Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act)
DPO	Data Protection Officer
GDPR	Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)
IS	Information Systems
Trade Secrets Directive	Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure

Contents

List of Figures

List of Tables

1

Introduction

This introductory chapter outlines the background of the research topic and presents the problem statement addressed in this thesis. Further, the research questions that serve as the framework for the conducted examination are presented. Lastly, this chapter describes the purpose of the study and its delimitations.

1.1 Background

Almost a decade ago, the term 'big data' was coined to describe a new and advanced stage of digitalization (Brynjolfsson & McAfee, 2014; Constantiou & Kallinikos, 2015; Varian, 2014). Big data was believed to be the key to creating a range of innovative services. Nearly ten years later, on March 21st, 2023, Bill Gates commented on Open AIs Chat GPT, and stated that "*the age of AI has begun*" (Gates, 2023). AI systems such as Chat GPT are trained on vast amounts of data characteristic of the 'big data' phenomenon. Such systems are the latest embodiment of the innovation made possible by the immense amounts of data generated by modern digital technologies.

While data plays a crucial role in these systems, the responsible companies are generally hesitant to reveal the precise origins of the data (Nield, 2023). However, indications point to sources like Wikipedia, public forums, Q&A sites, and tutorials (Thoppilan et al., 2022). This suggests that AI models rely heavily on these sources to obtain data without compensating those who created the resources. As a response, communities such as Reddit and StackOverflow have expressed intentions to introduce fees to access their material (Dave, 2023; Vigliarolo, 2023). At the time of writing, this is probably the most recent, real-life example of how data-driven technologies can cause disruptions in established business practices and open up new paths for economic growth, requiring businesses to adapt accordingly (Gokalp et al., 2016).

Simultaneously, data has become of vital interest to regulators, as evident by the introduction of privacy regulations such as the GDPR in the EU and the CCPA and BIPA in the US. The introduction of the EU AI Act and the EU Data Act, expected to enter into force in 2023 and late 2024, respectively, exemplify how regulators are now moving outside the scope of personal data and extending their focus to data in general. Thus, the complexity of the legal landscape related to data only stands to increase.

While AI currently is the most visible manifestation of the potential of data-driven innovation, there are innumerable other opportunities to be explored. Such opportunities do not only relate to groundbreaking AI solutions. Instead, the ability to leverage data affects numerous major indicators across businesses, such as revenue growth, sustainability, lead times, and attrition of employees (Franke et al., 2023). Companies with these abilities outperform competitors by larger and larger margins, turning data strategy, governance, and management into key activities for companies across all sectors, from the most traditional industry entities to the most innovative digital platforms. Failure to do so could lead to competitive disadvantages, regulatory risks, and lost business opportunities.

1.2 Problem Statement

Despite the heightened attention given to data, executives at even the most progressive firms continue to encounter difficulties in utilizing data to its fullest potential across a broad range of endeavors (Franke et al., 2023). The complexity of proficiently leveraging data is often attributed to the dual nature of the objectives sought through data-driven activities. Data-oriented initiatives often prioritize two inherently paradoxical goals, namely risk management and value creation (Davidson et al., 2023; Vial, 2023). Organizations tend to prioritize one of the two objectives over the other, failing to realize the full potential of data (Black et al., 2023; Davidson et al., 2023). Consequently, technological and legal developments pressure firms to develop dynamic capabilities to manage data in a way that facilitates digital innovation in an increasingly complex legal reality. Failure to address these challenges can lead to severe consequences, ultimately threatening the success of the business case (Franke et al., 2023). Thus, as Constantiou & Kallinikos (2015) and Plotkin (2020) argue, firms must understand how data assets relate to current business strategies and how to translate strategic data objectives into organizational action. How to go about that is the overarching topic of this thesis.

1.3 Project Purpose

This thesis will analyze how data assets are governed depending on the maturity of the data strategy employed at the organization in question. The thesis project will include an analysis and investigation of three cases. Thus, the research provides a comparative analysis of how data assets are governed depending on the maturity of the data strategy employed in different cases. This project aims to develop a framework for assessing the maturity level of an organization's data strategy, as well as identifying the corresponding practical mechanisms needed to support that strategy.

This entails establishing a connection between data strategy and data governance – an attempt that has not yet been undertaken to the best of our knowledge. By doing this, we bridge the gap between the high-level considerations associated with data strategy and the practically oriented realm of data governance. This enables a comprehensive overview of managing data assets, both from a broader 'what'-perspective, which concentrates on defining the overarching goal and strategy, and from a more detailed 'how'-perspective, which focuses on actionable steps to achieve that goal. By connecting the two, this project seeks to provide a roadmap outlining the key considerations necessary for transitioning from one maturity level of data strategy to another, along with the mechanisms that must be implemented to facilitate such a shift. The intention is that this will also contribute to the ongoing debate on managing the opposing goals of value creation and risk minimization in a conscious manner, where the trade-offs are intentional and minimized.

1.4 Research Questions

Below, the main research question and the sub-research questions used to answer the main research question are presented and briefly explained.

Main Research Question

How do firms govern data assets depending on the maturity of their data strategy?

The main research question aims to examine the variances in data governance practices depending on the maturity level of the data strategy implemented in a given case. By doing so, this study seeks to establish a correlation between the higher-level data strategy and the more practical perspective of data governance.

Sub Research Question 1

How are data assets governed across firms?

This question seeks to assess which governance mechanisms that are utilized by organizations. This assessment is conducted by examining the roles, responsibilities, relationships, and documentation involved in governing data.

Sub Research Question 2

What different maturity levels of data strategies exist?

This question aims to delineate the various data strategies employed, along with their defining characteristics, to establish distinct levels of maturity for each data strategy.

Sub Research Question 3

How does data governance support the objectives of data strategies on different maturity levels?

This question seeks to define the correlation between data governance practices and each level of maturity in data strategy, which is achieved by consolidating the findings from Sub Research Questions 1 and 2. This study will provide an overview of the governance mechanisms required, along with the objective or purpose thereof, to execute the data strategy at each level of maturity effectively. Furthermore, this question will provide a basis for outlining the practical measures necessary to transition from one level of maturity to another.

1.5 Delimitations

This thesis will be limited to examining three cases, two at the project level and one at the company level. For a more detailed description of the cases and the considerations made concerning them, please refer to Chapter 3.

This thesis allocates a significant portion of its content to analyzing the present and future regulatory frameworks related to data. The legal frameworks considered in this thesis are confined to the jurisdiction of the EU. Given the authors' competence and the time constraints of this project, this limitation is deemed appropriate and necessary. However, regulations beyond the EU jurisdiction may be considered in a complementary manner to emphasize the significance of particular legal considerations from a global perspective.

This thesis partially builds on the conceptual framework for data governance proposed by Abraham et al. (2019). This framework encompasses six dimensions: governance mechanisms, organizational scope, data scope, domain scope, antecedents, and consequences. In this project, we only focus on governance mechanisms, which represent the core dimension of the framework and encompass structural, procedural, and relational mechanisms. The other dimensions are not considered.

1.6 Previous Research

In the initial stages of the research project, a literature review was conducted to map previous research related to the thesis topic. As the fields this thesis addresses are relatively young and highly active, relevant publications have emerged throughout the project. Consequently, the body of literature has been continuously reevaluated as warranted throughout the project. This thesis primarily builds on two academic fields: data strategy and data governance.

Regarding data strategy, McAfee et al. (2012) were early to implicitly address the topic by examining how the ability to exploit big data can improve company performance. Constantiou and Kallinikos (2015) build on this and discuss how big data attributes challenge established rules of strategy making. Dallemule and Davenport (2017) focus on trade-offs between 'defensive' and 'offensive' data strategies and between control and flexibility in their use. Medeiros et al. (2020) build on the work of Dallemule and Davenport (2017) to analyze how data strategy affects the achievement of competitive advantage. Mazzei and Noble (2017) identify how big data improves organizations' functional capabilities, shapes new industries, and is critical to innovative and disruptive strategies. Wallis (2021) presents a practitioner's view on planning, developing, and implementing a data strategy.

Concerning data governance, Weber et al. (2009) transfer concepts from IT governance and organizational theory to the, in their own words, previously largely ignored field of data governance. Thereby, they start a scientific discussion on the topic. A year later, Khatri and Brown (2010) presented the first overall framework for data governance, aimed at providing researchers with tools to focus on critical data governance issues and practitioners with tools to develop an effective data governance approach, strategy, and design. Tallon et al. (2013) build on the theory of IT governance, which focuses on the governance of physical IT artifacts, to uncover the structures and practices used to govern information artifacts, i.e., data. Alhassan et al. (2016, 2018, 2019) explore and compare literature on data governance to provide a comprehensive analysis of the activities involved in data governance and outline the critical success factors for such activities. Al-Ruithe et al. (2018, 2019) use systematic literature reviews to synthesize state-of-the-art research in data governance and develop a taxonomy that defines the different attributes of data governance.

Abraham et al. (2019) synthesize the literature and presents a conceptual framework for data governance, which this thesis builds on to a large extent. Gupta and Cannon (2020), Plotkin (2020), and Ladley (2012, 2020) provide practitioners' views on how to implement data governance in an organization. In the most recent literature on the topic, Black et al. (2023) focus on the role of data governance in the secondary use of data. Davidson et al. (2023) reflect on established foundations in data governance research and highlight possible future directions for scholarship on data governance across multiple levels to enhance digital innovations for transformation and societal good. Vial (2023) identifies practitioner data governance issues and translates these into proposed research themes. One of the issues addressed by Vial (2023) is the overemphasis often put toward data governance based on compliance and control at the expense of value creation and innovation, which has inspired this thesis.

2

Theoretical Background

This section details the theories used to answer the research questions. It includes information from relevant literature on both data strategy and data governance. Additionally, it explores legal frameworks surrounding data, including the GDPR, the AI Act, and the Data Act.

2.1 Data Strategy

Strategy scholars have long been interested in data as information concerning the strategic fit of firms relative to their environments (Constantiou & Kallinikos, 2015). Models such as Porter's five forces and competitor analysis are both examples that aim to use external data to create a snapshot of the competitive environment to provide a basis for strategic decision-making (Constantiou & Kallinikos, 2015; Porter, 1979). Additionally, the resource-based view highlights the strategic importance of collecting internal and external data (Constantiou & Kallinikos, 2015). Since competitive advantage is considered to stem from resources controlled by the firm (Barney, 1991), identifying such resources is necessary to understand the firm's competitive position. More recently, the need for collecting both external and internal data to aid decision-making has been highlighted, or at least strongly inferred, by Teece (2007) and his work on dynamic capabilities which emphasizes the importance of understanding both the internal and external landscape. The necessity to incorporate data in strategy-making has only grown with the emergence of 'big data,' which McAfee et al. (2012) dubbed a 'management revolution.'

By some scholars, big data is also considered to change the relationship between data and strategy as a whole (Constantiou & Kallinikos, 2015; Lee et al., 2014). Constantiou and Kallinikos (2015) argue that big data brings about shifting business parameters that challenge the deductive approach to data gathering championed by traditional models of strategy-making. Instead of data being collected reactively to serve strategic decision-making, decisions are made about what already exists and is

available as data. This view is shared by Mazzei and Noble (2017), who argue that whereas strategy used to guide what data to collect and use, it is becoming more common that the data available to a firm guide the strategy. This shift is visualized below in Figure 2.1.

Figure 2.1: *Old and New Perception of the Relationship between Strategy and Data. Adapted from Mazzei and Noble (2017).*

To understand big data as something that reaches beyond improving traditional firm capabilities, Mazzei and Noble (2017) propose a three-tiered framework for how data creates value for firms. This is outlined in Table 2.1.

Table 2.1: *The Three Tiers of Data Value Creation. Adapted from Mazzei and Noble (2017).*

DATA AS A TOOL	Managers are able to solve traditional value chain problems more efficiently and effectively; existing capabilities are improved through real-time, customized decision making for individual consumers.
DATA AS AN INDUSTRY	Spin-offs and new ventures are created to specialize in acquisition, storage, and analysis of data, construction of infrastructure, and development of software devoted to handling big data.
DATA AS A STRATEGY	Visionary leaders develop companies dedicated to building data resources to allow them to develop radically innovative business models that wed traditional and modern strategic thought.

In a similar vein, Davenport and Redman (2020), heavily influenced by the resource-based view of the firm (Barney, 1991), identify 'proprietary data'[1] as data that is (1) unique to a particular firm and (2) can be used to gain sustainable competitive advantage. As many scholars do in recent works, Davenport and Redman (2020) argue that these data assets require specific strategic consideration and stress the need

[1]Although we use the definition offered by Davenport& Redman (2020), we will refrain from the term 'proprietary' as it implies that data can be claimed as property. As data is not covered by formal intellectual property laws, ownership is by no means obvious. Instead, we have opted to simply use the term 'data assets'. For more information on the ownership of data, see Appendix A.

for distinct 'data strategies' (see e.g., Constantiou & Kallinikos, 2015; Dallemule & Davenport, 2017; Davenport & Redman, 2020; Hagiu & Wright, 2020; Mazzei & Noble, 2017; Medeiros et al., 2020). Data strategy is a relatively new term that lacks a commonly accepted definition (Grossman, 2018; Medeiros et al., 2020). Adapting from Grossman (2018), Dallemule and Davenport (2017), and Davenport and Redman (2020), we define it as *the long-term decisions firms make to organize, govern, analyze, and deploy data assets.*

2.2 Offensive and Defensive Data Strategies

Building on the work of Dallemule and Davenport (2017), Medeiros et al. (2020) distinguish between defensively positioned and offensively positioned data strategies. The former emphasizes minimizing downside risk, especially that stemming from compliance issues, and reducing costs. The latter focuses on digital innovation and increasing revenues (Dallemule & Davenport, 2017). Medeiros et al. (2020) specifically studied how the existence of a formal data strategy impacted sustainable competitive advantage and found that (1) regardless of defensive or offensive positioning, the implementation of a data strategy positively impacts firms' competitive advantage, (2) the creation of either a defensive or offensive position means developing an organizational, analytical capacity which allows for the reconfiguration from a defensive to an offensive stance or vice versa, and (3) defensive positions are usually developed prior to offensive ones. These are visualized in Figure 2.2.

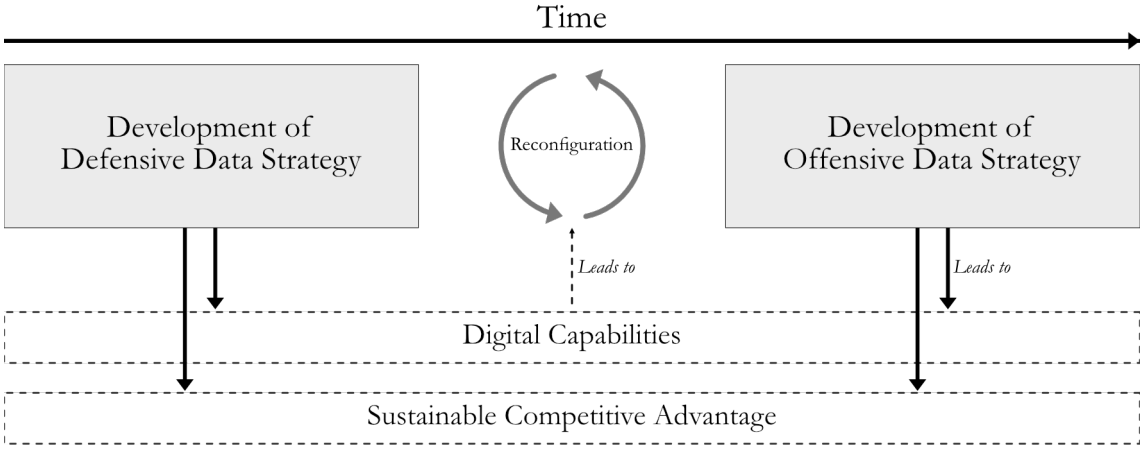


Figure 2.2: *Relationship between Developing Data Strategies, Digital Capabilities, and Competitive Advantage. Adapted from Medeiros et al. (2020).*

At its core, an offensive strategy focuses on leveraging data and its application to achieve market dominance and increase revenue (Dallemule & Davenport, 2017; Gupta & Cannon, 2020). Typically, this also entails a substantial focus on analytics

to gain deeper insight into customers or markets. Several examples of activities that characterize such a strategy are provided in the literature to clarify the practical implications of an offensive data strategy. However, these examples tend to be fragmented and vary in detail, ranging from broad objectives to specific tasks (see, e.g., Gupta & Cannon, 2020; Medeiros et al., 2020). The lack of a comprehensive overview of what practically constitutes an offensive data strategy can likely be attributed to the immaturity of this field, which has received limited academic attention and is often informed by practitioners' perspectives rather than academic research (Dallemule & Davenport, 2017; Gupta & Cannon, 2020; Wallis, 2021).

Despite these limitations, it can be observed that all the examples of activities pertinent to an offensive data strategy align with the overarching objectives of using data to either (1) increase revenue or (2) foster innovation and product development. Thus, these two objectives can serve as the principal criteria for an offensive data strategy. The challenges encountered in defining the objectives and activities of an offensive strategy are also attributable to a defensive strategy. Two main objectives can be identified for the defensive strategy, namely (1) cost reduction and (2) risk minimization, as visualized in Figure 2.3 (Dallemule & Davenport, 2017; Gupta & Cannon, 2020; Wallis, 2021).

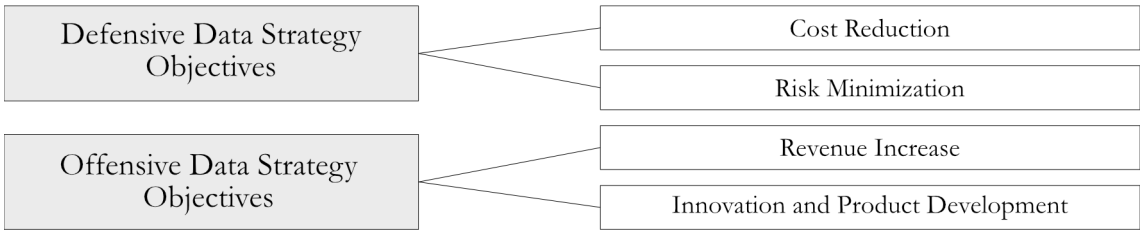


Figure 2.3: *Objectives of Offensive and Defensive Data Strategies.*

2.3 Data Governance

Data governance, sometimes referred to as information governance, is primarily covered by information system (IS) literature (Al-Ruithe et al., 2019; Alhassan et al., 2016; Benfeldt et al., 2020; Black et al., 2023; Davidson et al., 2023; Nielsen, 2017) but is also referenced as an essential component to data strategy in management literature (Constantiou & Kallinikos, 2015; Dallemule & Davenport, 2017; Medeiros et al., 2020). Within IS literature, there is no agreed-upon definition of the concept. Therefore, we have opted to use the definition which IS scholars seemingly most frequently reference:

> *The exercise of authority, control, and shared decision-making (planning, monitoring, and enforcement) over the management of data assets.*
>
> *- DAMA International (2017)*

Data governance is generally seen as having two primary objectives: (1) maximize the value of a firm's data assets and (2) manage data-related risks (Abraham et al., 2019; Brous et al., 2016; Davidson et al., 2023; Jagals et al., 2021; Lillie & Eybers, 2019; Vial, 2023). Data becoming key in driving value for innovative firms has led to the former being discussed in terms of digital innovation (Davidson et al., 2023; Jagals et al., 2021; Schüritz et al., 2017), while the increasing complexity of the judicial landscape has led to the latter being primarily concerned with regulatory compliance (Black et al., 2023; Davidson et al., 2023; Lillie & Eybers, 2019). Vial (2023) argues that while the dual objectives of data governance make sense in theory, they are paradoxical in that favoring one hinders the other. Problematically, the high risk exposure that data entails leads firms to generally emphasize the need for compliance, thus impeding digital innovation and the creation of new value offerings. Exactly how to cope with the opposing duality of data governance is an area that has received little attention beyond acknowledging the issue's existence (Vial, 2023).

In practice, data governance is considered by scholars to be conducted through the deployment of a number of *governance mechanisms* (Abraham et al., 2019; Davidson et al., 2023; Lis & Otto, 2020; Tallon et al., 2013; Vial, 2023). In their paper, Abraham et al. (2019) reviewed 145 articles and defined governance mechanisms as:

> *. . . formal structures connecting business, IT, and data management functions, formal processes and procedures for decision-making and monitoring, and practices supporting the active participation of and collaborations among stakeholders.*
>
> *- Abraham et al. (2019)*

It is also common practice in the literature to distinguish between three overarching categories of mechanisms: (1) *structural*, (2) *procedural*, and (3) *relational* (Abraham et al., 2019; Davidson et al., 2023; Lis & Otto, 2020; Tallon et al., 2013; Vial, 2023).

2.3.1 Structural Governance Mechanisms

Structural governance mechanisms regulate governance bodies, accountabilities, and reporting structures (Abraham et al., 2019; Borgman et al., 2016). Abraham et al. (2019) identify roles, responsibilities, and the location of decision-making authority as structural mechanisms.

2.3.1.1 Roles and Associated Responsibilities

According to Abraham et al. (2019), the key roles are executive sponsor, data governance leader, data owner, and data steward. The key governance bodies are the data governance council, data governance office, data producer, and data consumer. Additionally, there are data producers and data consumers, both of which are roles that may be held by a single individual or group (Abraham et al., 2019; Kooper et al., 2011). A summary of the roles and their associated responsibilities is found in Table 2.2.

Executive Sponsor – The executive sponsor is a high-level executive responsible for providing (1) funding, (2) strategic direction, and (3) prioritization toward data management (Abraham et al., 2019; Weber et al., 2009).

Data Governance Leader – Data governance leaders manage data governance on a day-to-day basis and are responsible for (1) providing guidance on the delivery, design, and maintenance of data assets, (2) ensuring compliance with data policies, and (3) coordinating the efforts of data stewards (Abraham et al., 2019).

Data Owner – Data owners have formal accountability for data assets within a specific business unit (Abraham et al., 2019). They are responsible for communicating general data requirements and risks within their business unit (Abraham et al., 2019).

Data Steward - Data steward is a role that has emerged in more modern literature and has been proposed to replace the role of data owners (Vial, 2023). Data stewardship is covered by many scholars (Abraham et al., 2019; Khatri & Brown, 2010; Plotkin, 2020; Tallon et al., 2013). It is not, however, a clearly defined role. Data stewards are sometimes discussed broadly (Khatri & Brown, 2010; Tallon et al., 2013). Other scholars distinguish between business data stewards (Abraham et al., 2019; Al-Ruithe et al., 2019; Plotkin, 2020; Weber et al., 2009), technical data

stewards (Abraham et al., 2019; Al-Ruithe et al., 2019; Plotkin, 2020; Weber et al., 2009), project data stewards (Dyché & Polsky, 2016; Plotkin, 2020), operational data stewards (Plotkin, 2020), and chief stewards (Al-Ruithe et al., 2019; Weber et al., 2009). However, the roles of operational, project, and chief steward relate to the organizational level of data steward responsibilities (Dyché & Polsky, 2016; Plotkin, 2020; Tallon et al., 2013). Therefore, what is of interest are the responsibilities carried out by business and technical data stewards.

Responsibilities connected to business data stewards identified in the literature are: (1) relating business requirements to technical specifications (Abraham et al., 2019), (2) knowing what data within their business unit means, is supposed to represent, and what business rules are associated with it (Plotkin, 2020), (3) decide how to use data to create competitive advantage (Tallon et al., 2013), (4) detailing data quality standards and business policies (Weber et al., 2009), and (5) ensuring compliance to new legal requirements (Tallon et al., 2013).

Technical data stewards are responsible for (1) translating business requirements to technical specifications (Abraham et al., 2019), (2) knowing how data is created, stored, manipulated, and moved throughout technical systems (Plotkin, 2020; Weber et al., 2009), (3) providing standardized definitions and formats for data elements (Weber et al., 2009).

Data Governance Council – a government body that overarches organizational hierarchies (Abraham et al., 2019; Otto, 2011b; Weber et al., 2009). Responsible for (1) determining the strategic direction of data governance, (2) aligning data governance with business goals, and (3) monitoring the performance of data governance (Abraham et al., 2019).

Data Governance Office – a governance body that carries out supportive functions for the data governance council and data stewards (Abraham et al., 2019). Responsibilities are (1) the establishment of communication channels, (2) preparing of meetings, (3) education of stakeholders, and (4) coordination of issue resolution (Abraham et al., 2019).

Data Producer – the data producer either creates the data or aggregates and maintains the data produced by others (Abraham et al., 2019; Kooper et al., 2011). Responsible for (1) creating data, (2) aggregating data, and (3) maintaining data.

Data Consumer – the data consumer is the user of data (Abraham et al., 2019; Kooper et al., 2011). Responsible for (1) specifying data requirements and (2) reporting data-related issues (Abraham et al., 2019).

Table 2.2: *Summary of Roles and Associated Responsibilities.*

Role	Responsibilities
Executive Sponsor	• Provide funding. • Provide strategic direction. • Provide prioritization toward data management.
Data Governance Leader	• Provide guidance on delivery, design, and maintenance of data assets. • Ensure compliance to data policies. • Coordinate efforts of data stewards.
Data Owner	• Communicate data requirements and risks within a specific business unit.
Business Data Steward	• Relate business requirements to technical specifications. • Understand the meaning, intended representation, and the rules associated with data assets in a specific business unit. • Decide how to use data assets to create competitive advantage. • Detail data quality standards and business policies. Ensure regulatory compliance.
Technical Data Steward	• Translate business requirements to technical specifications. • Understand how data is created, stored, manipulated, and flows. • Provide standardized definitions and formats for data elements.
Data Governance Council	• Determine strategic direction for data governance. • Align data governance with business goals. • Monitor performance of data governance.
Data Governance Office	• Establish communication channels. • Prepare meetings. • Educate stakeholders.
Data Producer	• Create data. • Aggregate data. • Maintain data.
Data Consumer	• Specify data requirements. • Report data-related issues.

2.3.1.2 Allocation of Decision-Making Authority

The allocation of decision-making authority refers to what organizational unit holds the right to action in relation to data governance (Abraham et al., 2019; Khatri & Brown, 2010). The literature distinguishes between the hierarchical and the functional positioning of decision-making authority (Abraham et al., 2019; Otto, 2011a). Hierarchical positioning refers to the level within an organization where the decision-making authority is situated. Functional positioning refers to which

business unit holds the authority. Further, whether decisions are centralized or decentralized is considered (Abraham et al., 2019; Tallon et al., 2013).

2.3.2 Procedural Governance Mechanisms

Abraham et al. (2019) identifies nine distinct procedural governance mechanism that are aimed at ensuring data assets are used effectively, recorded correctly, stored securely, and shared appropriately. These are (1) the data strategy, (2) policies, (3) standards, (4) processes, (5) procedures, (6) contractual agreements, (7) performance measurement, (8) compliance monitoring, and (9) issue management. Table 2.3 goes through the definition of each of these as described in IS literature. It should be noted that while IS literature identifies 'data strategy' as a governance mechanism, it is a research field in of itself covered in management literature. For that reason, we will return to data strategies in the next section which have dedicated entirely to the subject.

Table 2.3: *Procedural Governance Mechanisms.*

MECHANISM	DEFINITION
DATA STRATEGY	IS literature defines data strategy as a broad plan of action that is aligned with the overall strategic goals of the firm (Abraham et al., 2019). It typically includes a vision statement, a business proposal, guiding principles, both long- and short-term objectives, and a roadmap for implementation.
DATA POLICY	Data policies outline overarching principles and regulations for how data should be created, obtained, stored, secured, maintained for quality, and used in an appropriate manner (Abraham et al., 2019; Alhassan et al., 2019).
DATA STANDARD	Data standards aim to guarantee uniformity in the way data is represented and processed across the entire organization, thus facilitating interoperability within the firm (Abraham et al., 2019).
DATA PROCESS	Data processes are considered fundamental to the success of data governance (Abraham et al., 2019; Alhassan et al., 2019). They are formalized, standardized, and documented ways of implementing data governance (Abraham et al., 2019). Examples include (1) processes for developing rules for data processing and (2) processes for mapping data lifecycles.
DATA PROCEDURE	Data procedures are documented ways to accomplish a specific task, and they describe specific techniques and steps for doing so (Abraham et al., 2019; DAMA International, 2017).

Table 2.3: *Procedural Governance Mechanisms* (Continued)

MECHANISM	DEFINITION
CONTRACTUAL AGREEMENTS	Contractual agreements refer to any formal agreement aimed at facilitating data sharing internally between departments or externally with other organizations (Abraham et al., 2019).
PERFORMANCE MEASUREMENT	As the name suggests, performance measurement refers to assessing the effectiveness of data governance (Abraham et al., 2019; Otto, 2011b; Weber et al., 2009). This is done by monitoring the level of goal attainment in relation to business objectives.
COMPLIANCE MONITORING	Compliance monitoring refers to ensuring conformance to regulatory requirements and internal policies and standards (Abraham et al., 2019).
ISSUE MANAGEMENT	Issue management aims to identify, manage, and resolve data-related issues, including formalized processes for doing so (Abraham et al., 2019).

2.3.3 Relational Governance Mechanisms

The relational mechanisms identified by Abraham et al. (2019) are (1) communication, (2) training, and (3) coordination of decision-making. They are aimed at facilitating collaboration amongst stakeholders. Each mechanism is defined in Table 2.4.

Table 2.4: *Relational Governance Mechanisms.*

MECHANISM	DEFINITION
COMMUNICATION	Communication refers to the practice of continuously creating awareness for data governance in order to create a shared commitment amongst stakeholders (Abraham et al., 2019).
TRAINING	Training is aimed at ensuring that relevant stakeholders have the proper skills to implement data governance (Abraham et al., 2019). It takes many forms including computer-based training, one-on-one coaching, and job-specific training. By continuously training stakeholders, they are helped to act in accordance with policies, processes, and standards (Abraham et al., 2019; Alhassan et al., 2019).
COORDINATION OF DECISION-MAKING	Abraham et al. (2019) defines the coordination of decision-making as the practices aimed at aligning different functions. The authors distinguish between the hierarchical approach, characterized by pyramid-like structure with the decision-making authority on top, and the cooperative approach, characterized by a collaborative behavior aimed at clarifying differences and solving problems. The latter makes use of both formal coordination mechanisms, including committees, and task forces, and informal coordination mechanisms such as job rotation and cross-business performance measures (Abraham et al., 2019; Borgman et al., 2016; Tallon et al., 2013; Weber et al., 2009).

2.4 Regulatory Frameworks Related to Data

As has been accounted for in the previous sections, regulatory perspectives to data are emphasized both in relation to data strategy and data governance (Dallemule & Davenport, 2017; Davidson et al., 2023; Gupta & Cannon, 2020; Medeiros et al., 2020; Vial, 2023; Wallis, 2021). The emerging regulatory landscape concerning data is often what pivots the attention within an organization to data in the first place (Wallis, 2021). The legislators' mandate to sanction non-compliant entities further turns regulation into a specific type of coercive institutional pressure that force organizations to change their practices accordingly (Hu et al., 2007; Krell et al., 2016). Recent literature identify regulatory risks as one of the main considerations in relation to data (Black et al., 2023; Davidson et al., 2023; Ladley, 2020; Vial, 2023). This is also motivated by the fact that various regulatory frameworks, such as the GDPR, encompass other data-related risks like data breaches and data quality (Bindley, 2019; Stephens, 2021, see also e.g., Art. 5 and Art. 33 GDPR). Consequently, compliance with such regulations has the potential to simultaneously mitigate other risks (Diamantopoulou et al., 2020a, 2020b). To this background, an in-depth exploration of the regulatory frameworks that are of relevance is warranted.

There are several regulatory perspectives on data that require consideration. The more personal, private, and sensitive the data is, the more complex the regulatory framework becomes for processing such data (Gregg et al., 2006; Siddiqa et al., 2016; van der Sloot, 2020). For the purpose of this thesis, the GDPR, the AI Act, and the Data Act will provide the basis for creating an overview of the regulations relating to data. From each regulatory framework, it is possible to derive a set of categories of data and assign roles and responsibilities required for the management of each category. This is illustrated below in Table 2.5. The regulatory frameworks are not meant to provide an exhaustive account of all regulatory perspectives of data. Sector specific regulations along with a still evolving regulatory landscape makes it impossible to give such an account. Rather, the categories offered here serve as examples of how regulatory perspectives regarding data create obligations and accountabilities that necessitate implementation for compliance purposes.

Table 2.5: *Data Categories and Roles Derived from Respective Regulation.*

Regulation	Data Categories	Roles(s)
GDPR	• Personal Data • Sensitive Data • Pseudonymized Data • Anonymized	• Data Controller • Data Processor • Data Protection Officer
AI Act	• Taining Data • Validation Data • Testing Data • Input Data	• Provider • Importer • Distributor • User
Data Act	• Data Produced by the Use of Products and Related Services	• Product Manufacturer • Service Supplier • Data Holder • Data Recipient

2.4.1 GDPR

The General Data Protection Regulation (GDPR) is a comprehensive privacy regulation that governs the processing of personal data of individuals within the EU and the export of such data outside the EU. The GDPR applies to any organization that processes personal data of EU residents, regardless of whether the organization is based in the EU or not. It provides individuals with various rights regarding their personal data, including the right to access, rectify, erase, restrict processing, and portability of their data.

2.4.1.1 Data Categories under the GDPR

The GDPR designates several data categories that serve as the basis for the rules that apply to each category. These categories, and their definition, are outlined below in Table 2.6 and are based on the descriptions provided in the Recitals and Art. 4 of the GDPR.

The categorization of data into personal and non-personal data is a fundamental legal distinction that carries significant implications for data governance. It determines the applicability of the GDPR and privacy regulations worldwide, e.g., the California Consumer Privacy Act of 2018 (CCPA) and Australia's Privacy Amendment (Notifiable Data Breaches) Bill 2017. The differentiation is challenging, particularly in cases of de-identified data, which involves converting personal data into anonymous data through deliberate modifications (Finck & Pallas, 2020).

Table 2.6: *Data Categories under the GDPR.*

CATEGORY	DESCRIPTION
PERSONAL DATA	Any information relating to a natural person ('data subject') who can be identified, directly or indirectly, by reference to an identifier such as a name, an identification number, location data, an online identifier, or factors specific to the physical, physiological, genetic, mental, economic, cultural, or social identity of that natural person.
SENSITIVE DATA	Data revealing racial or ethnic origin, political opinions, religious or philosophical beliefs, trade union membership, genetic data, biometric data, and data concerning a person's health, sex life or sexual orientation.
PSEUDONYMIZED DATA	Personal data that no longer can be attributed to a specific data subject without the use of additional information, provided that such additional information is kept separately and is subject to technical and organizational measures to ensure that the personal data are not attributed to a natural person.
ANONYMIZED DATA	Personal data that has been rendered anonymous in such a way that the data subject no longer is identifiable and reidentification is impossible.

Information rendered anonymous so that the data subject is no longer identifiable does not fall within the scope of the GDPR (Recital 26 GDPR). However, it is necessary to distinguish between anonymous data and pseudonymized data. Pseudonymization is a privacy protection measure that is technically reversible (*Opinion 4/2007 on the concept of personal data,* 2007, p. 18). Pseudonymity allows for the reidentification of the data subject and therefore stays inside the scope of the GDPR (*Opinion 05/2014 on Anonymisation Techniques*, 2014, p. 10). Following Recital 26 of the GDPR, the critical factor for determining whether data is pseudonymous or anonymous is the ability to identify the data subject. To assess this, all possible identification methods should be considered, as well as the available technology at the time of the processing and future technological developments. Three criteria guide the assessment of whether anonymization has occurred, namely (1) if it is still possible to single out an individual, (2) if it is still possible to link records relating to an individual, and (3) if information concerning an individual can still be inferred (*Opinion 05/2014 on Anonymisation Techniques*, 2014, p. 3).

As indicated above, the legal definition of anonymous data is ambiguous. The definitions offered in the GDPR, by the Article 29 Working Party, and by national supervisory authorities vary considerably. The issue is further complicated given that, in reality, data exists on a spectrum between being clearly personal, clearly anonymous, and anything in between, which contrasts the binary perspective adopted by

the regulation (Finck & Pallas, 2020). Up until recently, the threshold for when personal data is considered anonymous has been primarily based on the *Breyer* ruling (C-582/14, Patrick Breyer v Bundesrepublik Deutschland, ECLI:EU:C:2016:779, hereinafter referred to as *Breyer*). The CJEU stated that a dynamic IP address obtained by an online media services provider qualifies as personal data because, combined with additional information from the internet service provider, the online media services provider can potentially identify the individual associated with the IP address. The Court acknowledged that internet service providers are generally prohibited from sharing this data with online service providers (*Breyer*, para 47). Despite this, the Court held that the data was considered pseudonymized rather than anonymized because:

> *(...) in the event of cyber attacks legal channels exist so that the online media services provider is able to contact the competent authority, so that the latter can take the steps necessary to obtain that information from the internet service provider (...)*
>
> - *Breyer*, para 47

This means that the distinction between pseudonymous and anonymous data depends on the presence of re-identification data or 'additional information.' For data to be anonymous, no additional information can be available to attribute the data to a specific individual. If such additional information exists, data cannot be considered anonymized. It will be deemed pseudonymized, with little or no regard given to the reasonable likelihood of an organization accessing such additional information (Castellanos et al., 2023). This has been the prevailing landscape for many years and sets a remarkably high standard for anonymizing data.

However, in April 2023, the General Court delivered a decision that indicates that the threshold between pseudonymous and anonymous data is becoming more nuanced (T-557/20, Single Resolution Board v European Data Protection Supervisor, ECLI:EU:T:2023:219, hereinafter referred to as *SRB*). It should be noted that the judgment can and is likely to be appealed to the CJEU (Castellanos et al., 2023). In this case, the Single Resolution Board (*SRB*), an EU organization, disclosed certain comments from shareholders and creditors to Deloitte within the context of a 'right to be heard'-process (*SRB*, para 13-15). Before this sharing, the comments were randomly assigned unique 33-digit alphanumeric identifiers. The SRB maintained a database that allowed them to connect these identifiers back to the original commenters. However, neither Deloitte nor the SRB staff handling the comments had access to this database.

SRB argued that even if the original processing entity maintained the potential for re-identification of the data and the information was not eliminated, the data could still be considered anonymous when shared with a third party. According to SRB's arguments, this presupposes that the data is shared in a form that does not allow re-identification or where re-identification is not reasonably likely. The Court agreed with the SRB and stated that whether information constitutes personal data shall be assessed from the perspective of each party involved (*SRB*, para 97-105). Consequently, what may be considered personal data for one company might not be the case for another. As put by Castellanos et al. (2023), the same data in different hands can qualify as both personal data and non-personal data, depending on the factual and legal circumstances in the specific scenario and the actual ability of each party to identify the data subject. Against this background, classifying data as pseudonymized or anonymized remains a subject of ongoing debate. The definitions are still evolving, with an inclination towards adopting a more nuanced and context-based approach, contrary to the previously prevailing case law.

Further, the GDPR distinguishes certain forms of personal data as sensitive. Information sensitivity is often determined by the magnitude and severity of risks associated with processing it (Fazlioglu, 2019). The GDPR imposes higher obligations upon the processing of these types of data, which is typically motivated by the likelihood and severity of the harms that can arise from misuse of the data (Fazlioglu, 2019; Skinner-Thompson, 2015, see also e.g., Art. 24-25 GDPR). The technological developments relating to data require legislators to constantly re-evaluate and expand the categories of data that are considered sensitive (Einav & Levin, 2014; Fazlioglu, 2019).

2.4.1.2 Roles and Responsibilities under the GDPR

Apart from the data categories accounted for above, the GDPR designates a set of roles that subsequently defines what responsibility the person or organization who fills that role is obliged to take. The primary roles, and their associated responsibilities, are outlined below in Table 2.7 and are derived from Art. 4 and Art. 37-39 in the GDPR, with support from other relevant provisions.

Notably, it is typical for an organization not exclusively to operate as the controller or processor but instead take on both roles in parallel (Lambrinoudakis, 2018; Sharma, 2019). For example, a company conducting market research on behalf of another company is a data processor. Still, when managing the personal data of their em-

Table 2.7: *Roles and Responsibilities based on the GDPR.*

ROLE	DEFINITION	RESPONSIBILITY
Data Controller	The natural or legal person which, alone or together with others, **determines the purposes and means** of the processing of personal data.	Main responsibility for ensuring that personal data is processed in accordance with the GDPR. Includes e.g., being able to demonstrate compliance with processing, data minimization and protection principles, establish legal basis for processing, and facilitate the exercise of data subject rights.
Data Processor	A natural or legal person which processes personal data on **behalf of the controller**. E.g., payroll firms, cloud service providers, and data analytics providers.	To only process personal data according to the instructions of the data controller. Is directly accountable for some parts of data protection, e.g., implementation of technical and organizational measures to ensure the security of personal data.
Data Protection Officer	A designated person with **professional qualities** and **expert knowledge** of data protection law and practices, who is involved in all issues which relate to the protection of personal data.	Responsible for ensuring that the strategy and implementation of data protection requirements are in accordance with the GDPR. Advises individuals or teams who carry out processing of their obligations pursuant to the GDPR.

ployees, they become data controllers. Both data controllers and processors must appoint a Data Protection Officer (DPO) in certain circumstances. These include when the processing is carried out by a public authority, when the processing activities include regular and systematic processing of a large number of data subjects, and when large-scale processing of sensitive data is carried out (Art. 37 GDPR). DPOs are responsible for ensuring that an organization's data protection strategy and implementation comply with the GDPR (Art. 39 GDPR, see also Sharma, 2019). The tasks of the DPO also include assigning responsibilities, raising awareness around data protection, and training staff involved in processing operations.

2.4.2 AI Act

The European Commission presented the proposal for the AI Act in April 2021. The Act aims to establish consistent regulation for creating, marketing, and utilizing AI systems, considering their characteristics and associated risks. This involves implementing prohibitions and a conformity assessment system similar to the EU's product safety legislation (Veale & Borgesius, 2021). Following the proposal of the AI Act, it is clear that once the regulation is adopted and enters into force, AI-related data will be subject to specific rules regardless of whether it falls within the scope

of data privacy regulation. The AI Act defines an AI system broadly as software that is developed using certain techniques and that can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with (Art. 3(1) AI-Act). AI systems are further split into four different risk categories based on the intended use of the system (Edwards, 2022). In practice, the AI Act is most concerned with 'high-risk AI,' which accounts for the vast majority of the prohibitions included in the regulation.

The inextricable link between AI and data generates a connection between the GDPR and the AI Act (Townsend, 2021). While the AI Act does not apply to controllers or processors, it will, in practice, apply in situations where personal data is used in AI systems. In cases where the AI Act is also applicable, it will impose obligations beyond GDPR. Therefore, some controllers or processors under GDPR will have additional obligations stemming from the AI Act and vice versa (Greenleaf, 2021).

2.4.2.1 Data Categories under the AI Act

The AI Act designates several data categories that serve as the basis for the rules that apply to each category. These categories, and their definition, are outlined below in Table 2.8 and are based on the definitions provided in Art. 3 of the AI Act.

Table 2.8: *Data Categories under the AI Act*

CATEGORY	DESCRIPTION
TRINING DATA	Data used for training an AI system through fitting its learnable parameters, including the weights of a neural network.
VALIDATION DATA	Data used for providing an evaluation of the trained AI system and for tuning its non-learnable parameters and its learning process, among other things, in order to prevent overfitting, whereas the validation dataset can be a separate dataset or part of the training dataset, either as a fixed or variable split.
TESTING DATA	Data used for providing an independent evaluation of the trained and validated AI system in order to confirm the expected performance of that system before it is placed on the market or put into service.
INPUT DATA	Data provided to or directly acquired by an AI system on the basis of which the system produces an output.

Recital 44 of the AI Act specifically addresses the need for high-quality data sets within the above-defined categories. Data quality is crucial for the proper functioning of various AI systems, especially those that use model training techniques. This ensures that high-risk AI systems operate safely and as intended without being a source of discrimination. Appropriate data governance and management practices must be implemented to achieve high data quality. The quality criteria the data sets must meet relate to relevance, representativeness, accuracy, completeness, and application-area-specific properties (Art. 10 AI Act). The criteria should be applied in view of the intended purpose of the system, meaning that the criteria should be assessed with consideration of the features, characteristics, or elements that are particular to the specific geographical, behavioral, or functional setting or context within which the AI system is intended to be used (Recital 44 AI Act, see also Veale & Borgesius, 2021).

2.4.2.2 Roles and Responsibilities under the AI Act

Just as the GDPR, the AI Act designates a set of roles that subsequently defines what responsibility the person or organization who fills that role is obliged to take. The primary roles, and their associated responsibilities, are outlined below in Table 2.9 and are derived from relevant provisions in Chapter 3 of the AI Act.

Table 2.9: *Roles and Responsibilities based on the AI Act.*

ROLE	DEFINITION	RESPONSIBILITY
PROVIDER	The natural or legal person that **develops** an AI system or that has an AI system developed with a view to **placing it on the market** or **putting it into service** under its own name or trademark, whether for payment or free of charge.	Holds the overall responsibility for ensuring that the AI system is transparent, accountable, and trustworthy. Specific tasks vary depending on the risk level of the AI system but include risk assessment, technical documentation, human oversight, appropriate use of data, and preventing harm and discrimination.
IMPORTER	A natural or legal person in the EU that places an AI system, which bears the name or trademark of a natural or legal person established outside the EU, on the market or puts it into service.	Obliged to ensure that high-risk AI systems bear the required CE marking and are accompanied by the required documentation and instructions of use, and that the provider has complied with the obligations set out in the AI Act.

Table 2.9: *Roles and Responsibilities based on the AI Act.* (Continued)

ROLE	DEFINITION	RESPONSIBILITY
DISTRIBUTOR	Any natural or legal person in the supply chain, other than the provider or the importer, that makes an AI system available on the EU market without affecting its properties.	Obliged to verify that high-risk AI systems bear the required CE marking and are accompanied by the required documentation and instruction of use, and that the provider and the importer have complied with the obligations set out in the AI Act.
USER	A natural or legal person using an AI system **under its authority**, except where the AI system is used in the course of a personal non-professional activity.	Obliged to use the system in accordance with the instructions of use, ensure that input data is relevant in view of the intended purpose, and monitor the operation of the system.

From this, it's clear that the provider of an AI system takes on the heaviest responsibility under the AI Act. However, an importer, distributor, or user can be considered as the provider if they place a high-risk AI system on the market under their name or trademark, modify the intended purpose of a high-risk AI system already placed on the market, or make a substantial modification to the high-risk AI system (Art. 28 AI Act).

2.4.3 Data Act

The proposed Data Act aims to regulate the use and access of data generated in the EU across all economic sectors. As with the AI Act, there will be situations where the Data Act overlaps with the GDPR. In such cases, GDPR still governs personal data. Suppose personal data is processed in connection with the rights and obligations in the Data Act. In that case, the data holder is considered a controller under the GDPR, and the actor is thus required to navigate the overlap of the regulations (Recital 24 Data Act). This could entail a complex practical situation, as the new data access and sharing rights could be affected by the high standards, legal uncertainty, and practical difficulties related to the GDPR's concept of consent (Leistner & Antoine, 2022).

Further, the relation to intellectual property rights and trade secrets protection should be touched upon. The protection of databases[2] provided in the Database Directive does not apply to databases containing data generated using a product or a related service (Art. 35 Data Act). The relation to trade secrets is addressed in Art. 4(3) and Art. 5(8), where it is stated that trade secrets shall only be disclosed to third parties when it is strictly necessary to fulfill the purpose agreed between the user and the third party, and provided that necessary measures are taken to preserve the confidentiality.

2.4.3.1 Data Categories under the Data Act

In principle, the Data Act only specifies one type of data the regulation applies to. 'Data,' on a general level, is defined as "*any digital representation of acts, facts or information and any compilation of such acts, facts, or information, including in the form of sound, visual or audio-visual recording*" (Art. 2(1) Data Act). Throughout the Data Act, it is clear that the provisions apply to data, as defined in Art. 2(1), that is generated by the use of products or related services.

In this context, 'product' is defined as "*a tangible, movable item, including where incorporated in an immovable item, that obtains, generates or collects, data concerning its use or environment, and that is able to communicate data via a publicly available electronic communications service and whose primary function is not the storing and processing of data*" (Art. 2(2) Data Act). Further, 'related service' is defined as "*a digital service, including software, which is incorporated in or interconnected with a product in such a way that its absence would prevent the product from performing one of its functions*" (Art. 2(3) Data Act).

2.4.3.2 Roles and Responsibilities under the Data Act

The Data Act designates a set of roles that subsequently defines what responsibility the person or organization who fills that role is obliged to take. The primary roles, and their associated responsibilities, are outlined below in Table 2.10 and are derived from relevant provisions in Chapter 2 of the AI Act.

[2] Database protection is a property right that exists to recognize the substantial investment that is made in obtaining, verifying, or presenting the contents of a database, even when this does not involve the creative aspect that is reflected by copyright.

In addition, Chapter 6 of the Data Act introduces minimum contractual, commercial, and technical requirements. These requirements apply to cloud, edge, and other data processing service providers to enable switching between such services. The act also includes several provisions relating to the differentiation between different types of entities (Dyck et al., 2022). This includes limiting obligations and providing protections for SMEs while excluding large 'gatekeeper entities' (as designated under the Digital Markets Act) from certain rights under the legislation.

Table 2.10: *Roles and Responsibilities based on the Data Act.*

ROLE	DEFINITION	RESPONSIBILITY
USER	A natural or legal person that owns, rents, or leases a product or receives a service.	N/A – responsibilities arise if the user qualifies as a data recipient, see below.
DATA HOLDER	Any legal or natural person who has the **right** or **obligation** to make certain data available in accordance with the Data Act or other EU legislation, *and*, if the data is non-personal, any legal or natural person who through **control of the technical design** of the product and related services has the **ability** to make certain data available.	Includes providing generated data **to a user** without undue delay or charge, providing data **to third parties** upon **request by a user**, imposing **fair**, **reasonable**, and **non-discriminatory** terms when obliged to make data available to a data recipient, and fulfilling **transparency requirements**, i.e., informing the user of what and how much data is generated using the product and how they may access that data.
DATA RECIPIENT	The legal or natural person, acting for purposes which are related to that person's trade, business, craft, or profession, to whom the data holder makes data available.	If the recipient is a **third party**, restrictions such as purpose limitations, onward sharing limits, data deletion, non-compete/ exclusivity requirements, and data protection compliance apply. If the recipient is a **user**, restrictions regarding disclosure of trade secrets and personal data and the prohibition on using the data to create competing products apply.
PRODUCT MANUFACTURER/SERVICE PROVIDER	The party who **manufactures** products, as defined above, that are placed on the market in the EU, *and* the party who **supplies related services**, as defined above, that are placed on the market in the EU. (For definitions, see *Section 2.4.3.1*)	Ensure that their connected products and services are **designed** with access to data (easy, secure, and where relevant, direct) addressed by default.

3

Methodology

In this chapter, the methodology employed throughout this thesis is outlined. This involves describing the primary considerations related to the selected research strategy and design and discussing the research method and the key aspects of the research's quality.

3.1 Research Strategy

The primary objective of this study was to explore whether data strategies can be expressed in terms of firm maturity and how governance mechanisms are used to realize strategies at each maturity level. The primary research strategy was a qualitative study, which is commonly used in business research, and aligns well with the primary data collection method used in this study, namely semi-structured interviews. Additionally, the difficulty in finding appropriate variables to quantify the results motivated conducting a qualitative rather than a quantitative study.

3.1.1 The Role of Theory

To achieve the thesis' purpose, both deductive and inductive approaches were employed. As outlined below, the application of each approach differs between the different Sub Research Questions.

To answer Sub Research Question 1, existing theory from IS literature on data governance mechanisms was used to create an initial framework, which was subsequently deductively tested using semi-structured interviews. Based on the findings, new theory was developed, which was then verified through new interviews at the studied firms. Theory played the same role concerning Sub Research Question 2, where theory on the strategic management of data assets was used to create an initial framework. Figure 3.1 depicts the research approach deployed in both cases.

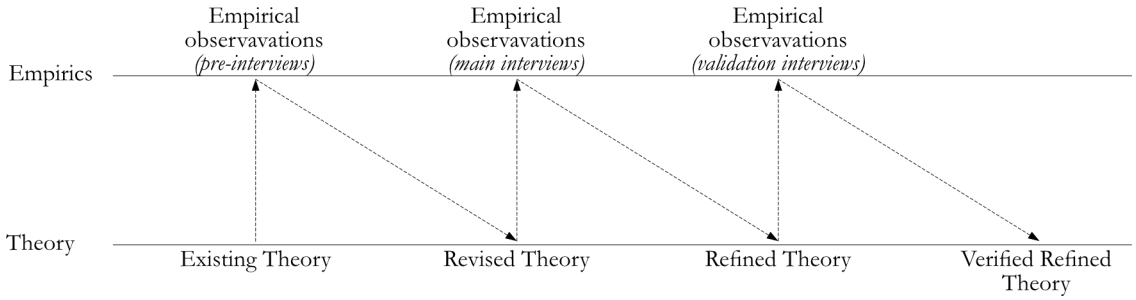


Figure 3.1: *The Role of Theory*

It could be argued that the stages of the research process, which include the revision of theory, entail that the research process should be considered inductive. As Bryman and Bell (2019) argues, deduction often entails an element of induction and vice versa. Fundamentally, deduction follows a process where theory leads to observations or findings, while induction follows the opposite process, where observations or conclusions lead to new theory development (Bryman & Bell, 2019). In this case, the empirical observations are used to revise existing theory, which can be done with a deductive approach. In determining whether the revision amounts to induction, it has to be considered that the revised theory remains closely linked to the original theory and merely incorporates minor adjustments based on the empirical findings. As a result, it would be an overstatement to claim that the new theory was developed solely based on empirical observations, as the received theory still constitutes the majority of the theoretical framework. Therefore, the research process, including the revision stages, is primarily considered deductive, while acknowledging that it includes some elements of induction.

Regarding Sub Research Question 3, the goal was to connect the identified governance mechanisms to each identified level of data strategy maturity. As evident from the role of theory in Sub Research Questions 1 and 2, theory on both the strategic management and governing of data assets was found in literature. However, no theory was identified during the literature review which expressly address the relationship between the two. Therefore, an inductive approach was used in relation to Sub Research Question 3, meaning that theory was created based on empirical investigations (Bryman & Bell, 2019). However, this is not to say that there is no element of deduction. We rely upon the findings in Sub-Research Questions 1 and 2 to answer the question. The answer, as such, is consequently greatly influenced by the theories used in previous inquiries.

3.1.2 Ontological and Epistemological Considerations

All phenomena studied in this thesis, namely *data assets*, *data strategy*, and *data governance*, are considered part of the socially constructed world. That is, they are 'made real' by the activities of humans and the meanings which observers attach to them and therefore are considered ontologically subjective (Bryman & Bell, 2019; Searle, 2007). Ontologically subjective phenomena are typically associated with interpretivist research, which is primarily concerned with understanding human behavior rather than explaining it (Bryman & Bell, 2019). Further, an interpretivist view of epistemology entails that knowledge concerning the studied phenomena can only be gained by understanding the social actors that give rise to their existence. However, the purpose of this thesis is not to question the meaning of the studied concepts. The meaning of data assets, data strategy, and data governance is largely taken for granted, as the concepts are considered to have a reified definition. While acknowledging that these concepts are indeed social constructions and thereby ontologically subjective, we consider them, in some measure, to exist objectively and externally, irrespective of the meanings that social actors assign to them. This represents a positivistic, or epistemologically objective, view of the studied subjects. Figure 3.2 illustrates the ontological and epistemological positioning of the Main Research Question.

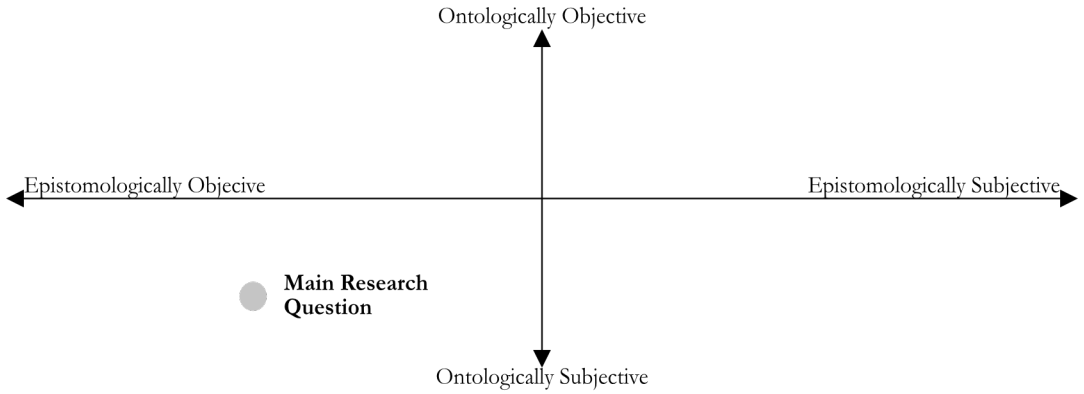


Figure 3.2: *Ontological and Epistemological Positioning*

The ambiguity expressed in the wording of the previous paragraph is warranted, as the meaning of these concepts can be affected by actions taken by social actors (Bryman & Bell, 2019). Viewing the studied subjects as epistemologically objective is not the only possible stance. Indeed, perhaps even more intuitively, one could argue that these concepts warrant an interpretivist view and thus are considered epistemologically subjective. The stance taken in this thesis is highly dependent

on the research questions and aims pursued, which motivates the aforementioned positioning. Therefore, the arguments put forward in this section seek to establish a positioning of the studied concepts that align with the objectives of this thesis rather than to provide a universal epistemological positioning. Therefore, the insights presented here should be interpreted as contextualized within the scope of this particular study.

3.2 Comparative Multiple Case Study

The research design employed in this thesis is a comparative multiple case study. Case study research concerns the complexity and particular nature of the case (Stake, 1995). The primary cases studied in this thesis are two distinct projects within an industrial company undergoing digitalization. The projects were the primary basis for selecting interviewees and gathering data. The projects represented two different perspectives: one being a digital-born product and the other being a digitalized mechanical product.

Moreover, a less comprehensive examination was conducted at one digital-born company. This study was less detailed than the main cases under investigation, why it was not possible to study specific projects at this company. Consequently, the research design, to some extent, draws on samples that combine different levels of analysis, i.e., project-level and company-level. To avoid potential misinterpretation, careful consideration should be given to whether something from one level is suitable to represent something at another level (Rousseau, 1985). The projects that served as the primary data source in this study were organized in a decentralized manner, granting each project the autonomy to establish individual data strategies and data governance practices. The projects were extensive, involving sizable teams, and were organized within distinct business units that functioned as independent legal entities. In contrast, the validation company employed a centralized approach, with mandates relating to data strategy and data governance situated at a higher organizational level. Given this organizational context, translating results between these levels is deemed appropriate.

Choosing cases that represent different perspectives, at project or company level, allowed for an idiographic approach, i.e., it was possible to highlight the unique features of each case (Bryman & Bell, 2019). The distinct characteristics of the cases also intend to contribute to their representativeness, as they exemplify three

diverse settings. Subsequently, the contrasting nature of the cases made it suitable to involve a comparative approach in the research design, as such an approach allows the distinguishing characteristics of the cases to act as a facilitator for theoretical reflections about contrasting discoveries (Bryman & Bell, 2019).

3.3 Research Method

This section outlines the research method used in this study, starting with a definition of the necessary data required to answer the research questions and explaining how the data is collected.

3.3.1 Required Data to Answer the Research Questions

The Main Research Question attempts to outline how data assets are governed at different levels of data-related strategic maturity. To answer this question, it has been broken down into three Sub Research Questions.

The first sub-research question focuses on how data assets are governed. To understand this, it is necessary to first understand what data assets exist and then gather data on what governance mechanisms are being deployed at each firm. To do so, a theoretical understanding of data governance is required, as outlined in Section 2.3.

The second sub-research question focuses on what levels of strategic maturity exist in relation to the management of data assets. To understand this, it must be possible to infer from the collected data what strategy is deployed at each studied firm. Thus, it is necessary to collect data on (1) what data assets exist, (2) why the data is being collected, and (3) how data is being strategically used and managed.

The third and final sub-research question aims at understanding what specific governance mechanisms enable the realization of different data strategies. This does not require additional data from Sub-Research Questions 1 and 2 regarding content. It does, however, require enough data to be collected so that the aggregation allows for identifying generalizable insights in the form of what governance mechanisms realize specific data strategies.

3.3.2 Data Collection

This thesis uses two primary methods to collect data. These methods are (1) interviews and (2) internal document analysis. Although it is not part of the formal data collection process, it is notable that an extensive literature review was conducted prior to the interviews and document analysis to define the theoretical framework underpinning each research question.

3.3.2.1 Interviews

The two main types of interviews used in qualitative research are unstructured and semi-structured interviews (Bryman & Bell, 2019). Unstructured interviews are characterized by using a selection of topics to cover a particular range of subjects or issues. The interviewer may pose only one question and allow the interviewee to answer freely, with the interviewer then responding accordingly to points worth exploring. A semi-structured interview involves a set of pre-determined questions, while also allowing both the interviewer and the interviewee to deviate from the script and make adjustments during the course of the interview.

The conducted interviews were divided into three stages: pre-interviews, main interviews, and validation interviews. The pre-interviews and main interviews were conducted as semi-structured interviews with a starting point in the theoretical framework developed from the literature review. As the research design is a comparative multiple case study, this amount of structure was deemed necessary to ensure cross-case comparability (Bryman & Bell, 2019). The pre-interview phase was characterized by open-ended questions, enabling interviewees to provide more expansive answers. Nonetheless, the pre-interviews were designed to cover relatively specific topics outlined in an interview guide, aligning with semi-structured interviews' characteristics (Bryman & Bell, 2019). Following the pre-interviews, the interviewee's abilities to provide valuable insights in relation to the research questions were assessed, and a selection of interviewees was chosen for the main interviews.

During the main interview phase, a more specific interview guide was used, as we now had gained a comprehensive understanding of the key concepts required to address the research questions within the designated research strategy. Before the interview, a copy of the interview guide was given to each interviewee. Questions were tailored to the interviewee's role; for instance, interviews with product managers or engineers focused more on technology and digital innovation, while interviews with

legal counsels or data protection officers were oriented toward risk management and compliance. However, questions relating to each field were presented to all interviewees to ensure a comprehensive understanding of all concepts from various relevant perspectives rather than solely relying on individuals with specific expertise in a particular area.

The validation interviews were conducted as unstructured interviews in a workshop format. Individuals with significant expertise and interest in relevant fields were selected as validation interviewees. The number of interviewees participating in the validation interviews was smaller than in the main interviews. To compensate for this, the validation interviews consisted of longer, interactive sessions where the interviewees were allowed to deep dive into the questions and elaborate on their answers and reflections. The validation interviews were conducted as an iterative process in the later stages of the study. During this phase, the selected validation interviewees were continuously updated about the impressions and findings from the study. Their comments were then considered, and adjustments were made accordingly.

Face-to-face interviewing is generally preferred over remote interviews (Bryman & Bell, 2019). Attempts were made to conduct as many face-to-face interviews as possible. However, due to geographical constraints, most interviews were conducted through Teams. The interviewees were selected based on their availability, willingness to participate, and ability to provide relevant insights and diverse perspectives on the topic. The main categories of interviewees are shown in Table 3.1. The full list of interviewees is disclosed in Appendix B.

Table 3.1: *Sample of Interviewees.*

ROLE	DESCRIPTION
Senior or Director Level Product Managers/Engineers	Has experience of working with data in products and projects in various ways, with great insight into how data is managed and governed within their teams.
Innovation Directors and Managers	Has experience of leading data driven innovation projects, with great insight into the requirements and perspectives that apply data in an innovation context.
Data Protection Managers and Senior Level Legal Councils	Responsible for ensuring that an organization complies with data protection regulations and thus has experience of education relating to data, advising the organization on data related issues, and monitoring data related activities.
Data Governance Directors	Responsible for establishing and implementing policies, procedures, and processes for the management of an organization's data assets. Has insight into developing data governance programs.
Software and Product Developers	Has experience of working with data science, software development and practical data management related to the use of data in products and projects.

3.3.2.2 Internal Document Analysis

In addition to the interviews, documentary data sources such as internal policies, guidelines, and corporate websites and reports have been collected and analyzed. These documents provided supplementary data to understand the investigated companies and projects further, thereby enriching the research. Additionally, documentary data is a valuable supplement to interview data, given its non-reactive nature. This compensates for the potential reactive effect in interviews, i.e., that the interviewees' responses may be influenced by their awareness of being studied (Bryman & Bell, 2019). Organizational documents should be evaluated based on four criteria: (1) *authenticity*, (2) *credibility*, (3) *representativeness*, and (4) *meaning* (Scott, 1990). Documents derived from private sources will likely be authentic and meaningful (Bryman & Bell, 2019). Instead, issues of credibility and representativeness are likely to be more complex. In this study, the use of documentary data as a supplementary source was accompanied by the inclusion of interview data from multiple independent sources. Further, a critical approach is applied to the documents, recognizing that organizational documents may be designed to reflect a favorable impression of the organization and therefore are likely to contain limited information about the company's problems. Combined, these actions are deemed to mitigate the risks relating to credibility and representativeness sufficiently.

3.4 Research Quality

In qualitative research, four criteria are primarily used to evaluate the research quality, namely (1) *credibility*, (2) *transferability*, (3) *dependability*, and (4) *confirmability* (Bryman & Bell, 2019; Guba & Lincoln, 1994; Lincoln & Guba, 1985). This section describes and discusses each of these criteria in relation to the research study.

3.4.1 Credibility

According to Bryman and Bell (2019), credibility refers to the possibility of addressing whether the research results are acceptable to others and perceived as consistent with reality. To achieve credibility, the research must be conducted with good practice and seek confirmation that the results are in accordance with the social world. In this study, respondent validation was used to ensure credible research results. By conducting validation interviews, the impressions and findings from the interviews were presented to selected interviewees with specific knowledge and interest in the study. Their comments were considered, and amendments or adjustments were made accordingly.

Further, triangulation was used to cross-check findings and ensure credibility. This entails using more than one method or source of data in the study of social phenomena (Bryman & Bell, 2019). The utilization of triangulation manifested in collecting data from multiple projects, multiple companies, and different sources (i.e., interviews and internal documentation), and subsequently cross-checking the findings from these sources with each other. For total transparency, it should be noted that time constraints limited this project's scope. As a result, it was not possible to include data from a wider range of sources or in different forms, which could have provided a more comprehensive understanding of the topic at hand. Despite these limitations, the approach taken in this section was designed to maximize the study's credibility within the research project's confines. Nevertheless, future studies could benefit from incorporating more diverse data sources to further enrich the analysis.

3.4.2 Transferability

Transferability refers to the applicability of findings and conclusions derived from one context to another (Leininger, 1994). Transferability is typically mitigated by producing what is referred to as thick description, i.e., rich accounts of the details of the studied culture to provide others with information for evaluating the possible

transferability of findings to other environments (Geertz, 1973; Lincoln & Guba, 1985). In this study, the categorization of projects and companies into 'digital born product' or 'digitalized mechanical product' and 'digital born company' provides insight into the specific context of the findings. This is further accompanied by descriptions of the hierarchy and organization of the companies, i.e. if a centralized or decentralized structure is employed.

3.4.3 Dependability

Dependability refers to the ability to repeat the generated research results (Bryman & Bell, 2019). Ensuring dependability involves adopting an 'auditing' approach to ensure that documentation is kept and can be accessed from all phases of the research process. This includes, e.g., documentation relating to problem formulation, selection of research participants, interview transcripts, and data analysis decisions (Bryman & Bell, 2019; Lincoln & Guba, 1985). In this research study, a considerable amount of data was collected through interviews at several companies, supplemented by internal documentation. Due to secrecy, much of this material is not susceptible to public disclosure. This may present challenges for other researchers seeking to replicate our findings to the fullest extent. The methodological approach described in this chapter can serve as a valuable guide for future researchers aiming to conduct similar studies. By using the detailed descriptions of the investigated companies as a starting point, researchers can identify comparable subjects and replicate the research process. Although there may be variations in the specific data collected, we believe that this approach can lead to similar results and contribute to a more comprehensive understanding of the topic.

3.4.4 Confirmability

Confirmability refers to the ability to corroborate data and challenge or affirm interpretation or theory (Drisko, 1997). The purpose is to show that the research has been conducted in good faith, i.e., that the researcher has not allowed their personal beliefs or theoretical biases to influence the research process and its resulting conclusions (Bryman & Bell, 2019). Repeated observations of the same phenomena enhance confirmability (Drisko, 1997). By studying a variety of projects and companies, repeated observations of the same phenomena were made possible. Further, Drisko (1997) proposes that feedback sessions with participants can confirm the accuracy of the interpretation of the data, thereby validating the researcher's interpretation and improving the confirmability of the study. The iterative validation

interviews served as a means to ensure the accuracy of our interpretation. Finally, consistency between what is reported from direct observation and other available sources further enhance the confirmability of the study. To this background, the study was anchored in established theoretical frameworks. This approach ensured that the findings were grounded in existing literature and that the interpretation of the data was consistent with established theories and concepts. By doing so, we believe that the study has achieved a high level of confirmability given the restrictions of the research project.

4

Results & Analysis

In this chapter, the empirical results from the data collection are presented. These results are derived from the interviews, relevant literature, and internal documentation. First, Section 4.1 presents the results of the investigation of data governance mechanisms. This account is purely descriptive, and the results are not further analyzed in their context. Section 4.2 proceeds to address the findings relating to strategic maturity. In this section, the distinction between pure results and analysis is blurred, which makes it challenging to describe the obtained results without analyzing them in parallel. Consequently, no designated section in this thesis will add an analytic layer to the investigation. Instead, Section 4.2 continuously embeds analysis in accounting for the results relating to Sub-Research Questions 2 and 3. Given the connection between the questions, this also includes a detailed analysis of the results of Sub-Research Question 1.

4.1 Implemented Governance Mechanisms

With regard to Sub-Research Question 1, the goal was to identify which governance mechanisms were deployed to ensure effective governance of data assets. The results are, in accordance with relevant literature, categorized into structural, procedural, and relational mechanisms. *P1* refers to the project involving a digital-born product. *P2* refers to the project involving a digitalized mechanical product. *C1* refers to the digital-born company.

4.1.1 Structural Governance Mechanisms

The identified structural mechanisms are presented in Table 4.2. The structural mechanisms, more than the procedural and relational mechanisms, tend to vary in their level of formalization. Employees were observed to take on data-related responsibilities outside the scope of their employment. In line with existing literature, the responsibilities outlined are represented across the companies. However,

the roles that take on the responsibilities differ from those commonly seen in the literature. One interesting finding is that the role of DPO informally takes on responsibilities outside the scope laid out in GDPR. It is also generally observable that informal responsibilities appear before formal ones in cases where executive sponsorship exists.

4.1.2 Procedural Governance Mechanisms

The procedural mechanisms identified through interviews are shown in Table 4.1. The mechanisms that appear first are those concerned with compliance issues: privacy policy, contractual agreements, and compliance monitoring.

Table 4.1: *Identified Procedural Governance Mechanisms.*

Mechanism	P1	P2	C1
Data strategy			
Defensive strategy			(X)
Offensive strategy	(X)		
Data policy			
Privacy Policy	X	X	X
Cybersecurity Policy			X
Retention Policy			X
Encryption Policy			X
Data standard			X
Data process			X
Data procedure			
Contractual agreements			
Internal Parties			X
External Parties	X	X	X
Performance measurement			X
Compliance monitoring	X	X	X
Issue management			X

X = formal mechanism (X) = informal mechanism

Generally, the findings related to procedural mechanisms support the existing literature on the topic. However, one finding of interest not mentioned in the literature is the apparent connection between the number of procedural mechanisms and the level of formalization of roles and responsibilities. Seemingly, companies with clearly defined roles concerning data governance tend to implement more procedural mechanisms.

Table 4.2: *Responsibilities Identified in the Studied Cases.*

Responsibility	**P1**	**Actual Role**	**P2**	**Actual Role**	**C1**	**Actual Role**
Provide funding.					X	General Counsel
Provide Strategic direction	X	Management			X	General Counsel
Provide prioritization toward data management	X				X	General Counsel
Provide guidance on delivery, design, and maintenance of data assets.					X	Data Governance Leader
Ensure compliance to data policies.					X	Data Governance Leader
Coordinate efforts of data stewards.					X	Data Governance Leader
Communicate data requirements and risks within a specific business unit.	X	Product Manager	X	Product Owner		
Relate business requirements to technical specifications.	(X)	R&D Manager / DPO			X	Team Manager
Understand the meaning, intended representation, and the rules associated with data assets in a specific business unit.			(X)	Various		
Decide how to use data assets to create competitive advantage.					X	R&D Team
Detail data quality standards and business policies.	(X)	R&D Manager / DPO				
Ensure regulatory compliance.	X	DPO	X	DPO		
Translate business requirements to technical specifications.	(X)	R&D Manager			X	
Understand how data is created, stored, manipulated, and flows.	(X)	R&D Manager / DPO	(X)	Various	X	Data Governance Team
Provide standardized definitions and formats for data elements.	(X)	R&D Manager				
Determine strategic direction for data governance.					X	Senior Management
Align data governance with business goals.					X	Data Governance Team
Monitor performance of data governance.					X	Data Governance Team
Establish communication channels.					X	Data Governance Leader
Prepare meetings.					X	Data Governance Team
Educate stakeholders.	(X)	DPO			X	Data Governance Leader
Create data.					X	R&D Team
Aggregate data.	(X)	Project team	(X)	Project team	X	R&D Team
Maintain data.	X	Project team	(X)	Project team	X	R&D Team
Specify data requirements.	(X)	Project team	(X)	Project team	X	Developer
Report data related issues.	X	Developer	(X)	Project team	X	Developer

X = formal role (X) = informal role

4.1.3 Relational Governance Mechanisms

Finally, the identified relational mechanisms are presented in Table 4.3. Notably, where executive sponsorship existed for data management, communication regarding data assets would happen informally. Privacy training to mitigate compliance issues relating to GDPR was also observed in both P1 and C1. The coordination of decision-making was found to be part of the company or project culture wherein which it was executed on either a formal or informal basis. Largely, these results support existing theory on relational mechanisms.

Table 4.3: *Identified Relational Governance Mechanisms.*

Mechanism	**P1**	**P2**	**C1**
Communication	(X)		X
Training			
Privacy training	X		X
Data classifcation training			X
Coordination of Decision-making			
Hierarchical		(X)	X
Cooperative	(X)		

X = formal mechanism (X) = informal mechanism

4.2 Levels of Data Strategy Maturity

This investigation aimed to delineate the various data strategies employed, along with their defining characteristics, to establish distinct levels of maturity for each data strategy. Through interviews and literature review, four distinct levels of data maturity were identified, as well as one foundational activity for managing data assets. These have been arranged in the hierarchical order depicted in Figure 4.1.

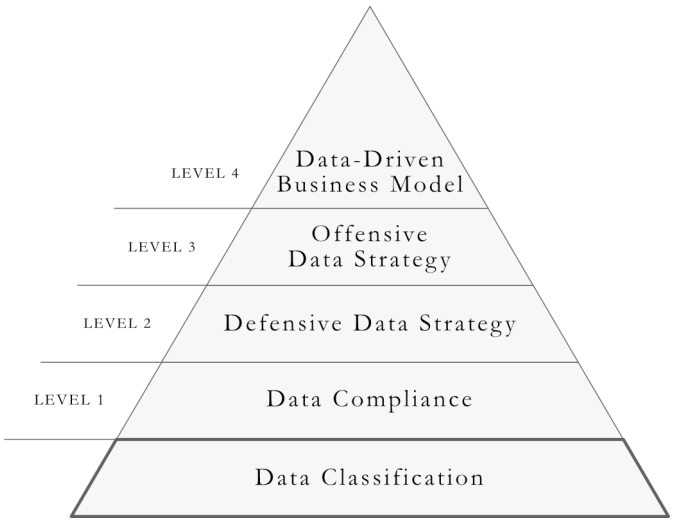


Figure 4.1: *The Data Maturity Pyramid*

In what follows, each level is described and analyzed independently and in relation to each other. Lastly, we discuss the foundational process of data classification. This step comes after describing the different levels of data strategy maturity, aiming to provide the reader with the necessary background to grasp what the foundational process entails in practice. Further, it should be noted that transitioning from one level to another does not mean abandoning the previous step. Instead, the organization incorporates the last level into its data management practices and builds upon it by acquiring additional competencies. For instance, when a company progresses from Level 1 to Level 2, they continue performing the activities of Level 1 while incorporating the additional practices of Level 2.

During the data collection process, it has become evident that interviewees consistently discuss data in terms of four distinct categories. Firstly, they distinguish between personal and non-personal data, aligning with the definition outlined in the GDPR as presented in Section 2.4.1. Further, the interviewees differentiate between internal and external data. In this context, internal data refers to data that, in one way or another, exists within the organization, e.g., production data, operational

data, financial data, or employee data. External data refers to data gathered from outside the organization, e.g., customer, competitor, market, publicly available, and partnership data. Both internal and external data can be of personal and non-personal character. Throughout this chapter, we consistently use these four data categories to present and analyze our findings. However, we do not make any claims regarding the general applicability of this categorization beyond the scope of this study.

4.2.1 Level 1: Data Compliance

Data compliance represents the initial and foundational level of strategic maturity concerning data. It involves managing data to mitigate risks associated with regulatory compliance. In the existing literature, compliance would qualify as part of a defensive data strategy. However, our findings suggest that the management of data assets to mitigate regulatory risk occurs regardless of whether intentional strategic consideration is afforded data or not. Therefore, we identify it as a distinct maturity level that any company must attain to operate at an acceptable level of risk.

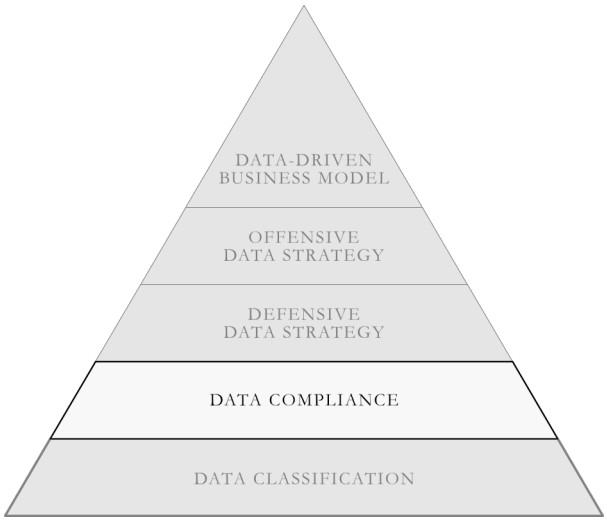


Figure 4.2: *Level 1: Data Compliance*

The need to identify data compliance as a separate level is likely attributable to two main factors. First, data assets are no longer solely collected for the explicit purpose of enhancing the strategic decision-making of firms. Instead, data exists within organizations regardless of intentional collection efforts. Following the widespread adoption of products and services that automatically gather data as an inherent part of their functionality, data is collected haphazardly. Thus, no matter the industry, data assets exist as part of the organizational capital of firms. Second, the legal landscape related to data is increasing in scope and complexity, with GDPR leading

the way and several upcoming legal frameworks worldwide. These two factors mean that all firms must manage data at least to a degree where regulatory requirements are met. Thus, it is only natural that the first level of maturity regarding the strategic management of data assets addresses regulatory compliance.

Level 1 companies are more concerned with mechanisms found in legislation rather than in data governance literature. Legal mechanisms are mandatory requirements intended to reinforce the specific interests behind the legislation. The data governance mechanisms derived from IS literature are voluntary measures to improve the efficiency of data management activities. Thus, while compliance with legal requirements serves a legitimate purpose, regulatory mechanisms are not aimed at value creation. As a result, Level 1 companies risk finding themselves in a situation where data is viewed solely as a cost without realizing the potential benefits of the data they possess.

Given that the GDPR is currently the only universally applicable data regulation, in addition to sector-specific regulations, it is understandable that the regulatory impact on data management activities is predominantly attributed to the GDPR. Consequently, Level 1 companies are primarily, if not exclusively, focused on managing personal data. While the data can theoretically encompass internal and external sources, it is likely to predominantly comprise internal data in practice. Within Level 1 companies, decision-making authority concerning data is typically centralized at higher levels of the organizational hierarchy, given its association with compliance. The subsequent sections will delve into specific findings relating to Level 1 in greater detail.

4.2.1.1 The Firm as Processor or Controller

From a legislative point of view, firms need to take on the role of either processor or controller in relation to each identified personal data asset. However, firms at this level do not consider this from a strategic perspective but from a cost perspective. Firms will view data assets as a costly byproduct of operations and attempt to minimize data management activities and reduce risk exposure. Consequently, firms express a desire to act as processors rather than controllers in relation to as many data assets as possible. The manifestation of this tends to be that firms act as controllers for data assets relating to subjects in the internal environment while acting as processors for all data assets relating to subjects in the external environment. This is visualized in Figure 4.3.

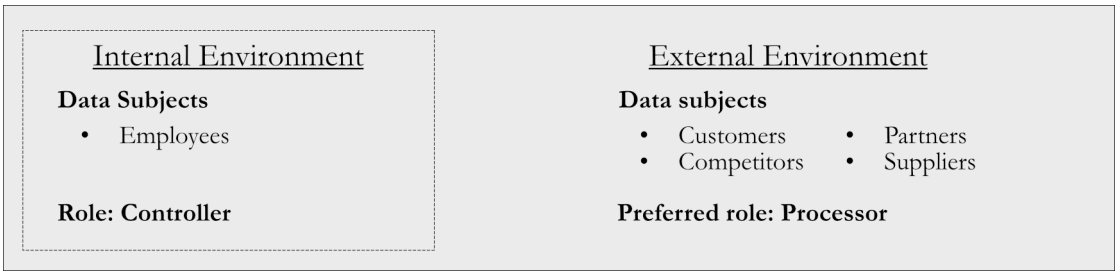


Figure 4.3: *Internal and External Data Subjects and the Preferred Role of the Firm.*

4.2.1.2 The DPO as a Structural Governance Mechanism

The introduction of the GDPR and, by extension, the DPO, clearly demonstrates how data governance is becoming part of legislation and, as such, a necessary firm capability. The DPO is, after all, a role with defined responsibilities and, thus, a structural governance mechanism. In relation to personal data, the DPO is legally obliged, indirectly or directly, to carry out responsibilities associated with multiple roles defined in IS literature. As shown in Table 4.4, these include the responsibilities of the data governance leader, the data owner, the business data steward, the technical data steward, the data government office, and the data consumer. Note that these responsibilities only extend to personal data assets.

Table 4.4: *Responsibilities Observed to be Taken by the DPO and Connected Theoretical Roles.*

Role in theory	Responsibility	Role in practice
Data Governance Leader	Ensure compliance to data policies.	**Data Protection Officer**
Data Owner	Communicate data requirements and risks within a specific business unit.	
Business Data Steward	Understand the meaning, intended representation, and the rules associated with data assets in a specific business unit. Ensure regulatory compliance.	
Technical Data Steward	Understand how data is created, stored, manipulated, and flows.	
Data Governance Office	Establish communication channels. Educate stakeholders.	
Data Consumer	Specify data requirements. Report data related issues.	

Since companies at Level 1 only classify their data based on compliance, the responsibility to ensure efficient management of data assets falls more or less solely on the DPO, who identifies and classifies personal data by educating stakeholders. This is supported by numerous interviewees who believe that the DPO is responsible for all compliance issues relating to the GDPR.

Due to how recently the GDPR was introduced, it is unsurprising that existing literature does not address DPOs. However, the introduction of a cross-hierarchical role may disrupt traditional governance structures. Therefore, future research should not only seek to include DPOs as structural mechanisms but investigate how other roles are impacted by their existence and how governance structures are created around them.

4.2.1.3 Anonymization and Pseudonymization

A common source of confusion at Level 1 firms is the anonymization of personal data. While everyone agrees on the usefulness of anonymizing data to manage data assets in compliance with the GDPR, there is no shared view of what constitutes anonymization. Legal professionals will use the definition in the GDPR, whereas the employees processing the data relies on their perception of the word 'anonymous.' Generally, what technical professionals consider anonymized, legal professionals consider pseudonymized. Consequently, training programs are instituted to create awareness of the legal definitions of the terms across the firm. In other words, the need to anonymize data triggers the use of a relational governance mechanism to decrease risk exposure. However, despite these efforts, confusion surrounding the topic tends to remain high among Level 1 and 2 companies. Thus, the source of confusion might not be a lack of education among employees.

Perhaps a more likely cause of the confusion on anonymization and pseudonymization is the evolving regulatory landscape. Since 2016, the distinction between anonymization and pseudonymization has been based on the *Breyer* ruling. It was established that if it is possible to identify a data subject in theory, then that data should be considered personal data. However, in 2023, the *SRB* ruling, while subject to appeal, the Court moved toward a narrower definition of the terms, stating that it must be possible to identify the data subject in practice. Given the uncertainty characterizing the legal field, it is perhaps unsurprising that firms find it challenging to educate their employees on the distinction between anonymized and pseudonymized data. This might also explain the lack of standardized processes for anonymizing data amongst Level 1 firms. Putting standards for managing anonymization in writing without absolute certainty that they are compliant may be considered to generate far too much risk exposure.

To conclude this section, it should be noted that because compliance has been derived from the defensive strategy level, the rationale described in this section also applies to companies operating at Level 2.

4.2.1.4 The Effect of the AI and the Data Act

Given the impact the GDPR has had on firms' data governance and data strategy, it is warranted to take a closer look at how future regulations might move the development. Note that since neither the AI Act nor the Data Act has entered into force, we can only speculate as to how these legislations will change the behavior of firms. As stated in the previous section, the reasoning relating to the AI Act and the Data Act applies to both Levels 1 and 2, given their similarities.

Firstly, regarding the AI Act, companies at Level 1 will likely only take on the user role as any other role, by definition, entails that data is used for purposes other than regulatory compliance. Taking on the role of the 'user' under the AI Act, the primary strategic consideration regards the implementation of high-risk AI systems into the firm's business activities. From a data governance perspective, the implementation of any high-risk system has one of two major implications: (1) the firm has to be able to incorporate a firm-wide capability to follow data-related directions provided by the instructions of use for the system or (2) incorporate a firm-wide capability to ensure that the input data is relevant to the intended use of the system. Thus, should high-risk systems become ubiquitous in the competitive landscape, Level 1 firms may be forced to develop data governance capabilities to preserve competitiveness and maintain an acceptable level of risk. However, how that need will be translated into specific governance mechanisms is impossible to know at the time of writing.

Secondly, regarding the Data Act, firms that offer products or services that generate data through their use will need to take on the role of the data holder. The consequence is that the firm will need to be able to (1) share generated data with users, (2) share generated data with third parties upon users' requests, and (3) inform the user of what and how much data is generated and how said data can be accessed. Given that firms at Level 1 have little to no knowledge of the business value of their data assets, we do not expect these firms to make strategic changes regarding what data is made available outside the organization. However, we expect these firms to face major data management challenges in the short term, necessitating stricter governance structures.

Data at level 1 firms tend to be generated hap-hazardously, and there are seldom structures in place that allow for identifying which data belongs to what user. Nor are there processes for making that data accessible to users or third parties. Therefore, these companies will likely need to increase their data classification efforts to maintain records. To do so, new governance mechanisms will likely need to be implemented to ensure the efficient classification of product-derived data assets. Thus, the Data Act might follow the GDPR in making governance mechanisms requirements for compliance rather than sources of competitive advantage.

Notably, none of the above-mentioned regulations include a role similar to that of the DPO in the GDPR. Therefore, firms will need to distribute responsibilities and decision-making authority to ensure compliance without clear guidance from the regulations. In other words, in addition to setting up processes to ensure compliance, firms must determine what roles and responsibilities are necessary to do so. From our perspective, we find it exceedingly likely that these roles resemble those found in data governance literature, although their responsibilities will relate to specific data assets defined in legislation rather than to data management as a whole.

4.2.2 Level 2: Defensive Data Strategy

Level 2 represents the first stage where deliberate strategic thinking is employed to leverage data assets effectively. Firms at this level focus on utilizing data assets for cost reduction and risk mitigation. Examples include process optimization and reducing data storage costs. This supports existing literature on the topic with the distinct difference of singling out compliance, as compliance can be carried out on its own without being associated with a defensive strategy (see e.g., Dallemule & Davenport, 2017; Davenport & Redman, 2020; Medeiros et al., 2020).

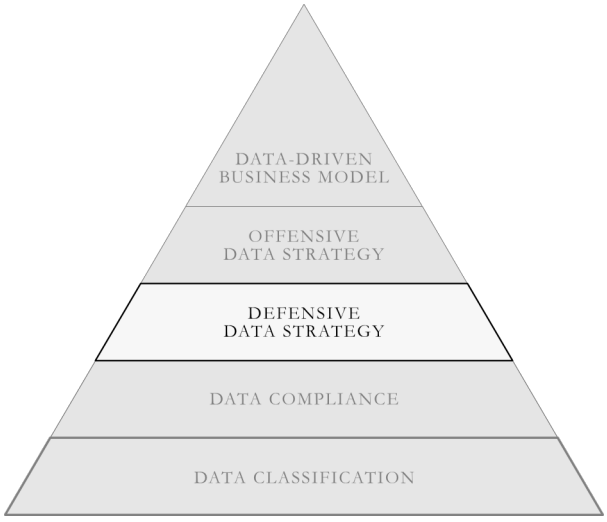


Figure 4.4: *Level 2: Defensive Data Strategy*

Moreover, the findings of this study reveal that, unlike Level 1, the data assets afforded consideration by Level 2 firms encompass more than just legal categories, incorporating additional categories of data that can effectively contribute to risk mitigation and cost reduction. This entails an emphasis on internal non-personal data, as the activities associated with a defensive strategy often revolve around, e.g., process optimization, where such data is especially useful. A defensive strategy thus tends to include active management of two distinct types of data assets: (1) personal data, as required by regulatory frameworks, i.e., the GDPR, and (2) internal non-personal data that can be used to support the objectives of a defensive strategy. The decision-making authority regarding the defensive strategy is commonly situated within the organization's lower hierarchical levels. The decisions relating to the implementation of the strategy, i.e., data governance, and the strategy itself, are determined at a hierarchically lower level compared to higher levels of strategic maturity.

4.2.2.1 Practical and Strategic Implications of the Cost Reduction Focus

The theory relating to defensive data strategy clearly outlines cost reduction as one of the primary objectives pursued within such a strategy. The findings in this study confirmed this. Cost reduction includes using data for internal process optimization. Optimization efforts are typically aimed at reducing operational or production costs. In this context, a defensive strategy does not center around the data itself but on leveraging data to support other company activities. It is worth noting that while data collection has become increasingly haphazardous compared to the past, this particular use of data aligns with a more conservative perspective where data plays a supportive role in the overall strategy rather than the other way around (see e.g., Constantiou & Kallinikos, 2015).

While data utilization for process optimization and related activities may appear abstract, companies with a defensive data strategy have also implemented measures to tangibly reduce costs directly associated with data. A concrete example of this is charging customers exceeding a specified data storage time frame. One executive at a company with a defensive strategy explained that:

> *The amount of data has only been growing and growing. Last year we implemented a limit, the [customer's] data is only saved for 1 year if they don't purchase a license. Then they can save it for 10 years. It became so slow to search the database, so we had to filter which data is saved automatically.* [our translation from Swedish]
>
> *- Global Product Manager, 2023*

This shows how the data collection facilitated by connected products and services puts pressure on data-related infrastructure. Naturally, this raises the question of who should bear the associated costs. In this case, the decision was made to pass on the costs to the customer. This choice inherently reflects a defensive strategy. The primary focus is on managing the expenses generated by the growing volume of data. The implemented actions are aimed to minimize or reduce those costs by shifting them onto the customer. No consideration is given to whether the increased quantity of data could hold inherent value in itself.

From an offensive perspective, it would be natural to evaluate whether it is worthwhile to bear the cost of increased storage space in exchange for the potential to utilize the data for revenue growth. This does not imply that employing a defensive

strategy and focusing on cost reduction is inherently the wrong choice. For certain companies, it may be the most strategically reasonable approach, yielding other valuable benefits. These benefits could include assuring customers that their data is securely stored and under their complete control. Particularly for organizations or projects managing highly sensitive data, this can be of very high value to the customer. Thus, the key takeaway is not that an offensive course of action intrinsically is preferred over a defensive one. However, the choice of which course of action to prioritize must be preceded by deliberate, strategic considerations to ensure that the choice makes business sense in the long run.

4.2.2.2 Defensive Data Strategies and Compliance

Although we recognize compliance as an independent level, separate from the defensive strategy, the influence of compliance is still evident in Level 2 firms. As described above, this level entails that data is strategically considered to some extent, in contrast to the mere compliance level. However, compliance continues to be a significant constraint on data-related activities. This can likely be attributed to the fact that a considerable proportion of the data that Level 2 firms manage is personal data. Although incorporating non-personal internal data enables a more flexible approach to some data-related activities, the coexistence of personal data necessitates compliance with regulatory frameworks. Thus, Level 2 firms demonstrate a complex position, as they find themselves simultaneously constrained by the regulatory boundaries inherent to Level 1 while also deploying a more intentional and strategic data management approach.

This is demonstrated by their careful consideration of the regulatory role the firm is willing to assume. Insofar as they relate to personal data, data-related activities are adjusted to align with the firm's intention to act as a processor rather than a controller. In practical terms, this implies that data use cases are adjusted following legislative frameworks, which serve as guidelines for determining the permissible use cases. As elaborated upon later, this stands in contrast to offensive strategies, where the intended use case is established upfront, and legislative frameworks are consulted with the purpose of finding a way to achieve the intended use in a compliant manner. Thus, the execution of a defensive strategy is typically adjusted reactively in response to legislative requirements, given that such requirements apply to the data assets at hand.

4.2.3 Level 3: Offensive Data Strategy

An offensive data strategy is characterized by a focus on revenue increase and product and service development. At this level, the firm's strategic position is to leverage data for various forms of value creation. This aligns with the existing theory by Dallemule and Davenport (2017) and Medeiros et al. (2020). Compared to a defensive strategy, implementing an offensive strategy entails several essential differences, which will be elaborated upon below, after some general remarks are made.

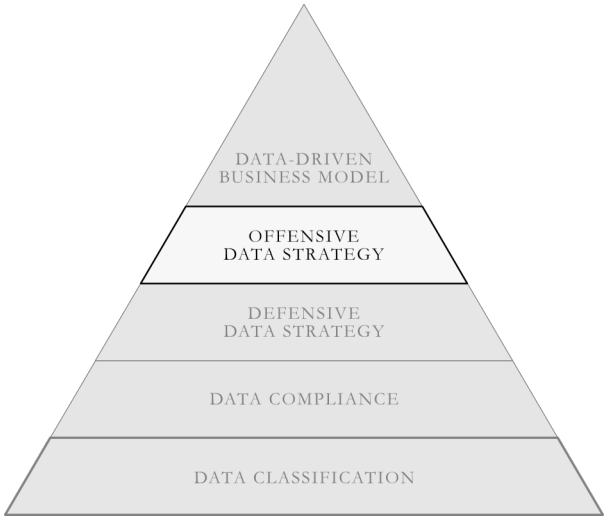


Figure 4.5: *Level 3: Offensive Data Strategy*

Our findings suggest that Level 3 firms tend to extend their data scope to include more external data of both personal and non-personal character, compared to Level 1 and 2 firms, where the emphasis is on internal data. This shift is quite significant, as it indicates that a lack of data management competence no longer constrains the data collection. Instead, organizations that embrace an offensive data strategy demonstrate their ability to effectively handle the increasing volumes of available data. Regarding decision-making authority, an offensive strategy involves a shift towards higher levels of the firm's hierarchy to determine the organization's strategy. Generally, the strategy is formulated at the executive level. However, when it comes to practical implementation decisions, there is often a lag compared to the strategy, and such decisions continue to be decentralized at lower hierarchical levels.

4.2.3.1 The Introduction of Informal Data Stewards

The fact that implementation decisions concerning the data strategy are made at lower hierarchical levels results in potential variations in the activities executed to implement the strategy, across different departments and sometimes even among

individuals. This means that, in contrast to companies with a data-driven business model, data management may still lack structure and uniformity. Level 3 companies exhibit a lack of consensus regarding the processes, structures, and formats employed to support the data strategy. Consequently, although the management of data is generally effective and relatively sophisticated at this level, the actual activities may appear fragmented and lack a cohesive structure. This implies a disconnect between the data strategy and data governance in Level 3 firms. While the organization as a whole has adopted an offensive approach towards leveraging data assets, which offers numerous benefits, there remains some confusion about how to execute its strategic insights in a coherent and well-organized manner.

This situation could potentially be attributed to the fact that certain aspects of strategy execution are carried out by individuals who assume the informal role of data stewards. The role of the data steward is a structural governance mechanism. The emphasis here is that the role is filled informally, meaning that the scope and boundaries of the role are determined at an individual level, often by highly competent individuals. However, the absence of formalized roles results in a lack of dedicated resources and structured support to fully enable the execution of the data strategy. This indicates that while the strategy is communicated formally, governance mechanisms for value creation are still implemented informally.

4.2.3.2 The Introduction of Executive Sponsorship

The findings of this study show that firms with an offensive data strategy are the first level with formal executive sponsorship. As suggested in theory, an offensive data strategy entails taking a more aggressive and proactive stance to the management of data assets (see e.g., Medeiros et al., 2020). This is necessary to facilitate the primary objectives of the offensive strategy, namely product and service development and increasing revenues.

Thus, an offensive data strategy necessitates an upfront investment, contrasting it with a defensive strategy that primarily aims to mitigate risks and minimize costs. The latter is uncontroversial and unlikely to face resistance from the executive level. When allocating funds towards generating future revenue, uncertainty and risk come into play, making executive sponsorship critical. The absence of executive sponsorship risks leading to insufficient resources being dedicated to effectively execute the strategic direction. One interviewee experienced the transition to an offensive strategy without dedicated executive sponsorship and stated that:

> *The strategy from the company side is that we need to get more active [in our data management], but when you ask the teams how they translate this vision to actual actions, they often experience that they get directions from the company, but then there are finite resources to implement it.*
>
> *- Data Protection Officer, 2023*

This relates to the informal data steward role discussed in the previous section. Such informal roles suggest that, despite executive sponsorship, the dedication of resources to create formal roles is not necessarily observable. However, to state that with certainty, and explore the possible reasons behind such a disconnect, requires an in-depth examination beyond the scope of this thesis.

Here, we will conclude by stating that executive sponsorship is a critical factor in enabling the successful execution of an offensive strategy. This distinguishes Level 3 from Levels 1 and 2, where the data strategy often inherently incorporates the necessary resources and mandates, given that the pursued objectives align with universally accepted company goals of cost minimization and compliance. The objectives of Level 3 firms require elements such as financial resources, decisional mandates, and risk acceptance, which are exceedingly difficult to obtain without the support of executive sponsorship.

4.2.3.3 The Shift from Incidental to Intentional Data Collection

A significant finding differentiating Level 3 firms from the previous levels is that the data collection moves from incidental to intentional. In Levels 1 and 2, firms acquire data incidentally through their operational activities. For instance, a Level 2 firm selling a connected product inherently collects data pertaining to its usage. This data collection is regarded as an outcome of the product's intended operation and is managed from a defensive standpoint, focusing on managing the data to minimize costs. In contrast, Level 3 firms exhibit a fundamentally different mindset, considering data collection as an independent objective.

One interviewee from a Level 3 firm shared their experience of selling a product that automatically collected some data during its regular use. However, they recognized that supplementing the incidentally collected data with additional data points could significantly enhance the commercial value offering built from that data. This involved installing extra sensors to gather data that was not essential for the product's operation. As a result, the Level 3 firm obtained data assets that were more

accurate and valuable. This example clearly illustrates how a Level 3 firm may willingly incur a temporary increase in costs, such as including "non-necessary" sensors, to generate future revenue based on data.

4.2.3.4 The Shift from Compliance as a Constraint to a Strategic Tool

Another significant distinction between Level 3 firms and the preceding levels lies in their perspective on legislation. It has been held throughout this thesis that the regulatory landscape relating to data is becoming increasingly complex. As previously discussed concerning Levels 1 and 2, regulations are frequently regarded as major constraints to data-related activities. Here, Level 3 firms demonstrate a shift in the perception of regulation and show that while legislation pressures firms' data management capabilities, it can also be leveraged to support the goals of an offensive strategy.

Firstly, it is evident that the implementation of legally required roles and processes, to some extent, mirrors the roles and processes related to data governance (see Section 4.2.1.2 for an elaborate discussion on this topic). Level 3 firms demonstrate how this can be utilized to promote the objectives of an offensive strategy. For example, the DPO of one Level 3 firm states that:

> *I take initiatives to understand new regulatory frameworks and implement measures. I help people with operations as well to make sure we are doing everything legally. Engineers want to use data in different ways, and I help them with that.*
>
> *- Data Protection Officer, 2023*

In our interviews with other team members within the firm, particularly those with engineering roles, it becomes apparent that the DPO's support and enabling mindset play a critical role in facilitating product development grounded in data. The interaction and collaboration between the engineering and legal sides are crucial in ensuring that innovation is supported but still carried out in a compliant manner. Level 1 and 2 firms express hesitance to involve their respective legal departments because the process is perceived as long, complicated, and constraining. Level 3 firms instead aim to include legal professionals early in the product development process.

Consequently, this shift allows the intended use case of the data to take center stage. This diverges from a defensive strategy where regulatory requirements take precedence, and the data use case is reactively adjusted to serve the regulatory frameworks. The defensive approach risks adopting an overly narrow interpretation of what regulatory frameworks permit. In an offensive data strategy, the data use case is established upfront, and legal frameworks are then consulted to ensure compliance while pursuing the desired use case. In practical terms, this involves engineers seeking the guidance and support of lawyers to carry out data-driven product development in a compliant manner. Legal departments play a crucial role in supporting such innovation initiatives, which may involve activities such as contract management and establishing legal constructs related to data ownership. Level 3 firms demonstrate that engineers and lawyers can, in collaboration, navigate the legal landscape while driving innovation and ensuring compliance simultaneously.

However, for an organization to fully embrace an offensive data strategy, it must be willing to assume legal responsibility. As highlighted in the discussion regarding Levels 1 and 2, the findings of this study reveal that firms often prefer to assume specific legal roles, such as opting to act as a processor rather than a controller under the GDPR. Going back to the wording of the legislation, the controller is the organization that determines the purposes and means of processing personal data. If there is a firm-wide perception that assuming this responsibility is undesirable, which Level 2 firms exhibit, it inherently restricts the possibilities for data utilization. The organization may end up constructing a situation where they are legally prohibited from determining the purposes for which data can be used, significantly limiting its ability to explore the full potential of data. Level 3 companies acknowledge the inherent value of data and, as a result, view disengaging from legal responsibility - which undoubtedly entails certain requirements - as an undesirable choice. As such, the ability to determine the purposes for which data is used is prioritized, and the legal requirements that follow are willingly managed.

This mindset also translates into more practical considerations, such as anonymization and pseudonymization. Compared to previous levels, the handling of anonymous data becomes more pragmatic for Level 3 companies. They actively evaluate the necessity of collecting personal data and, whenever possible, prefer to avoid it to foster greater freedom in their innovation activities. However, in cases where the collection of personal data is unavoidable, Level 3 firms exhibit slightly lower confusion regarding anonymization compared to Level 1 and 2 companies. This manifests

a higher threshold for when data is considered anonymous. Often, only highly aggregated, processed data, which implies less presence of personal data, is considered anonymous. Where there is confusion regarding anonymization vs. pseudonymization, the organization tends to act on the basis that the data is pseudonymized.

Level 3 companies approach the AI Act in a similar way. Firms with an offensive data strategy are heavily involved in data-driven product development, including AI. Thus, they bear greater responsibility than Level 1 and 2 firms when it comes to ensuring the quality of input data, validation data, and training data. However, Level 3 firms perceive this heightened responsibility as a potential advantage rather than a mere drawback, as it enhances the overall quality of their data assets. This, once again, demonstrates their willingness to embrace legal responsibility.

4.2.4 Level 4: Data-Driven Business Models

Our findings suggest that, in line with existing theory, the highest level of data maturity is characterized by the ability to leverage data assets to create new business models. Companies in this category have business models where the revenue-generating activities entirely depend upon data asset collection and utilization.

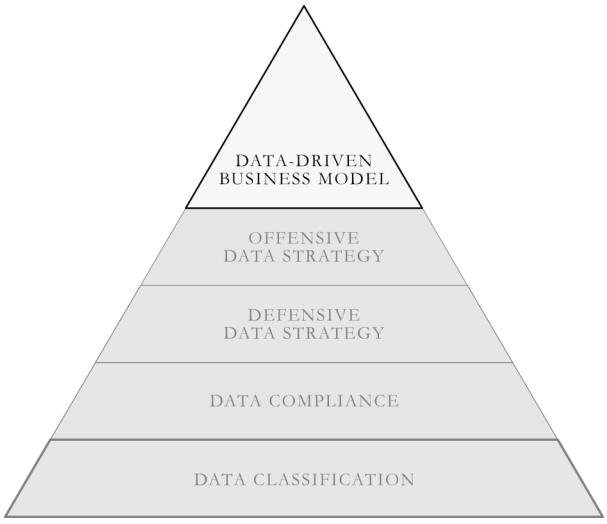


Figure 4.6: *Level 4: Data-Driven Business Model*

Level 4 firms will generally not have a data strategy separate from the overall business strategy. Instead, they are one and the same. As stated by one interviewee:

> *There is no separate data strategy, there is our business strategy which is extremely reliant on data.*
>
> *- Associate Director of Data Governance, 2023*

As is often the case with level 3 firms, strategic consideration is given to data assets by top management. The difference is that the 'data strategy' determines the strategic direction of the firm at large, necessitating decision-making to be placed at the absolute highest level. Firms at this level will also position the decision-making authority relating to data governance hierarchically higher than other firms and exhibit a much higher degree of formalization of governance mechanisms.

4.2.4.1 Formalization of Data Governance

The most apparent difference between Levels 4 and 3 is the degree to which governance mechanisms are formalized. Firstly, data-related responsibilities are clearly defined across the firm and generally situated at a hierarchically high level in the organization. Secondly, Level 4 firms prolifically implement structural governance mechanisms which are, by definition, formal. These include, but are not limited to, documented policies, standards, processes, and procedures. Level 4 firms often create incentive programs for managers to ensure that the governance mechanisms are enforced. How well teams perform their data governance activities is carefully tracked and the results are reflected in the likelihood of promotion and/or pay raise for managers.

Based on our findings, we see three reasons behind the apparent need for formalization. First, level 4 companies have reached a maturity level where data assets are considered inherently valuable. As such, level 4 firms will seek to collect as much data as possible. Since data is intangible and intrinsically difficult to define and manage, maintaining and deriving value from large and diverse volumes creates a management headache requiring strict governance. Second, increasing data volumes naturally means managing more data subject to regulation. Therefore, these firms experience a high level of risk exposure, necessitating the careful and intentional identification of data assets. Third, at the time of writing, the nine highest fines under the GDPR have been issued to Amazon, Meta, WhatsApp, and Google, all of which fall under data maturity level 4. Thus, level 4 firms are more likely to face litigation from the EU than those at levels 1 to 3. To conclude, for level 4 firms to both create value and manage risk in relation to data assets, the formalization and strict implementation of governance mechanisms are perceived as necessary.

4.2.4.2 Embedding Compliance in the Governance Structure

The formalization of governance at level 4 firms is, in part, a way to embed compliance into the firm's structure. As opposed to level 3 firms, where compliance is dealt with primarily through the knowledge contained in the minds of employees, level 4 firms embed the knowledge into a high number of formal governance mechanisms. This transfers compliance responsibilities from engineers that process data to data governance professionals and managers who develop and approve the policies, processes, standards, and procedures. Thus, the allocation of decision-making is generally situated at a hierarchically higher level than in any of the previous three maturity levels.

One interviewee emphasized the importance of individualizing data governance to the organization to ensure that the structural governance mechanisms make sense in relation to how data is actually being processed. Data governance professionals, therefore, need to continuously communicate with development teams to align the governance structure with developers' workflow. This may be the key to balancing compliance and innovation, a question often addressed in data governance literature. It is not hard to imagine that if governance mechanisms are misaligned, the firm will (1) hinder employees in their innovative endeavors and (2) create a less attractive workplace for innovative employees. This indicates the need to allocate multiple responsibilities to the same role to successfully implement data governance that manages both innovation and regulatory risk. Specifically, the role responsible for communicating with developers should also be responsible for developing structural governance mechanisms. Whether placed on a group or individual employee, such a role would require a high degree of both technical and legal expertise. Therefore, we might expect the borders between traditional IT and legal departments to be blurred as data management becomes increasingly legally complex.

4.2.5 Foundation: Data Classification

Finally, data classification has been identified as a prerequisite for any strategy-making regarding firms' data assets. Data classification refers to the identification and tagging of data assets in a manner that enables strategic management and governance. As such, we view it as a foundational activity that necessarily needs to be carried out, although to varying degrees of complexity depending on the strategic maturity.

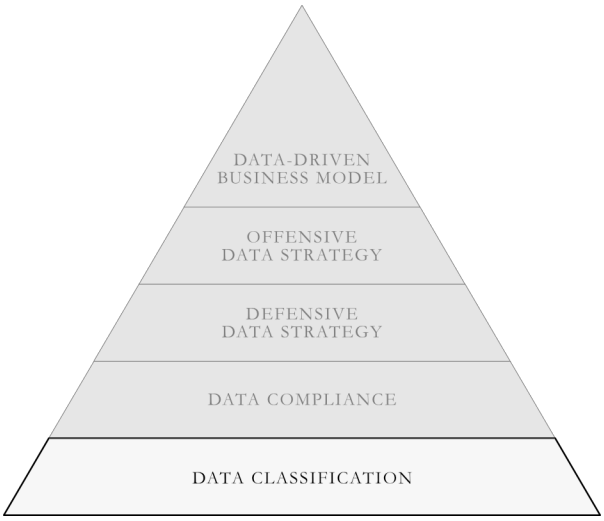


Figure 4.7: *Foundation: Data Classification*

We want to emphasize that data classification is not necessarily a formal process but an activity performed at differing degrees of formalization depending on the data strategy. To clarify, consider the following. Data classification at a level 1 firm may simply entail identifying personal data. The identification is likely performed informally, meaning an employee simply acknowledges the existence of personal data, without necessarily recording it. By contrast, a level 4 company will employ highly sophisticated and searchable databases wherein employees enter metadata according to predetermined parameters.

As visualized in Figure 4.8[1], data classification becomes more complex as data maturity increases. The reason for this is the increasing number of parameters that needs to be known about data assets the more mature a firm is. A Level 1 company

[1]To ensure there is no confusion, the graph is a very simplified version of the truth. We have no data that support that the correlation is linear or otherwise.

only needs to know if a certain data asset is subject to regulation, while a Level 4 firm may look at parameters such as use case, source, or technical traits. This likely pushes the firm to create formal governance mechanisms to manage the increasing complexity.

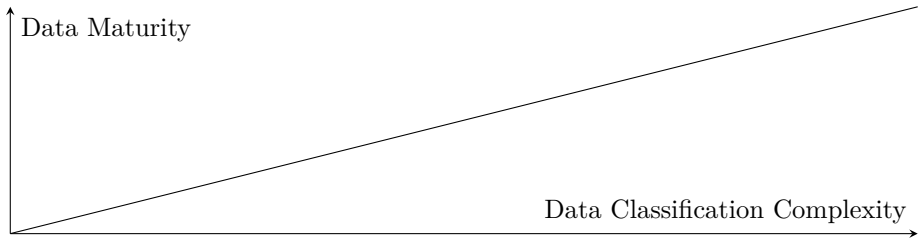


Figure 4.8: *Relationship between Strategic Maturity and Classification Complexity.*

The result shows that the parameters can be organized into four distinct categories. These are use case, legal, structural, and technical parameters. Use case refers to the intended use of the data, e.g., AI development or product improvement. Legal tagging refers to classification based on the regulatory frameworks that apply, e.g., if the data is personal data and thus triggers the GDPR's application. Structural tagging involves adding metadata to describe the structure and contents of the data semantically. It can also be used to indicate relationships between different data sets. Technical tagging refers to classifying data based on technical details, such as encoding, file format, and other technical specifications.

The number of parameters firms use to classify data assets generally increase with the level of strategic maturity. In the studied cases, Level 1 firms only classified data assets based on legal parameters. Level 2 companies performed legal and structural tagging. Level 3 firms add use cases to their data tagging efforts. Lastly, companies with a data-driven business model classified their data based on legal, structural, and technical parameters and intended use cases. Consequently, as visualized in Table 4.5, the amount of metadata an organization retains concerning its data assets grows as strategic maturity increases.

Table 4.5: *Exemplification of Classification Parameters Used at Different Levels of Data Maturity.*

Classification Parameters	**Data Compliance**	**Defensive Strategy**	**Offensive Strategy**	**Data-Driven Business Model**
Use Case Tagging			X	X
Legal Tagging	X	X	X	X
Structural Tagging		X	X	X
Technical Tagging				X

To conclude, classification allows firms to gain a more comprehensive understanding of their data assets, enabling them to make well-informed decisions about their data strategy. As a result, the relationship between data strategy and data classification can be seen as interdependent, with the level of data strategy affecting the complexity of data classification efforts within a firm, and the complexity of classification efforts, in turn, influencing the suitability of their chosen strategy.

4.3 Summary of Results

The framework presented in this thesis is a snapshot of how data maturity levels can be categorized today. It includes four distinct maturities underpinned by the foundational activity of data classification. By reviewing each individual level from a governance perspective, key characteristics of the individual levels, as well as key differences between them, have been shown.

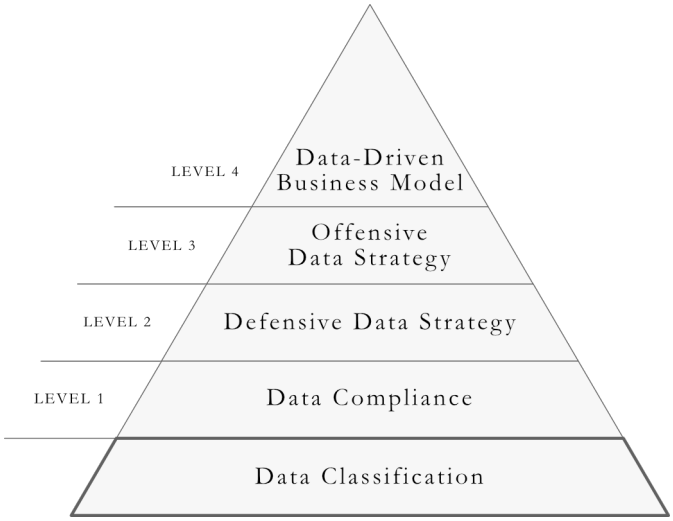


Figure 4.9: *The Data Maturity Pyramid*

Key Take-aways

- ✓ The foundation of the pyramid, *data classification*, is a prerequisite for both data strategy and data governance and is carried out at an increasing level of formalization as data maturity increases.
- ✓ **Level 1**, *data compliance*, is inhabited by firms only seeking to be compliant with data-related regulations.
- ✓ **Level 2**, *defensive data strategy*, is inhabited by firms that, in addition to achieving compliance, seek to exploit data assets to reduce costs.
- ✓ **Level 3**, *offensive data strategy*, is inhabited by firms that seek to exploit data assets to create new revenue streams.
- ✓ **Level 4**, *driven business models*, is inhabited by firms that seek to exploit data assets to create new and innovative business models.
- ✓ *Data governance mechanisms* exhibit a higher degree of formalization the more mature the firm is.
- ✓ Regulations such as the GDPR, Data Act, and the AI Act, are legally defining governance mechanisms, thus making data governance a necessary element of compliance, as well as a source of competitive advantage.

5

Discussion

The framework delineated in this thesis encompasses four distinct strategic maturities, firmly grounded in the fundamental practice of data classification. Through an examination of each level, this study unveils the essential characteristics unique to each level, and significant distinctions among them. This chapter aims to explain how the results contribute to existing theory, discuss the limitations of the data maturity model, and give suggestions for future research.

5.1 Contribution to Theory

This study contributes to the current data strategy and governance literature. It expands existing models of data strategy by differentiating between offensive data strategies (see e.g., Dallemule & Davenport, 2017; Medeiros et al., 2020) and data-driven business models (see Mazzei & Noble, 2017). Additionally, legislation and interview findings were used to extract data compliance as its own level of maturity. Regarding data governance (see e.g., Abraham et al., 2019; Davidson et al., 2023; Vial, 2023), the study introduces the idea of legally defined governance mechanisms. To the best of our knowledge, this is the first study to recognize legally defined roles and responsibilities as parts of data governance programs. On a broader level, this study bridges data strategy and data governance by viewing the latter as being the execution of the former and by highlighting an apparent correlation between strategic maturity and the formalization of governance mechanisms.

5.2 Limitations of the Model and Future Research

As all models do, the data maturity pyramid exhibits limitations. First, there is the question of what data use cases fit into the model. The model is based on the rather modern view that data assets influence strategy, not vice versa. Therefore, data collected and used for strategic decision-making is somewhat difficult to fit into the model. For it to slot into the model easily, using data for strategic decision-making

would need to be considered a capability that appears at some arbitrary level of data maturity and increases in sophistication the higher up in the pyramid you go. However, it is easy to imagine a firm that focuses on selling a service completely unrelated to data that still collects vast amounts of data about its competitive landscape that it uses in its strategic decision-making. This problem arises because our model assumes an increasing level of digitalization in relation to firms' business models as we move up the pyramid. By contrast, it seems unlikely there would be a correlation between a firm's level of digitalization and the degree to which it uses data in its strategic decision-making. Therefore, a firm at the top of the pyramid might still be immature in relation to using data in strategy-making. It would be interesting for future research to explore possible data maturity scales in relation to strategic decision-making and point out possible correlations to the maturity levels presented in this thesis.

Second, there is the question of whether level 1, data compliance, actually exists in reality. Do firms ever truly only manage data to mitigate legal risks? From a theoretical point of view, we believe this to be true. Throughout this thesis, we have maintained that data is generated by firms unintentionally, making regulatory risk stemming from data assets an inherent quality of doing business. Therefore, the lowest amount of data processing a firm theoretically can do, and still operate within the confines of the law, is to manage haphazardly collected legally defined data assets. However, we recognize that this is likely not true in the practical sense. Data assets not subject to legal requirements are probably used freely within organizations without supervision before or at the same time that structures are created for identifying and classifying legally defined assets. Perhaps a better way to view the model, in relation to levels 1 and 2, is to distinguish between processing data assets of differing logics: those defined in legislation and those not. It would be interesting for future research to take a deeper look into strategically immature firms to see how different assets are viewed and managed over a more extended time period.

This brings us to our third point, what will happen to level 1 as legislators define more and more data assets? The AI Act and Data Act have been used in this thesis to exemplify the increasing legal complexity forcing companies to manage a higher number of defined data assets. Over time, firms may find themselves in a legal environment where it is nonsensical to differentiate between legally defined data and freely used data. Thus, rather than being a distinct maturity level, compliance would be a prerequisite to having a strategy in the first place.

Fourth, is the prevalence of formal governance mechanisms observed at level 4 companies associated with data strategy or intentionality? We found that firms that deploy a data-driven business model generally exhibit a high degree of formalization of governance structures. This could be a result of implementing sophisticated strategies. However, it could also result from the intentional, rather than unintentional, strategic consideration afforded data assets. Essentially, we might expect firms that intentionally deploy a defensive strategy to also develop formal governance mechanisms. Or perhaps a correlation exists between intentionality and strategic sophistication, thus making formal structures an inherent trait of offensively positioned firms. It would be very interesting to see future research explore whether firms can execute an intentional defensive data strategy, by deploying formal governance mechanisms, or if affording data intentional strategic consideration naturally pushes the firm upwards in the pyramid.

It would also be interesting to see future research address if and how firms at different levels of strategic maturity will design new governance mechanisms to manage future legislation. The GDPR created the DPO, a role which we have observed to take on more responsibility at firms of higher strategic maturity. Will the DPO also take on responsibilities in regard to the Data Act and the AI Act, or will entirely new roles be created?

6

Conclusion

This thesis sought to answer the main research question: How do firms govern data assets depending on the maturity of their data strategy? The findings of this study reveal that the strategic data maturity of firms can be categorized into four distinct levels underpinned by the foundational data classification process. The four levels are (1) data compliance, (2) defensive data strategy, (3) offensive data strategy, and (4) data-driven business model.

First, it can be concluded that firms at all levels employ legally defined governance mechanisms, which is why we argue that legally defined roles and responsibilities should be considered as parts of data governance structures. Further, key characteristics and notable distinctions between each level were identified by examining each level from a governance perspective. As firms ascend to higher levels, the intentionality of data collection and the formalization of governance structures increases. Additionally, the view on legislation varies among firms at different levels. At Levels 1 and 2, legislation is perceived as a constraint determining how data can be used. Conversely, at Levels 3 and 4, the primary focus is the intended use case, with legislation seen as a tool to facilitate that use case in a compliant manner. Correspondingly, higher-level firms are more willing to embrace regulatory responsibilities rather than trying to evade them.

Data classification is unanimously recognized as a fundamental process for effectively managing data assets within organizations at all levels. Irrespective of a firm's strategic maturity, data assets are subjected to classification. However, as firms progress towards higher levels of strategic maturity, the complexity and the number of parameters used for data classification increase. In addition to classifying data from a legal perspective, strategically mature firms expand their classification practices to include use case tagging, structural tagging, and technical tagging. Thus, the quantity of metadata an organization possesses regarding its data assets increases in correlation with its strategic maturity.

The expanded knowledge of which data assets the firm possess equips firms with a broader foundation for making informed decisions regarding their data strategy. Consequently, the relationship between data strategy and data classification can be considered circular. The level of data strategy influences the complexity of a firm's data classification efforts, and the complexity of the classification efforts, in turn, affects the accuracy of their chosen strategy. Situated between classification and strategy, data governance acts as a practical guide, encompassing a range of mechanisms for executing the strategy effectively. At an operational level, data classification is performed by employees, establishing reporting structures that inform data governance and dictate the specific mechanisms to be employed based on the identified data assets.

Bibliography

[1] R. Abraham, J. Schneider, and J. vom Brocke, "Data governance: A conceptual framework, structured review, and research agenda," *International Journal of Information Management*, vol. 49, pp. 424–438, 2019.

[2] I. Alhassan, D. Sammon, and M. Daly, "Data governance activities: an analysis of the literature," *Journal of Decision Systems*, vol. 25, no. 1, pp. 64–75, 2016. doi: 10.1080/12460125.2016.1187397.

[3] I. Alhassan, D. Sammon, and M. Daly, "Data governance activities: A comparison between scientific and practice-oriented literature," *Journal of enterprise information management*, vol. 31, no. 2, pp. 300–316, 2018.

[4] I. Alhassan, D. Sammon, and M. Daly, "Critical success factors for data governance: A theory building approach," *Information Systems Management*, vol. 36, no. 2, pp. 98–110, 2019. doi: 10.1080/10580530.2019.1589670.

[5] M. Al-Ruithe, E. Benkhelifa, and K. Hameed, "Data governance taxonomy: Cloud versus non-cloud," *Sustainability*, vol. 10, no. 1, p. 95, 2018.

[6] M. Al-Ruithe, E. Benkhelifa, and K. Hameed, "A systematic literature review of data governance and cloud data governance," *Personal and Ubiquitous Computing*, vol. 23, no. 5, pp. 839–859, 2019.

[7] J. Barney, "Firm resources and sustained competitive advantage," *Journal of Management*, vol. 17, no. 1, pp. 99–120, 1991.

[8] O. Benfeldt, J. S. Persson, and S. Madsen, "Data governance as a collective action problem," *Information Systems Frontiers*, vol. 22, pp. 299–313, 2020.

[9] P. Bindley, "Joining the dots: how to approach compliance and data governance," *Network Security*, vol. 2019, no. 2, pp. 14–16, 2019.

[10] S. Black, M. Davern, S. B. Maynard, and H. Nasser, "Data governance and the secondary use of data: The board influence," *Information and Organization*, p. 100447, 2023.

[11] H. Borgman, H. Heier, B. Bahli, and T. Boekamp, "Dotting the i and crossing (out) the t in it governance: New challenges for information governance," *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pp. 4901–4909, 2016.

[12] P. Brous, M. Janssen, and R. Vilminko-Heikkinen, "Coordinating decision-making in data management activities: A systematic review of data governance principles," *Electronic Government*, pp. 115–125, 2016.

[13] E. Brynjolfsson and A. McAfee, *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company, 2014.

[14] C-582/14, judgment of 19 October 2016, Patrick Breyer v Bundesrepublik Deutschland, ECLI:EU:C:2016:779.

[15] S. D. A. Castellanos, G. Gállego, and J. R. Robles, "Sending personal data, receiving non-personal data: Recent eu judgment reinforces the power of pseudonymization," 2023.

[16] California Consumer Privacy Act, CA Civ Code § 1798.192 (2018).

[17] I. Constantiou and J. Kallinikos, "New games, new rules: Big data and the changing context of strategy," *Journal of Information Technology*, vol. 30, pp. 44–57, 2015.

[18] L. Dallemule and T. H. Davenport, "What's your data strategy?," *Harvard Business Review*, vol. 95, no. 3, pp. 112–121, 2017.

[19] D. International, *DAMA-DMBOK: data management body of knowledge*. Technics Publications, LLC, 2017.

[20] P. Dave, "Stack overflow will charge ai giants for training data," 2023.

[21] T. H. Davenport and T. C. Redman, "Your organization needs a proprietary data strategy," *Harvard Business Review Digital Articles*, pp. 2–5, 2020.

[22] E. Davidson, L. Wessel, J. S. Winter, and S. Winter, "Future directions for scholarship on data governance, digital innovation, and grand challenges," *Information and Organization*, vol. 33, no. 1, p. 100454, 2023.

[23] V. Diamantopoulou, A. Tsohou, and M. Karyda, "From iso/iec27001: 2013 and iso/iec27002: 2013 to gdpr compliance controls," *Information & Computer Security*, vol. 28, no. 4, pp. 645–662, 2020.

[24] V. Diamantopoulou, A. Tsohou, and M. Karyda, "From iso/iec 27002: 2013 information security controls to personal data protection controls: guidelines for gdpr compliance," *Computer Security: ESORICS 2019 International Workshops*, pp. 238–257, 2020.

[25] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure.

[26] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

[27] J. Drexl, "Designing competitive markets for industrial data," *J. Intell. Prop. Info. Tech. & Elec. Com. L.*, vol. 8, p. 257, 2017.

[28] J. W. Drisko, "Strengthening qualitative studies and reports," *Journal of Social Work Education*, vol. 33, no. 1, pp. 185–197, 1997.

[29] J. Dyché and A. Polsky, "Models for data stewardship," *SAS best practices white paper*, 2016.

[30] P. V. Dyck, F. Miotto, and E. Keeling, "The eu data act - what could this mean for you?," 2022.

[31] L. Edwards, "The eu ai act proposal: A description of the significance of the eu ai act, its scope and main points," 2022.

[32] L. Einav and J. Levin, "The data revolution and economic analysis," *Innovation Policy and the Economy*, vol. 14, no. 1, pp. 1–24, 2014.

[33] M. Fazlioglu, "Beyond the nature of data: Obstacles to protecting sensitive information in the european union and the united states," *Fordham Urb. LJ*, vol. 46, p. 271, 2019.

[34] M. Finck and F. Pallas, "They who must not be identified—distinguishing personal from non-personal data under the gdpr," *International Data Privacy Law*, vol. 10, no. 1, pp. 11–36, 2020.

[35] M. R. Franke, M. Grebe, D. Kleine, V. Lukic, L. Quarta, and K. Rogers, "Any company can become a resilient data champion," *Boston Consulting Group + Google*, 2023.

[36] B. Gates, "The age of ai has begun," 2023.

[37] C. Geertz, *Thick description: Toward an interpretive theory of culture*, pp. 41–51. New York: Basic Books, 1973.

[38] M. O. Gokalp, K. Kayabay, M. A. Akyol, P. E. Eren, and A. Kocyigit, "Big data for industry 4.0: A conceptual framework," IEEE.

[39] G. Greenleaf, "The 'brussels effect'of the eu's 'ai act'on data privacy outside europe," 2021.

[40] R. L. Grossman, "A framework for evaluating the analytic maturity of an organization," *International Journal of Information Management*, vol. 38, no. 1, pp. 45–51, 2018.

[41] E. G. Guba and Y. S. Lincoln, "Competing paradigms in qualitative research," *Handbook of qualitative research*, vol. 2, no. 163-194, p. 105, 1994.

[42] U. Gupta and S. Cannon, *A Practitioner's Guide to Data Governance : A Case-Based Approach*. Emerald Publishing Limited, 2020.

[43] A. Hagiu and J. Wright, "When data creates competitive advantage," *Harvard Business Review*, vol. 98, no. 1, pp. 94–101, 2020.

[44] E. C. Hettinger, "Justifying intellectual property," *Philosophy & Public Affairs*, pp. 31–52, 1989.

[45] Q. Hu, P. Hart, and D. Cooke, "The role of external and internal influences on information systems security–a neo-institutional perspective," *The Journal of Strategic Information Systems*, vol. 16, no. 2, pp. 153–172, 2007.

[46] P. B. Hugenholtz *et al.*, "Data property: Unwelcome guest in the house of ip," 2017.

[47] M. Jagals, E. Karger, and F. Ahlemann, "Already grown-up or still in puberty? a bibliometric review of 16 years of data governance research," *Corporate Ownership and Control*, vol. 19, no. 1, pp. 105–120, 2021.

[48] W. Kerber, "A new (intellectual) property right for non-personal data? an economic analysis," *Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil (GRUR Int)*, vol. 11, pp. 989–999, 2016.

[49] V. Khatri and C. V. Brown, "Designing data governance," *Communications of the ACM*, vol. 53, no. 1, pp. 148–152, 2010.

[50] M. N. Kooper, R. Maes, and E. E. O. R. Lindgreen, "On the governance of information: Introducing a new concept of governance to support the management of information," *International Journal of Information Management*, vol. 31, no. 3, pp. 195–200, 2011.

[51] K. Krell, S. Matook, and F. Rohde, "The impact of legitimacy-based motives on is adoption success: An institutional theory perspective," *Information & Management*, vol. 53, no. 6, pp. 683–697, 2016.

[52] J. Ladley, *Data Governance: How to Design, Deploy and Sustain an Effective Data Governance Program*. Elsevier Science Technology, 2012.

[53] J. Ladley, *Data Governance : How to Design, Deploy, and Sustain an Effective Data Governance Program*, vol. 2. Academic Press, 2020.

[54] C. Lambrinoudakis, *The General Data Protection Regulation (GDPR) Era: Ten Steps for Compliance of Data Processors and Data Controllers*, pp. 3–8. Springer International Publishing, 2018.

[55] D. Lametti, "The concept of property: relations through objects of social wealth," *The University of Toronto Law Journal*, vol. 53, no. 4, pp. 325–378, 2003.

[56] Y. Lee, S. Madnick, R. Wang, F. Wang, and H. Zhang, "A cubic framework for the chief data officer (cdo): Succeeding in a world of big data," vol. 13, 2014.

[57] M. Leininger, *Evaluation criteria and critique of qualitative studies*. Newbury Park, CA: Sage, 1994.

[58] T. Lillie and S. Eybers, "Identifying the constructs and agile capabilities of data governance and data management: A review of the literature," in *Locally Relevant ICT Research* (K. Krauss, M. Turpin, and F. Naude, eds.), pp. 313–326, Springer International Publishing.

[59] Y. S. Lincoln and E. G. Guba, *Naturalistic inquiry*. CA: Sage, 1985.

[60] T. A. Lipinski and J. Britz, "Rethinking the ownership of information in the 21st century: Ethical implications," *Ethics and information technology*, vol. 2, no. 1, p. 49, 2000.

[61] D. Lis and B. Otto, *Data Governance in Data Ecosystems – Insights from Organizations*. 2020.

[62] M. J. Mazzei and D. Noble, "Big data dreams: A framework for corporate strategy," *Business Horizons*, vol. 60, no. 3, pp. 405–414, 2017.

[63] A. McAfee, E. Brynjolfsson, T. H. Davenport, D. Patil, and D. Barton, "Big data: the management revolution," *Harvard business review*, vol. 90, no. 10, pp. 60–68, 2012.

[64] M. M. d. Medeiros, A. C. G. Maçada, and J. C. d. S. Freitas Junior, "The effect of data strategy on competitive advantage," *The Bottom Line*, vol. 33, no. 2, pp. 201–216, 2020.

[65] T. W. Merrill, "Property and sovereignty, information and audience," *Theoretical Inquiries in Law*, vol. 18, no. 2, pp. 417–445, 2017.

[66] D. Nield, "How chatgpt and other llms work—and where they could go next," 2023.

[67] O. Nielsen, "A comprehensive review of data governance literature (selected papers of the iris, n. 8)," *Atlanta, GA: Association for Information Systems*, 2017.

[68] Opinion 4/2007 on the concept of personal data (01248/07/EN WP 136). 2007. Article 29 Working Party.

[69] Opinion 05/2014 on Anonymisation Techniques (0829/14/EN WP216). 2014. Article 29 Working Party.

[70] B. Otto, *A morphology of the organisation of data governance*. 2011.

[71] B. Otto, "Organizing data governance: Findings from the telecommunications industry and consequences for large service providers," *Communications of the Association for Information Systems*, vol. 29, p. Article 3, 2011.

[72] J. E. Penner, "The bundle of rights picture of property," *UcLa L. rev.*, vol. 43, p. 711, 1995.

[73] U. Petrusson, *Intellectual Property & Entrepreneurship: Creating Wealth in an Intellectual Value Chain*. Center for Intellectual Property Studies (CIP), 2004.

[74] D. Plotkin, *Data Stewardship: An Actionable Guide to Effective Data Management and Data Governance*. San Diego: Elsevier Science Technology, 2020.

[75] M. E. Porter, "How competitive forces shape strategy," *Harvard Business Review*, vol. 57, no. 2, pp. 137–145, 1979.

[76] Privacy Amendment (Notifiable Data Breaches) Act 2017 (No. 12, 2017).

[77] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (COM/2021/206 final).

[78] Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act) (COM/2022/68 final).

[79] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation).

[80] J. Ritter and A. Mayer, "Regulating data as property: a new construct for moving forward," *Duke L. & Tech. Rev.*, vol. 16, p. 220, 2017.

[81] D. M. Rousseau, "Issues of level in organizational research: Multi-level and cross-level perspectives," *Research in organizational behavior*, 1985.

[82] R. Schüritz, S. Seebacher, G. Satzger, and L. Schwarz, *Datatization as the Next Frontier of Servitization – Understanding the Challenges for Transforming Organizations*. 2017.

[83] J. Scott, *A matter of record : documentary sources in social research*. Cambridge : Polity, 1990.

[84] J. R. Searle, *The construction of social reality*. Free Press, 2007.

[85] S. Sharma, *Data privacy and GDPR handbook*. John Wiley Sons, 2019.

[86] A. Siddiqa, I. A. T. Hashem, *et al.*, "A survey of big data management: Taxonomy and state-of-the-art," *Journal of Network and Computer Applications*, vol. 71, pp. 151–166, 2016.

[87] S. Skinner-Thompson, "Outing privacy," *Nw. UL Rev.*, vol. 110, p. 159, 2015.

[88] R. E. Stake, *The art of case study research*. Thousand Oaks, CA: Sage, 1995.

[89] I. Stepanov, "Introducing a property right over data in the eu: the data producer's right–an evaluation," *International Review of Law, Computers & Technology*, vol. 34, no. 1, pp. 65–86, 2020.

[90] E. Stephens, "The intersection of data quality and compliance," *Forbes Innovation*, 2021.

[91] T-557/20, judgment of 26 April 2023, Single Resolution Board v European Data Protection Supervisor, ECLI:EU:T:2023:219.

[92] P. P. Tallon, R. V. Ramirez, and J. E. Short, "The information artifact in it governance: Toward a theory of information governance," *Journal of Management Information Systems*, vol. 30, no. 3, pp. 141–178, 2013.

[93] D. J. Teece, "Explicating dynamic capabilities: the nature and microfoundations of (sustainable) enterprise performance," *Strategic Management Journal*, vol. 28, no. 13, pp. 1319–1350, 2007.

[94] R. Thoppilan *et al.*, "Lamda: Language models for dialog applications," *arXiv pre-print server*, 2022.

[95] B. Townsend, "Decoding the proposed eu ai act," American Society of International Law.

[96] B. van der Sloot, *Regulating non-personal data in the age of Big Data*, pp. 85–105. Routledge, 2020.

[97] S. van Erp, "Ownership of data: the numerus clausus of legal objects," *Brigham-Kanner Prop. Rts. Conf. J.*, vol. 6, p. 235, 2017.

[98] H. R. Varian, "Beyond big data," *Business Economics*, vol. 49, no. 1, pp. 27–31, 2014.

[99] M. Veale and F. Z. Borgesius, "Demystifying the draft eu artificial intelligence act — analysing the good, the bad, and the unclear elements of the proposed

approach," *Computer Law Review International*, vol. 22, no. 4, pp. 97–112, 2021.

[100] G. Vial, "Data governance and digital innovation: A translational account of practitioner issues for is research," *Information and Organization*, vol. 33, no. 1, p. 100450, 2023.

[101] B. Vigliarolo, "Reddit: If you want to slurp our api to train that llm, you better pay for it, pal," 2023.

[102] I. Wallis, *Data Strategy : From definition to execution*. BCS Learning & Development Limited, 1st ed., 2021.

[103] K. Weber, B. Otto, and H. Oesterle, "One size does not fit all - a contingency approach to data governance," *ACM Journal of Data and Information Quality*, vol. 1, p. Article 4, 2009.

A

On the Ownership of Data

The concepts of ownership and property are fundamental to our legal and societal systems, as it establishes a framework for assigning rights and responsibilities (Hettinger, 1989). The foundations of our intellectual property systems are based on fundamental notions of ownership, even though the property in question is intangible (Lipinski & Britz, 2000). In recent years, as digital technology has become more ubiquitous and data has become a valuable commodity, the notion of data ownership has increased in both actuality and complexity. While individuals and organizations may create, collect, and store data, it is not always clear who has the right to control or profit from it. Ownership of data refers to information matter over which one has stated rights, including the ability to use, possess, sell, or prevent others from accessing it (Lametti, 2003; Lipinski & Britz, 2000; van Erp, 2017). The rights that are assigned to the matter is what qualifies the matter, i.e. the data, as property (Merrill, 2017; Penner, 1995). While data is increasingly being treated as a form or property, which is manufactured, transferred, licensed, and sold, it has to be considered that there is no explicit, legal mechanism assigned to data that clarifies the control and access rights to such assets (Ritter & Mayer, 2017). Thus, the question of whether data can be owned in the first place is yet to be settled (Stepanov, 2020).

This does not mean that the area is unregulated, but that the regulatory landscape is fragmented. The *sui generis* protection of databases offer some protection for data-related assets, but lacks in both scope and clarity (Hugenholtz, 2017). The GDPR grants rights to data subjects where personal data is concerned. However, no privacy or data protection laws expressly define which entity owns personal information. On the contrary, the GDPR, with reference to the European Charter of Fundamental Rights, denotes personal data as rights subject to special consideration over which no property rights could be exercised (Stepanov, 2020). The same can be said for imminent legislations, e.g., the Data Act, which imposes significant responsibilities to 'data holders', but fails to define ownership of data.

The lack of ownership mechanisms in existing regulatory frameworks result in a higher dependency on trade secrets and contractual mechanisms. There is no legal barrier preventing data from qualifying as a trade secret. Here, the issue is instead that trade secrets protection fails to assign a property right to the objects of trade secret protection. Instead, the rules are focused on liability for tortious acts, i.e., sanctioning unlawful access to the data (Art. 4 Trade Secrets Directive). Contract law, in itself, does not assign a property right to data. Nevertheless, by regulating data access and limiting how data can be used and shared, contracts can be used as a tool for 'simulating' a legal relationship similar to property protection (Drexl, 2017; Kerber, 2016). The obvious drawback is that these rights only apply to the contracting parties, which is where contractual relationships profoundly differ from property rights.

The dependency on trade secret management and contractual mechanisms for de facto regulation of data ownership provides a significant insight. The basic prerequisite for using trade secrets and contracts to construct 'data ownership' is that the data holder has effective control over the data. Otherwise, the data holder is in no position to negotiate contracts or employ trade secret qualification practices, as the data is already publicly available or outside the data holder's control. Interdependently, trade secret and contract management contribute to facilitating such control, along with other control mechanisms. To this background, this thesis will not build on the notion of data ownership, as it is to fragmented and unclear to provide a basis for treating data as an asset. Instead, the thesis will focus on the idea of data control facilitated through legal, technical, and business-related mechanisms as the primary tool for qualifying data as an asset (Petrusson, 2004).

B

List of Interviewees

Role	Pre-Interview	Main Interview	Validation Interview
Senior or Director Level Product Managers/Engineers			
Global Product Manager	X	X	X
Head of Product IT	X	X	X
PLM Program Manager	X		
Product Manager	X		
R&D Cloud Manager	X		
Innovation Directors and Managers			
Concept Innovation Manager	X	X	
Head of Digital Eco-Systems	X	X	X
R&D Software and Test Manager	X	X	X
Senior Director of Innovation	X		
Data Protection Managers and Senior Level Legal Councils			
Data Protection Manager #1	X	X	
Data Protection Manager #2	X	X	X
Head of Legal	X	X	
Manager, Regulatory Affairs	X		
Senior Trademark & IP Counsel	X	X	
Staff Engineer, IP R&D	X	X	
Trademark Counsel	X	X	
Data Governance Directors			
Associate Director of Data Governance	X	X	
Executive/VP Level			
Chief Technology Officer	X	X	
VP & Head of R&D Services	X		
Software and Product Developers			
Enterprise Architect	X		
FW Developer	X		
Product & Software Director	X		

DEPARTMENT TECHNOLOGY MANAGEMENT AND ECONOMICS
DIVISION OF ENTREPRENEURSHIP AND STRATEGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY