Master’s Programme in Information and Service Management

Perceptions of Data Contracts and Data Sharing Agreements

Case UPM

Iisa Paananen

Master’s thesis
2025

Aalto-yliopisto
Aalto-universitetet
Aalto University

Author Iisa Paananen

Title of thesis Perceptions on Data Contracts and Data Sharing Agreements

Programme Master of Science (Economics and Business Administration)

Major Information and Service Management

Thesis supervisor Dr Riitta Hekkala

Thesis advisor(s) Sakari Himanen

Collaborative partner UPM Kymmene Oyj

Date 27.10.2025 **Number of pages** 104 **Language** English

Abstract

In today's data driven world, new developments in data management are constantly sought after. Data Contracts and Data Sharing Agreements are presented as a potential resolution to some issues facing the data management world, including data security, data sharing, and facilitating improvements in data culture. This thesis contributes to the existing literature in Data Contracts and Data Sharing Agreements by examining how they are seen by different employees in a case company. By conducting semi-structured interviews with 10 participants in data-related roles in three separate business units within the same company, this thesis set out to find hopes, fears, value and pain points related to the use and implementation of Data Contracts and Data Sharing Agreements, both within the business areas and the organization as a whole.

The main findings of this study, identified through thematic analysis, are focused on the forces that the participants feel are hindering the adoption of Data Contracts and Data Sharing Agreements. These include factors such as the pull between local autonomy and organizational steering, challenges in balancing data protection with data democratization, as well as unclarity in roles, ownership and the leading of data culture. This thesis presents concrete managerial suggestions to improve the situation, such as clarifying roles and responsibilities and prioritizing groundwork in data work.

Keywords Data Contract, Data Sharing Agreement, Data Governance, Data Management, Data Culture, Data Democratization, Data Protection

Aalto-yliopisto
Aalto-universitetet
Aalto University

Tekijä Iisa Paananen	
Työn nimi Näkemyksiä datakontrakteista ja datanjakosopimuksista	
Koulutusohjelma Kauppatieteiden Maisterin tutkinto	
Pääaine Tieto- ja palvelujohtaminen	
Vastuuopettaja/valvoja Tohtori Riitta Hekkala	
Työn ohjaaja(t) Sakari Himanen	
Yhteistyötaho UPM Kymmene Oyj	
Päivämäärä 27.10.2025 **Sivumäärä** 104.	**Kieli** Englanti

Tiivistelmä

Tiedonhallinta ottaa uusia kehitysaskeleita jatkuvasti nykypäivän datavetoisessa maailmassa. Datakontrakteja ja datanjakosopimuksia on esitetty mahdollisina ratkaisuina tiedonhallinnan ongelmiin, kuten tietoturvaan, tiedonjakoon ja datakulttuurin parantamiseen. Tämä opinnäytetyö täydentää olemassa olevaa datakontrakteja ja datanjakosopimuksia käsittelevää kirjallisuutta tutkimalla, miten eri työntekijät näkevät ne tapausyrityksessä. Tutkielmassa haastateltiin kymmentä dataan liittyvissä tehtävissä toimivaa henkilöä kolmessa liiketoimintayksikössä saman yrityksen sisällä. Näin pyrittiin löytämään toiveita, pelkoja, arvoja ja kipukohtia, jotka liittyvät datasopimusten ja tiedonjakosopimusten käyttöön ja täytäntöönpanoon eri alueilla, mutta myös koko organisaation laajuisesti.

Pro Gradu- tutkielman keskeisimmät tulokset, jotka tunnistettiin temaattisen analyysin avulla, liittyvät haastateltavien kokemuksiin datakontraktien ja datanjakosopimusten käyttöönottoa hidastavista tekijöistä. Näitä tekijöitä ovat esimerkiksi liiketoimintayksiköiden autonomian ja organisaatiotason ohjauksen välinen ristiriita, haasteet tietosuojan ja datan demokratisoinnin tasapainottamisessa sekä epäselvyydet roolien, omistajuuden ja datakulttuurin johtamisen kanssa. Tämä opinnäytetyö esittää tulosten pohjalta konkreettisia kehitysehdotuksia, kuten roolien ja vastuiden selkeyttämistä sekä datatyön pohjatyön priorisointia.

Avainsanat Datakontrakti, datanjakosopimus, tietohallinta, datahallinta, tietojohtaminen, datajohtaminen, datakulttuuri, datademokratisaatio, tietosuoja

Table of contents

Acknowledgements

I want to sincerely thank my Thesis Supervisor Dr. Riitta Hekkala, whose guidance, direction and support in writing this thesis were invaluable. Additionally, I want to thank my Thesis Advisor Sakari Himanen, who ideated, inspired and believed in this thesis – and me – from day one.

This thesis would not have seen the light of day without the unwavering support I received from my partner, family, friends and colleagues. My people make me who I am, and I am forever grateful to be of you.

Helsinki, 27 October 2025
Iisa Paananen

Abbreviations

ACER	EU Agency for the Cooperation of Energy Regulators
ADKAR	Awareness, Desire, Knowledge, Ability, Reinforcement
AI	Artificial Intelligence
API	Application Programming Interface
BA	Business Area
BI	Business Intelligence
CCPA	California Consumer Privacy Act
CDP	Common Data Platform
CIA	Confidentiality, Integrity, Availability
CPRA	California Privacy Rights Act
CSPG	Culture, Staffing, Processes, Governance
DaaP	Data as a Product
dbt	Data Build Tool
DC	Data Contract
DCat	Data Catalog
DD	Data Democratization
DDMO	Digital and Data Management Office
DG	Data Governance
DLT	Distributed Ledger Technologies
DP	Data Product
DPM	Data Product Manager
DSA	Data Sharing Agreement
ERP	Enterprise Resource Planning
FAIR	Findable, Accessible, Interoperable, Reusable
FC	Finance and Controlling
FERPA	Family Educational Rights and Privacy Act
GDPR	General Data Protection Regulation
HIPAA	Health Insurance Portability and Accountability Act
HoD	Head of Data
IA	Information Architect
IAWG	Information Architects' Working Group
IT	Information Technology
LGPD	Brazilian General Data Protection Law
OCDS	Open Contracting Data Standard
PIPA	South Korean Personal Information Protection Act
RFID	Radio frequency identification
SLA	Service Level Agreement
UPM	UPM Kymmene Oyj

1 Introduction

This thesis looks into Data Contracts and Data Sharing Agreements, as they are seen by different employees in a case company. By interviewing 10 different people in data-related roles within the same company, this thesis aims to find hopes, fears, value and pain points related to the use and implementation of Data Contracts and Data Sharing Agreements.

1.1 Background and Motivation

This Master's Thesis is conducted in collaboration with a Finnish forest industry company UMP Kymmene Oyj (UPM). The company is comprised of several different business areas (BAs), which operate under the same conglomerate. Within this thesis, when discussing the whole organization, it is referred to as the Group. Some functions are global and serve the entire Group, including the information technology (IT), legal and sourcing departments. UPM is vertically integrated, as it owns multiple stages of production and the supply chain, from raw materials to final products. As the different business areas can act as each other's suppliers and competitors in different instances, information sharing principles must consider ethical and competitive limitations and avoid possible conflicts of interest.

Similarly to how shared functions serve multiple business areas, some data in UPM is also utilized by multiple BAs. Data development in the Group is done both 'in-house' by the business areas, but also in a centralized manner under the global IT department. This means that data pertaining to separate business areas, which could be deemed sensitive or in breach of competition law if shared, can be managed and shared forward by the same IT entity. Drawing the line between useful information sharing (e.g. sharing best practices or shared semantics) within business areas and shielding sensitive information (e.g. financials or personnel data) from being misplaced is not an easy task. Through Data Contracts (DC) and Data Sharing Agreements (DSA), data flows can be technically and contractually safeguarded from misuse, alleviating the risk of competitive breaches or loss of value.

The growing complexity of modern data ecosystems has solidified the need for robust, transparent, and interoperable governance mechanisms (Hassani et al., 2025; Abraham et al., 2019). As organizations increasingly adopt decentralized data architectures, traditional models of control and accountability are being redefined (Dehghani, 2019; Ahlawat et al., 2023; Ravva, 2025). In this evolving landscape, Data Contracts and Data Sharing Agreements have emerged as possible instruments for formalizing and contractually agreeing on expectations around data quality, access, as well as rights and responsibilities between producers and consumers. These solutions, both in their own ways, aim to balance the need for agility and innovation with the need for control, compliance, and trust (Khatun, 2024; Segner, 2022; Sassoon, 2023; Petersen, 2019).

Examining DCs and DSAs within a large, vertically integrated organization such as UPM provides a valuable opportunity to observe how these emerging governance instruments interact with complex organizational realities, such as a plethora of stakeholders, misaligning business semantics, as well as unclear accountability. Studying and shedding light on these phenomena contributes to a broader academic conversation on data governance evolution, organizational change and trust in data ecosystems.

DCs and DSAs, which are relative newcomers in the data scene, have not been thoroughly studied academically. Much of the existing literature is produced by companies and individual actors, such as data platform businesses and consulting companies. Therefore, much of the focus is more on the managerial and practical topics. While this focus can be very useful in providing concrete examples and real-world implementations, the material does not go through academic rigor. This thesis aims to participate in closing the research gap evident here by providing an academical case study into this world. The key focus of this thesis and its findings is on the factors hindering DC and DSA adoption.

1.2 Research Questions and Objectives

This study aims to find commonalities and differences between how Data Contracts and Data Sharing Agreements are perceived in different business areas under UPM, which tools are best suited to their needs, and to provide suggestions on future data governance tasks based on these findings.

This thesis aims to answer the following research questions:

1. What value do Data Contracts and Data Sharing Agreements provide?
2. How can this value be realized?

To the best of the author's knowledge, no case studies aiming to explore the early-stage implementation of Data Contracts and Data Sharing Agreements have been conducted, underscoring a notable gap in the literature.

1.3 Structure of the Thesis

This thesis is structured into six main chapters, each contributing to a comprehensive understanding of how Data Contracts and Data Sharing Agreements are applied and perceived within the context of data governance and organizational change. The structure follows a logical progression from theoretical grounding to empirical analysis and interpretation.

Chapter 1 establishes the background, motivation, and research objectives of the study. It outlines the relevance of examining Data Contracts and Data Sharing Agreements in the evolving landscape of data governance and articulates the central research questions guiding the investigation.

Chapter 2 provides the theoretical foundation for the study. The literature review begins with data governance and data security principles before turning to organizational data practices, including data sharing, democratization, and the conceptualization of data as a product. The chapter then introduces Data Contracts and Data Sharing Agreements, tracing their conceptual origins, legal and technical dimensions, examining existing implementations, and related case studies. The second chapter concludes in organizational factors of trust and resistance that influence data governance adoption.

Chapter 3 describes the empirical design of the study. It introduces the company and case context, outlines the qualitative research methods employed, and explains the procedures for data collection and analysis. Considerations regarding research ethics are also addressed. This chapter establishes the methodological transparency and rigor necessary for evaluating the study's findings.

Chapter 4 presents the empirical results derived from the case study. The findings are organized around three thematic areas: system-level governance tensions, organizational readiness and roles, and data sharing paradigms. Each theme explores the dynamics and challenges observed in the early implementation of Data Contracts and Data Sharing Agreements within the organization.

Chapter 5 interprets the findings in light of existing literature and theoretical perspectives. It discusses how the observed patterns align with or diverge from prior research, evaluates how the findings address the research questions, and outlines both managerial and practical implications for organizations seeking to operationalize data governance through Data Contracts and DSAs.

Chapter 6 summarizes the key insights and contributions of the thesis. It reflects on the study's limitations and proposes avenues for future research, particularly in relation to the broader applicability of Data Contracts and Data Sharing Agreements across different organizational and sectoral contexts.

2 Literature review

To fully understand and appreciate the complex context behind Data Contracts and Data Sharing agreements, a thorough look back to previous phases of Data is necessary. Throughout the human existence, we have been eager to understand and collect the things we know, essentially collecting data all the way from 21 000 years ago (Huylebrouck, 2019). For this thesis, the relevant historical breakthroughs are much more recent. In their bibliographical study into the history and evolution of Big Data and Analytics, Batistič and van der Laken (2019) found that the first foundational Big Data articles have been written in the 1980s' and 90's. According to Kitchin (2014), even the term 'Big Data', though first coined by John Masey in the mid-1990s, didn't begin to take off before the 2010s. Given the very recent emergence of this concept, the current explosion of interest into all things data is even more impressive.

This literature review synthesizes existing research and industry thinking on data governance, organizational data practices, and the emerging role of data contracts within enterprise environments. The review aims to build a conceptual foundation for understanding data contracts not only as technical tools but also as mechanisms that support trust, accountability, and change in complex data ecosystems.

2.1 Foundations of Data Governance

In their article Weill and Ross (2004) define IT Governance as '*the decision rights and accountability framework for encouraging desirable behaviors in the use of IT*' (pp. 4). This definition serves well in Data Governance (DG), too. Many researchers of data governance have leaned on this definition, including Wende and Otto (2007) and Khatri and Brown (2010). In their paper, Weill and Ross (2004i) suggest measuring governance performance by the effectiveness of the four following IT categories: costs, asset utilization, growth, and business flexibility.

These four criteria work well in the context of data as well: data related costs and data solutions' effectiveness are often mentioned as major drivers behind data development (Grande et al., 2020; Ahlawat et al., 2023). Additionally, expanding data capabilities are skyrocketing data-driven innovations in businesses (Zolnowski et al. 2016). Many scholars and industry experts consider the recognition of data as a strategic asset to be the primary driver for implementing DG initiatives (Alhassan et al., 2016; Khatri and Brown, 2010; Otto, 2011).

Data Governance has been defined in numerous ways by practitioners of both academia and industry. The Data Management Association (DAMA) defines data governance as "the exercise of authority, control and shared decision making (planning, monitoring and enforcement) over the management of data assets" (DAMA, 2017, pp. 69). According to Khatri and Brown (2010), governance refers to the decisions made to ensure effective management, as well as defining who is responsible for those decisions.

Central to these definitions are several core principles. In their review of data governance literature, Brous et al. (2016) identified four overarching themes: Organization, Alignment, Compliance, and Common Understanding. These principles aim to ensure that data governance structures support organizational goals, regulatory compliance, and shared understanding across diverse stakeholders. The Data Management Association (DAMA), on the other hand, articulates governance principles across three main domains: People (Roles & Responsibilities, Organization & Culture), Processes (Activities and Techniques), and Technology (Tools and Deliverables) (DAMA, 2017). While Brous et al. (2016) focus on high-level organizational alignment, the DAMA framework is more pragmatic, offering guidance on operationalizing governance in real-world contexts.

An increasingly influential addition to the data governance discourse is the FAIR principles of data being Findable, Accessible, Interoperable, and Reusable, proposed by Wilkinson et al. (2016). Originally developed in the context of scientific research data management, these principles have gained popularity in broader governance contexts due to their emphasis on both human and machine-actionable data quality. The FAIR principles provide guidance for improving governance models by offering a lens through which data assets can be evaluated for usability and sustainability (Jacobsen et al., 2020). In order to be findable, data must have rich metadata and be indexed in searchable resources. To be accessible, data must be retrievable via standardized protocols and access conditions. Interoperable data must use standardized formats, vocabularies, and ontologies to enable integration with other datasets. Finally, to be reusable, data must be well-documented, licensed, and conform to domain-relevant standards (Wilkinson et al., 2016).

These principles align closely with the goals outlined by Brous et al. (2016) and DAMA (2017). For instance, ensuring that data is reusable and interoperable supports organizational alignment and common understanding. Similarly, making data findable and accessible enhances data stewardship and accountability, aligning with DAMA's focus on people and processes. In this way, the FAIR framework can be seen not as a replacement for traditional data governance approaches, but rather as an enhancement, particularly in distributed, data-intensive, and collaborative environments (Jacobsen et al., 2020). As organizations aim to build modern data ecosystems, incorporating FAIR principles into data governance strategies and data collection at the very beginning ensures data is not only well-managed but also truly usable across teams and over time (Endapally, 2023).

These principles, whether from Brous et al. (2016), DAMA (2017), or the FAIR framework (Wilkinson et al., 2016), underscore shared themes such as clarity of roles, alignment with business needs, compliance with legal and ethical norms, and a shared culture around data. Each of these themes will be explored further in this literature review.

The importance of strong governance structures becomes even more pronounced in distributed and decentralized data environments, such as those following a data mesh architecture (Dehghani, 2022). In such contexts, governance ensures that despite the decentralization of data ownership across domains, organizations can still maintain consistent quality, compliance, and discoverability standards. Data Mesh principles will be further elaborated on in the chapter 2.2.3.

2.2 Foundations of Data Security

Data Security is the practice of keeping data protected, managed, and used in a way that does not subject the data to unauthorized access or leave it vulnerable to attacks or theft. One of the leading Data Security frameworks, which also lends well to the topics of this thesis, is the Confidentiality, Integrity, and Availability or CIA triad framework (Samonas and Coss, 2014).

Confidentiality ensures that information is accessible only to those who have the necessary permissions. This principle focuses on protecting sensitive data from unauthorized access or disclosure, often through access controls and authentication mechanisms (Chandramouli and Suchithra, 2018). By maintaining confidentiality, organizations can safeguard privacy and mitigate the risk of data breaches. Closely related, data integrity involves maintaining the accuracy, consistency, and trustworthiness of data throughout its lifecycle (Goswami et al., 2024). The principle of integrity aims to ensure that information is not altered or corrupted, either accidentally or maliciously (Irwin, 2018). Techniques such as authorization, access control, encryption, hashing and version controls are commonly used to support data confidentiality as well as integrity, allowing data users to rely on the data (Paulus and Fauzan, 2018; Duggineni, 2023; Hosseini et al., 2023).

Availability ensures that authorized users can access information and systems when needed (Howard et al, 2023). This principle focuses on minimizing disruptions caused by hardware failures, cyberattacks, or natural disasters. High availability solutions, such as backup systems, failover mechanisms, and robust network architectures, are critical to maintaining uninterrupted access to essential data and services (Ekiye et al., 2023).

While the three cornerstones of the CIA have for decades been presented in literature separately and not under this specific term (Saltzer and Schroeder, 1975), the triad gained popularity in the 1990's (Samonas and Coss, 2014). The CIA Triad has since become a widely recognized standard for understanding and implementing Data Security practices. However, it has also faced criticism for being overly simplistic in addressing the complexities of modern information systems as well as the soft issues of information security (Lundgren and Möller, 2017).

Donn B. Parker introduced an expanded framework in his book, Fighting Computer Crime (1998). Parker's 'Security Hexad' or 'Parkerian Hexad' adds three additional dimensions to the CIA Triad: Authenticity, Possession or Control, and Utility. This broadened perspective highlights the complex nature of data protection and emphasizes the need for a more holistic approach.

When it comes to data security legislation, different jurisdictions have vastly differing regulations. EU, for example, leads the way with the General Data Protection Regulation (GDPR). This data protection legislation focuses on Lawfulness, Fairness, and Transparency, Purpose limitation, Data minimization, Accuracy, Storage limitation, Integrity and Confidentiality, and Accountability. The GDPR also dictates that individuals have rights to access, rectify, object to processing of their individual data and to also be forgotten. (European Union, 2016). Across the world many countries have followed the GDPR's suit in setting high standards for data legislation (Chauhan et al., 2022). Awareness of global data protection law is crucial for companies that have functions in many regions and are therefore impacted by multiple legislative guidelines, like UPM.

Brazil's General Data Protection Law (LGPD), for instance, follows the GDPRs example in many ways, including a wide range of data processing activities, data subjects, and their information, similar mechanisms for transferring data to international organizations and third-party countries. Under the LGPD, there are stronger protections in certain respects: e.g. specific consent is required for processing children's personal data. LGPD also includes explicit treatment of anonymized data and provides rights such as anonymization, blocking, deletion. In contrast, GDPR generally emphasizes correction, erasure, and pseudonymization rather than mandating anonymization in all cases. (ANPD, 2018; European Union, 2016)

Meanwhile the US, for example, does not have one comprehensive regulation on data protection, but rather specific areas or demographics are covered by individual laws: Health Insurance Portability and Accountability Act HIPAA protects healthcare information; Family Educational Rights and Privacy Act FERPA details accesses to student education records; California Consumer Privacy Act CCPA and its amendment California Privacy Rights Act CPRA grant California residents rights similar to GDPR, including data access and deletion (Klosowski, 2021).

Some countries have so-called adequacy agreements with the EU, which state that the data protection laws in the partnering countries are considered to be on a suitable level (European Commission, 2024). While originally enacted in 2011, the South Korean Personal Information Protection Act (PIPA) was revised in 2020 and subsequently in 2023, and now closely resembles the GDPR (PIPC, 2023).

Data Security is a major consideration even in internal contexts. Effective data sharing requires robust access controls, data classification protocols, and secure integration pathways (Shaikh and Sasikumar, 2015; Zubair et al., 2024). Alhassan et al. (2016) argue for explicit governance mechanisms that ensure both security and usability, emphasizing the need for clear accountability frameworks when implementing internal data-sharing initiatives.

Vertically integrated corporations face unique data governance challenges due to the complex legal and competitive environments in which they operate. According to Yan (2023), organizations must implement information barriers, often called "Chinese walls," to prevent the flow of sensitive or competitively significant information between different units. These barriers are particularly crucial in structures where internal conflict-of-interest relationships exist. Yan (2023) highlights that failure to maintain proper data segregation can lead to serious legal consequences, including antitrust violations and regulatory penalties. The concept of "need-to-know" access forms a possible solution to this visibility dilemma.

Makhecha and Harker (2024) emphasize that even within a single corporate entity, information access should be strictly limited based on functional responsibilities and legitimate business purposes. This principle serves multiple objectives: it prevents conflicts of interest, preserves competitive integrity, and ensures compliance with various regulatory frameworks. In vertically integrated companies, these technical controls must be complemented by clear policies, employee training, and regular compliance auditing to maintain the integrity of information barriers.

2.3 Organizational Data Practices

Organizational data practices sit at a crossroads of human, technological and societal needs and constraints. Solid data practices are crucial for data-driven decision-making, innovation, and maintaining compliance, as well as addressing challenges in data quality, security, and governance. In this section, the discussion is focused on four key aspects of organizational data practices: internal data sharing, democratization of data, the development of data as products, and modern data architecture approaches.

2.3.1 Data Sharing within Organizations

Data sharing, in essence, means granting others access to a company's or an individual's data, generally in return for some compensation, be it reciprocal data or monetary (Scaria et al., 2018). In addition to holding technical value in how this data transfer is enacted, this process also holds business value (Fassnacht et al., 2023; Eastwood, 2021). While the feared downsides of losing business-critical advantages of information are evident, companies are waking up to the upsides of data flowing more freely between companies. Data Sharing enables new force in collaboration (van den Broek and van Veenstra, 2015; Thuermer et al., 2019), developing products and processes and creating competitive advantage from effective and trusted data sharing networks (Thuermer et al., 2019). This chapter looks into key organizational information sharing topics: centralised versus decentralized architecture, data silos, data quality, and finally the risks and barriers as well as incentives and motivations for internal data sharing.

A key enabler of sustained internal data sharing is effective governance. Nielsen et al. (2019) propose a model that combines structural, procedural, and relational practices. Structural elements like formal roles and committees provide a foundation, while standardized procedures manage access and usage. Relational practices, which are more organizational and non-technical in nature, such as communication, collaboration, and mutual trust are essential to balance control with flexibility.

While much of the literature on data sharing focuses on inter-organizational contexts, the same principles apply internally, with slightly different constraints and opportunities. Internal data sharing is shaped by architectural decisions, organizational incentives, and governance frameworks (Weber et al., 2009). One of the central architectural debates revolves around centralized versus decentralized data systems. Decentralized systems, particularly those using distributed ledger technologies (DLT), offer enhanced transparency and security by removing the need for a central authority (Antal et al., 2021; Savadatti et al., 2025).

However, these benefits often come at the cost of scalability and efficiency, as these consensus algorithms slow down processing and consume more resources (Bhardwaj and Datta, 2020). Centralized databases, on the other hand, provide operational efficiency and streamlined management, but at the expense of resilience and transparency, introducing risks like single points of failure and reduced trust (Bhardwaj and Datta, 2020).

A comparative study by Naganathan (2018) illustrates these trade-offs clearly. Centralized architectures tend to offer better governance and consistency but limit departmental flexibility and responsiveness. The centralized approach often creates bottlenecks in data access, as all requests must be routed through a core authority. In contrast, Ravva (2025) argues that while decentralized models promote agility and responsiveness, they may suffer from data inconsistency and governance fragmentation.

To resolve this tension, hybrid models like Dehghani's (2022) Data Mesh architecture propose a middle path. In this model, data is treated as a product (more on Data Products in chapter 2.2.3) managed by domain experts while adhering to standardized, organization-wide governance protocols. Dehghani emphasizes four key principles—domain ownership, self-service infrastructure, federated computational governance, and interoperability through standardization—as the foundation for scalable and sustainable data sharing in large, complex enterprises. This approach helps organizations avoid the inefficiencies of centralization while mitigating the possible stress of uncoordinated decentralization (Futurice, 2021).

Despite the advancements in data storage systems, data silos remain a barrier to internal data sharing. Silos, isolated data repositories inaccessible to other systems or departments, arise from a combination of historical, organizational, and technical factors. According to Cisco's 2023 Artificial Intelligence (AI) Readiness Index, which surveyed 8,000 business leaders responsible for AI in large organizations, 81% of respondents acknowledged that data exists in silos across their organizations, prompting Cisco to identify this as the most pressing issue currently facing this clientele.

Origins of these silos can be attributed to a lack of forward planning when a company is facing growth lack of shared understanding between data providers and consumers (Sadiq and Indulska, 2017), non-integrated technologies or narrow team-centric focus that doesn't account for collaborative benefits (Skinner, 2022). According to Günther et al. (2017), many of these silo-inducing factors reinforce one another, creating strong barriers to enterprise-wide data integration.

Breaking down silos requires both technical and organizational change. Routes to overcoming these issues can include centralized data systems, metadata management, standardizing data formats and tools, implementing integration architecture, and improved training and communications can promote more integrated data practices (Kumar, 2023). Skinner (2022) echoes many of these points, by stating that fixing silos has to happen at both the personnel and technological levels. These scholars warn against looking for solely technical solutions, as organizational culture and human-to-human dynamics must also be addressed.

In the context of data sharing within a company, the quality and visibility of data are also very important. Wang and Strong's (1996) widely cited framework identifies intrinsic, contextual, representational, and accessibility dimensions of data quality—all of which influence whether shared data is trusted and used. Their research underscores the importance of delivering data that is accurate, timely, understandable, and secure. Without meeting these standards, internal data-sharing efforts are likely to fail.

Building on this, Sadiq and Indulska (2017) stress that quality assurance must be embedded into the whole life cycle of data processes, not treated as an afterthought. They highlight the importance of standardized quality dimensions and metrics as well as a shared understanding the context of the data at hand as vital to maintaining trust and usability in shared data. These goals ensure that the required level of quality persists, and its value-adding potential can be realized as as data travels across departments and systems.

Khatri and Brown (2010) offer a foundational framework for understanding broad risks and barriers to internal data sharing, beginning with regulatory compliance. Internal sharing must account for privacy laws, data retention rules, and audit requirements. Building on this foundation, Abraham et al. (2019) provide updated insights on how these challenges have evolved in the era of GDPR and other modern privacy regulations.

Organizations must also navigate more political challenges rising within their individual contexts, including interdepartmental power dynamics, unclear resource allocation, mismatched performance metrics, and general risk aversion. Both Khatri and Brown (2010) and Abraham et al. (2019) emphasize that governance must strike a balance between the value of open information flows and the need to manage risk effectively, with Abraham et al. highlighting how this balance has become increasingly complex in today's data-intensive business environment.

Motivation and incentives also play a critical role. Günther et al. (2017) identify value perception, reciprocity, recognition, and reduced friction as major motivational factors. According to Günther et al., when employees and departments see tangible benefits in data sharing, participation increases. Ransbotham et al. (2019) found that organizations excelling in AI adoption also exhibited strong internal data-sharing cultures. These organizations typically benefit from executive support, clear communication of value, structured recognition, and incentive systems that reward collaboration. Dedicated socio-technical systems, models of good practice, training programs, incentive mechanisms and support services are essential for enabling effective sharing (Fassnacht et al., 2023; Uren and Edwards, 2023; Woods and Pinfield, 2022). Without these, even the best intentions may fall short.

Looking forward, the future of internal data sharing is being shaped by several converging trends. Organizations are increasingly embracing domain-oriented architectures like Data Mesh (Dehghani, 2022), embedding governance directly into their platforms, emphasizing quality as a foundation for sharing (Wang and Strong, 1996; Sadiq and Indulska, 2017), and aligning incentives with collaborative behaviors (Günther et al., 2017; Ransbotham et al., 2019). A framework presented by Grossman and Siegel (2014) embodies this well: their CSPG framework stands for Culture, Staffing, Processes, and

Governance, all of which they deem are crucial in designing an organization apt for data capabilities.

The challenge remains to continuously balance competing demands: control and flexibility, standardization and specialization, security and accessibility. Organizations that succeed in navigating these tensions will be best positioned to realize value from their data assets while mitigating associated risks.

2.3.2 Data Democratization

Data Democratization (DD), the practice of making data usable and accessible to a broader audience within an organization, signifies a change in how data management and decision-making are starting to be seen (Samarasinghe and Lokuge, 2023). Originally imagined in fields such as medicine or urban planning, data democratization seeks to empower non-data native situations to be looked at from a data-driven point of view and therefore enhance results, be it in patient care or community planning (Lefebvre et al., 2021). Its core ambition is to dismantle any technical and organizational barriers that have confined data access to secluded data teams, empowering a freer flow of knowledge between technical staff – who understand data – and the non-technical staff – who understand the context – to co-utilize data in decision-making and innovation (Samarasinghe and Lokuge, 2023; Achanta, 2023).

In their study, Zeng and Glaister (2017) found that while the data scientists were the ones responsible for orchestrating data usage across the company, assuming that they alone would be sufficient in creating value might not be the case. Instead, they propose that the likelihood of increasing value creation with data grows along with the breadth to which data is being disseminated within the organization.

Data democratization can be defined as the process by which all individuals within an organization, regardless of their technical background, gain the ability to access, understand, and utilize data with sufficient independence, while keeping within reasonable limitations (Zeng and Glaister, 2017; Hyun et al., 2020; Bellin et al., 2010). This approach aims to separate itself from situations where data is siloed within teams and departments and to relieve the often-overworked data specialists from the most basic tasks (Lefebvre et al., 2021). Data Democratization that has been well embedded into a company's culture is characterized by diverse opinions, knowledge and experiences being shared openly and interactively (Hyun et al. 2020).

The overarching goals of democratization are to tear down barriers around data (Shubladze, 2023), accelerate the pace of decision-making (Huyn et al., 2019), cultivate data-driven culture spanning from grassroots to top management (Díaz et al., 2018) and expanding analytics skills beyond data scientists, especially to managerial positions (Carillo, 2017).

This for-the-people, by-the-people initiative is closely linked to the emergence of self-service analytics, which provides the technological solutions for democratization. Self-service analytics allows users to analyze, report and visualize data independently. These non-technical business users are often dubbed 'citizen data scientists' or 'citizen developers' (Lebens et al., 2021; Gade, 2021). Self-service analysis allows for broader participation in data processes, which, as discussed earlier, is one of the core benefits of data democratization proposed by Bellin et al. (2010). Data democratization and self-service tools together have transformed how data is used within organizations (Gade, 2021).

The technical infrastructure supporting data democratization has advanced significantly, driven by the rise of sophisticated business intelligence (BI) platforms and low-code development environments (Achanta, 2023). BI platforms such as Tableau, Microsoft's Power BI, and Qlik have democratized access to data through user-friendly features, including intuitive visual interfaces (Sallam, 2019), natural language query capabilities that interpret user questions in plain language (Gao et al., 2015), and automated insights that surface patterns, anomalies and clusters with minimal user intervention (Sallam, 2019).

Low-code and no-code platforms take data democratization beyond analytics and into application or data development. These tools enable business users to build custom applications or data solutions tailored to specific tasks and workflows without traditional programming (Rokis and Kirikova, 2022). They streamline operations through visual, drag-and-drop interfaces, while still allowing for the same end results as traditional coding. Cloud-based services have strengthened these capabilities by removing infrastructure constraints and offering scalable access to computational resources (Columbus, 2018).

Despite the advantages of democratization, it introduces complex governance challenges. The primary tension lies in finding the right equilibrium between open access to data and the enforcement of security, privacy, and quality controls (Khatri and Brown, 2010; Shubladze, 2023). Without sufficient governance, data democratization can generate significant organizational risks. Misinterpretation of data by untrained or uninformed users can lead to inaccurate conclusions (Webber, 2020). Inconsistent or scattered datasets can undermine data quality (Haneem et al., 2017). Redman (2013a, 2013b) argues that data issues such as poor quality, lack of trust or overwhelming data management requirements can push managers towards intuitive rather than data-driven decision making.

Additionally, mishandling of sensitive information may breach regulatory frameworks such as GDPR (Chapter 2.2). Data Democratization may inadvertently waste resources and produce a lack of trust in data-derived decisions (Cutler, 2024). It is not enough for users to have access to data, but they must also possess the capability to use data wisely.

Data democratization is a transformative idea with the potential to significantly enhance how organizations leverage data for competitive advantage, but its success depends on more than just technological solutions (Foote, 2023). The literature shows that cultural change, trust, governance, communication and training are equally essential. The most effective initiatives combine broad access with responsible controls, creating data democracy that empowers rather than confuses (Cutler, 2024; Foote, 2023; Achanta, 2023). As the scale and complexity of data environments continue to grow, especially with AI systems gaining more and more traction, organizations must continually adapt their governance practices and no longer view democratic data as a luxury but as a necessity (Achanta, 2023).

2.3.3 Data as a Product and Data Products

Mentions of data or information as products are not new in the literature, spanning even decades into the past (Wang et al., 1998; Pierce, 2005). However, this phenomenon has gained wider traction from the 2010s onwards, as the data science world has gotten sufficiently established (Meierhofer et al., 2019; Davenport and Kudyba, 2016). This chapter explores the fundamental principles behind Data as a Product (DaaP), the resulting Data Products (DPs), the rise of specialized Data Product management roles, required operational frameworks and how this concept relates to the broader data topics, such as the data mesh architecture.

Drawing on the paper by Huang et al. (2015), one of the first that defines the concept of "Data as a Product" names the challenges posed by the ever-growing volume of Big Data as the main starting point for why the DaaP-shift was necessary. To alleviate these growing pains, the DaaP approach aims to separate raw data from data that's ready to be consumed. In essence, the DaaP concept is an approach to manage vast amounts of data by creating smaller, refined Data Products that facilitate easier access and understanding for users. This ease-of-access thinking ties well with the data democratization principles of chapter 2.3.2.

Data Products can be understood as the outputs of a DaaP-oriented data development processes. While the literature on the topic has not yet been condensed into a single, universally accepted definition, Data Products are commonly seen and described as being engineered to deliver actionable insight and benefit to another entity, often named as customers (Dehghani, 2022; Meierhofer et al., 2019; Mucci and Stryker, 2024). Notably, the customers can be internal as well as external, and human as well as non-human (Dehghani 2022; Hasan and Legner 2023).

Both DaaP and DPs include a clear focus on consumer-centricity, which means designing Data Products with a deep understanding of the consumers' needs, challenges, and objectives (Hasan and Legner, 2023). The mindset emphasizes value orientation, which links data products directly to business value (Meierhofer et al., 2019), and incorporates lifecycle management, treating DPs as evolving assets with defined lifecycles (Hasan and Legner, 2023; Mucci and Stryker, 2024).

Loukides (2011) takes the Data Product definitions a step further, by categorizing DPs into 'overt' and 'covert' Data Products based on their final result: is the data *providing* the end result (covert), or *is* the data the end result (overt)? Examples of these could include a social media feed offering data on your followers' activities, which is overtly a Data Product, and a hotel room reserved to you, building on data in a covert way.

The emergence of the DaaP mindset has catalyzed the development of specialized product management roles focused on data assets. According to Deshpande (2023), a Data Product Manager (DPM) serves as the critical bridge between technical and business teams. A DPM is responsible for defining and communicating the strategy and vision for data products. They gather and prioritize requirements from stakeholders to ensure the product meets business needs. Managing the product roadmap, coordinating across team and function limits, overseeing the end-to-end lifecycle, staying on top of relevant quality, privacy and security topics and promoting adoption and literacy around the product are all key aspects of any product managers, and therefore of the DPM as well (Gupte, 2023). The skills required include a blend of technical and domain knowledge, problem-solving, communication skills, and analytical thinking (Deshpande, 2023).

User documentation is a critical enabler of data product usability and therefore a critical task of the DPM. Effective documentation provides an overview and purpose of the data product, describes its schema, outlines data lineage, offers access instructions, presents sample queries and use cases, explains known limitations, and includes a record of changes.

A company's Data Product can be accessed via a Data Catalog (DCat). Data Catalogs, simply defined by Olensen-Bagneux (2023) as *organized inventory of the data in your company* (pp.3), offer an overview of available data products, enabling efficient data discovery, verification and upkeep while lowering the time needed for searching for data (Jahnke and Otto, 2023; Oliveira et al., 2024). Requesting and granting accesses can be done efficiently though a Data Catalog as well, since initially only metadata such as column names are shown to the DCat user, while no actual data values are exposed without sufficient accesses (Olensen-Bagneux, 2023).

Organizations face several challenges when adopting the DaaP approach. According to Loukides (2011), these include cultural resistance to change, skill gaps in the workforce, difficulty justifying return on investment, governance complexity, and the challenge of integrating with legacy systems.

2.3.4 Modern Data Architecture approaches

Architectural choices and visions play a key role in bringing DP and DaaP thinking to life. The choice of architectural framework to follow is not always straightforward. A recurring theme in the evolution of data architecture is the tension between monolithic and microservice designs. According to Blinowski et al. (2022), monolithic solutions are simplistic, fast and reliable to operate, as they require only a single process to run. However, monolithic approaches are not very flexible when it comes to changes in size and complexity. The bigger the monolith becomes, the harder it is to upkeep, deploy, and develop, as any single change requires the whole monolith to be rebuilt

(De Lauretis, 2019). To overcome the issues of monolithic approaches, different decentralized Data Architecture solutions have been introduced.

Microservices, spearheaded by Lewis and Fowler (2014), aim to break the monolith apart to small, independent services, that in communication support the same processes as a monolith would, only in smaller units allowing for more flexibility, scalability, and differentiation. As with any solution, the microservices are not without their critique. For example, Moses (2023) attributes Accountability, Transparency, Privacy as well as Bias and Fairness as some of the challenges these scattered solutions face: who is responsible for the services, their observability, data security or any systemic biases the services may hold?

Borrowing from software engineering, hexagonal architecture, also known as "ports and adapters", provides another useful architecture possibility. In this model, first introduced by Cockburn (2005), the core data is safeguarded from external systems, while standardized interfaces, or ports, and adapters facilitate interaction with diverse consumers and producers. This architecture promotes testability, resilience, and flexibility, aiming that changes in upstream sources or downstream applications cannot disrupt the data at the core. This safeguarding of core data is closely tied to the idea of Data Contracts, presented in the following chapter 2.4.

Lastly, the Data Mesh paradigm represents a synthesis of many of the architectural ideas discussed above. Introduced by Dehghani (2019), it reframes data architecture as a socio-technical system rather than merely a technical one. Its principle of domain-oriented decentralized data ownership, decentralized ownership, federated governance, and self-service infrastructure seek to address the same coordination challenges that DaaP and Data Product thinking respond to (Bode et al., 2023; Araújo Machado et al., 2022). Data Mesh enhances contextual relevance and accountability by aligning Data Products with business domains, while shared governance structures maintain uniformity and interoperability across the enterprise (Blohm et al., 2024).

These modern structures aim to empower business domains to take responsibility for their data products while maintaining overarching standards and consistency through coordinated governance (Dehghani, 2019; Araújo Machado et al., 2022). The self-service infrastructure principle enables business domains to independently manage the lifecycle of their data products by providing them with the necessary technical capabilities (Caserta et al., 2023). Finally, interoperability through standardization is essential to avoid creating or strengthening data silos and bottlenecks. In a case presented by Loukiala et al. (2021), the adoption of a distributed data platform architecture was found to eliminate many of the bottlenecks associated with traditional, centralized data warehouses and data lakes.

2.4 Data Contracts and Data Sharing Agreements

As discussed in the chapter 2.3.1 on Organizational Data Practices, trust plays a crucial role in all data sharing. Without robust methods to hold the data sharing parties accountable, sharing information and knowledge could well prove detrimental to many companies. For example, in their research into data sharing willingness, Dahlberg and Nokkala (2019) found that their study groups were unfamiliar with technological trust providers and that they would require significant actions and behavioral changes to have these adopted.

Data Sharing Agreements (DSA) and Data Contracts (DC) have been offered as a solution to the question of how data sharing can be done safely and in a way that enforces trust between collaborating parties. Since Data Contracts and Data Sharing Agreements are relatively young concepts and not thoroughly researched academically, much of the knowledge in this field comes from private entities: companies providing services in, or individuals working on these topics.

Data Contracts are a technology-first approach. Data Contracts are implemented in platform-agnostic YAML format, and they define features such as structures, semantics, quality, and terms of use between data providers and their consumers (Sassoon, 2023). In effect, Data Contracts can transform any agreements made between the data provider and receiver into a sort of gate, through which the data does not flow if it does not meet the preset criteria.

Data Sharing Agreements, on the other hand, can be much more versatile in how they are constructed, as they are less technical in nature. DSAs define how data is shared and handled, whereas DCs define qualities of the data itself. Data Sharing Agreements can dictate, for example, who is allowed to work with the shared data, and through which means the data is shared (Petersen, 2019).

While Data Contracts provide a set structure for the outports of data products, they allow for different methods of getting to that final result. This means that data products can be gathered and perfected in differing ways under different organizations, while still enabling seamless integration (Sassoon, 2023; Khatun, 2024). Data products need to be flexible when it comes to their creation, since all data cannot be used in the same way by the same tools (compare, for instance, IoT data and personnel data), but the end use must be standardized and tool agnostic to facilitate effective cross-referencing (Sassoon, 2023). In a diversified conglomerate such as UPM, this means that every separate area or function can choose whichever data development method best suits their needs as long as their output matches the preset limitations of a Data Contract.

2.4.1 Definitions and Conceptual Origins

The concepts of Data Contracts and Data Sharing Agreements have emerged as mechanisms for governing the increasingly complex practice of data exchange. While they serve related purposes in facilitating trust and accountability, they arise from different practices and therefore address different aspects of data governance and interoperability.

A Data Sharing Agreement is rooted in the legal and regulatory realms. It is a formalized, legally binding document that defines the terms under which data is shared between entities, typically organizations or departments. DSAs specify who is permitted to access data, under what circumstances, for what purposes, and with which constraints, to allow for controlled and collaborative data exchange (Matteucci et al., 2012). They have become especially important in the context of privacy regulations such as the GDPR and other compliance frameworks.

The primary function of a DSA is to formalize the governance of the behavior of the human actors involved in data exchange – who could, depending on the situation, occupy roles such as a lawyer, a business expert, and an end-user – and to provide legal recourse in cases of misuse, breach, or noncompliance (Costantino et al., 2017).

Data Contracts, by contrast, emerge from the realm of software engineering and data architecture. They are primarily machine-readable specifications that encode expectations about the structure, semantics, and quality of data exchanged between producers and consumers. Rather than governing how data may be legally used, Data Contracts ensure that the data is technically correct, well-structured, and aligned in meaning with the needs of downstream consumers (Dehghani, 2022).

Typically written in formats like YAML or JSON, Data Contracts enforce constraints on schema, field types, validation rules, and data freshness, thereby reducing misunderstandings and integration errors across systems (Khatun, 2024). Although not inherently legal documents, they can support compliance efforts by offering transparency and traceability into the flow and transformation of data.

Conceptually, Data Contracts are an extension of long-standing software engineering practices, such as interface contracts and Application Programming Interfaces (APIs) specifications. Interface contracts, as conceptualized in Meyer's (1992) Design by Contract methodology, establish explicit agreements between software components by defining acceptable input parameters, expected output formats, and necessary conditions that must be met before functions can be properly executed. Similarly, APIs serve as structured communication frameworks between distinct software systems, establishing standardized protocols for how information requests should be formatted, what specific data elements may be requested, and in what format the requested information will be delivered to the receiving system.

However, such traditional contracts are often too limited in today's data landscapes. They assume, rather than ensure, semantic alignment, meaning that they verify that a system can transmit data, but not that the data will be meaningful or valid in its new context. The changing of data, column names, or any number of other properties of the data, known as Schema Evolution or Schema Drift, is common but destructive if not facilitated correctly (How, 2020; Osarenren, 2024).

Data Contracts extend the scope of interface contracts by addressing not only the format of exchanged data but also its meaning, provenance, and quality. As noted by Dong et al. (2019), while API contracts govern the syntax of exchange, Data Contracts govern the semantics and integrity of the information being transmitted. This distinction is critical in modern data environments, where technical compatibility does not guarantee trustworthy or meaningful data.

According to Snodgrass and Soon (2019), revisions and updates done on the API's side can result in major, resource-intensive problems and fixes on the receivers' side, potentially harming multiple downstream dependents. To combat this, Data Contracts introduce a layer of semantic rigor previously missing from most technical specifications.

The historical emergence of Data Contracts as a distinct technical practice can be linked to several developments in the data ecosystem. One such development was the introduction of the Data Mesh paradigm by Dehghani (2019), which called for domain-oriented ownership of data. This approach advocated treating data as a product, complete with well-defined interfaces, service-level objectives, and user documentation.

In such a distributed architecture, where autonomous teams manage their own data domains, informal agreements and tribal knowledge proved insufficient to maintain reliability and trust (Dehghani, 2019; Data Mesh discussed in in Chapter 2.3.1, Data Products in Chapter 2.3.3). Data Contracts emerged as a formal mechanism to define and enforce the expectations between these decentralized producers and consumers.

Another influential development was the creation of open data standards like the Open Contracting Data Standard (OCDS), which represents a significant milestone in the evolution of standardized data contracts for cross-institutional collaboration. According to the OCDS documentation, the standard was developed to address the challenge of data being provided in differing formats by separate publishing entities, hindering cross-publisher analysis (Open Contracting Data Standard, 2024). The standard establishes a comprehensive schema that defines how contracting information should be structured, documented, and shared to enable systematic analysis and transparency across different jurisdictions and organizations.

Implementation experiences have provided practical insights for managing data contracts in complex environments, especially with public-sector governmental challenges. In Colombia, the adoption of OCDS through the SECOP II procurement platform has enabled unprecedented transparency and analysis capabilities (Richter Gade et al., 2020). The World Bank documentation (Richter Gade et al., 2020) reveals that Colombia's implementation of OCDS has had a positive impact on monitoring public procurement processes and the analysis of procurement data for policy-making purposes.

However, in the UK, implementing OCDS on its Contracts Finder platform highlighted a different challenge: how to balance standardization with the practical requirements of pre-existing ways of working (Crown Commercial Service, 2020). These contrasting experiences bring out a fundamental tension in data contract design: there is need for standards rigid enough to enable interoperability while remaining flexible enough to accommodate diverse contexts.

What makes the Colombian implementation particularly compelling is how it transformed analytical capabilities that were previously constrained by data fragmentation. The Open Contracting Partnership documents how Colombian authorities are now better equipped to recognize patterns and trends in procurement, aiding in both sustainable policy decisions and anti-corruption aims (Richter Gade et al., 2020; Open Contracting Partnership, 2024).

The broader significance of OCDS extends beyond procurement transparency to demonstrate principles applicable across data sharing contexts. The approach taken by both Colombian and UK implementers, combining detailed technical specifications with practical implementation guidance, suggests a model for developing data contracts that acknowledge the messy realities of organizational change while maintaining the technical precision necessary for meaningful data exchange. This balance has made OCDS a promising standard to follow for other many other public sector actors too: countries and cities including Moldova, Uganda, Ukraine, Mexico City and Montreal are publishing data based on the OCDS (Alliance for Integrity, 2018)

2.4.2 Legal and Technical Perspectives

The effective management of data exchange in modern organizations requires addressing both legal compliance and technical reliability (Petersen, 2019). Data Sharing Agreements (DSAs) and Data Contracts (DCs) represent complementary approaches to these challenges. From a legal perspective, DSAs serve as formal instruments for managing data exchange under regulatory and contractual obligations. In the context of regulations such as the General Data Protection Regulation (GDPR) in the European Union or the Health Insurance Portability and Accountability Act (HIPAA) in the United States, organizations must provide assurances that shared data will be processed lawfully, stored securely, and accessed by authorized parties only. DSAs fulfill this role by defining permissible uses, data ownership, liability in the event of a breach, and protocols for data subject access requests or audits (Liu et al., 2018). They often include confidentiality clauses, retention schedules, and jurisdictional boundaries to ensure compliance with legal mandates. For example, GDPR Article 28 requires formal agreements when processing personal data on behalf of another organization, making DSAs a legal necessity in data processing relationships across many industries (European Commission, 2016).

However, the legal protections offered by DSAs are insufficient on their own to guarantee the technical fidelity of the data exchanged. This is where Data Contracts play a critical role. While not typically embedded in legal documents, Data Contracts define the enforceable specifications of data quality and structure that systems must uphold. These specifications include schema definitions, expected data types, nullability constraints, and validation rules that are applied at the point of data production or transformation (Dehghani, 2022). In distributed or domain-oriented data architectures, such as those promoted by the data mesh paradigm, Data Contracts are indispensable in aligning expectations between teams that may otherwise operate in isolation (Pal, 2024).

One of the most prominent technical domains where Data Contracts have taken root is in analytics engineering, particularly through the dbt (data build tool) ecosystem. dbt allows teams to specify the expected structure of datasets, including column names, data types, and constraints, within YAML configuration files. These contracts are enforced during model execution, where if the output deviates from the defined contract, the model fails to build, thereby preventing the flow of erroneous data downstream (dbt Labs, 2025a). By embedding these controls within the transformation layer, dbt enables teams to maintain data quality and consistency across complex data workflows, mitigating the risks associated with data inconsistencies and governance fragmentation (dbt Labs, 2025a; dbt Labs, 2025b; Atlan, 2024).

Technical enforcement in dbt is further enhanced through the use of custom tests, which evaluate data against expectations such as uniqueness, referential integrity, and value ranges. These tests are automated, meaning that contract violations, such as missing columns or unexpected null values, can be detected early in the pipeline and addressed before data reaches downstream consumers. Such feedback loops are critical in maintaining trust in data products, particularly in fast-moving environments where data is continuously ingested, transformed, and visualized (Zhang et al., 2022).

In addition to dbt, a growing ecosystem of tools supports the implementation of Data Contracts. Platforms like Monte Carlo, Great Expectations, and Soda.io offer data observability solutions that monitor pipeline health, detect anomalies, and enforce schema consistency at scale (Abraham et al., 2019). These tools enable organizations to operationalize Data Contracts beyond static schema definitions by embedding continuous validation, lineage tracking, and alerting mechanisms into production workflows. By automating these checks, teams can prevent data drift and ensure that changes in source systems do not silently break analytical models or reports.

While Data Contracts do not replace legal agreements, they augment them by translating abstract obligations into executable specifications. For instance, a DSA might require that personal data fields be pseudonymized before sharing, whereas a corresponding Data Contract would define the technical method of pseudonymization and validate its consistent application. In this way, Data Contracts serve as the technical instantiation of legal intentions, creating a bridge between compliance requirements and system-level enforcement. (Segner, 2025; Atlan, 2024)

The convergence of legal and technical governance mechanisms reflects a broader shift in data management culture—from informal coordination and reactive fixes to proactive specification and accountability. As Seiner (2020) argues, effective data governance demands clarity not only in policy but also in execution. DSAs and DCs, when used in tandem, create a robust framework in which organizations can securely and reliably share data across boundaries without sacrificing control or compliance.

The technical architecture of OCDS incorporates principles that have influenced broader data contract development. The standard employs what the official documentation describes as "a data model that describes the structure and semantics of contracting data" (Open Contracting Data Standard, 2024). This approach defines not just data structure but also the meaning and relationships between different data elements, enabling automated validation and quality checking while maintaining flexibility for local adaptations. The standard's approach to versioning ensures that "publishers can adopt new versions of the standard while maintaining backward compatibility with existing data consumers."

2.4.3 Comparing Data Contracts, SLAs and Data Sharing Agreements

As organizations increasingly formalize their data exchange practices, they must navigate a complex ecosystem of governance instruments, including Data Contracts, Data Sharing Agreements, and Service Level Agreements (SLAs). While all three aim to establish clarity and accountability in collaborative data environments, they differ significantly in scope, orientation, and mechanisms of enforcement.

Data Contracts are primarily technical constructs that define the expected structure, semantics, and quality of data exchanged between systems or teams. They are typically implemented as machine-readable specifications, enabling validation and enforcement through automated workflows. As Dehghani (2022) notes, Data Contracts emerged from the need to treat data as a product in distributed architectures, where autonomous teams produce and consume data across organizational boundaries. A typical use case involves a data producer agreeing to deliver a dataset with specified fields, types, and quality metrics, which are validated continuously using tools such as dbt. This approach ensures that consumers can rely on the consistency and fitness-for-use of the data without needing to re-verify it each time (Segner, 2025).

In contrast, Data Sharing Agreements are legal documents that define the terms under which data may be accessed, used, stored, or transferred. They specify rights, responsibilities, and liabilities associated with data exchange, often in compliance with external regulations such as the GDPR, HIPAA, or industry-specific standards. DSAs are typically negotiated between legal or compliance teams and address questions of intellectual property, data privacy, data retention, and dispute resolution (Liu et al., 2018). A common scenario for DSAs arises when organizations share sensitive customer data with third-party vendors or research partners. In such contexts, the agreement may stipulate, for example, that data must be anonymized, stored within specific jurisdictions, or deleted after a fixed period (European Commission, 2016).

Service Level Agreements occupy a third space, focusing on operational performance guarantees between service providers and consumers. In the realm of data, SLAs often define acceptable thresholds for metrics such as data freshness, system availability, query response times, and incident resolution windows (Kostic, 2024). While not specific to data content or legal compliance, SLAs play a vital role in managing expectations and accountability in data services. For example, a cloud data warehouse provider might commit to data refresh interval of 15 minutes for streaming updates. Breaches of such guarantees typically result in penalties, making SLAs a key mechanism in vendor management and operational risk mitigation (Nicolazzo et al., 2024).

Although distinct in their primary functions, these three instruments frequently intersect in practice. A robust data sharing framework may rely on a DSA to define legal boundaries, a Data Contract to enforce technical specifications, and an SLA to govern service-level performance. The overlaps can lead to confusion if not carefully coordinated. For instance, a DSA might specify that personally identifiable information must be pseudonymized, but without a corresponding Data Contract, there may be no guarantee that the pseudonymization logic is technically applied or validated. Likewise, a Data Contract may ensure schema stability, but if freshness guarantees are not explicitly stated in an SLA, consumers may still encounter stale or delayed data. As Luis et al. (2024) argue, effective data governance arises from the alignment of legal, technical, and operational perspectives—not from any single domain or document.

Determining when to use which mechanism depends largely on the context, stakeholders, and risks involved. For internal exchanges between closely coordinated teams, lightweight Data Contracts may suffice, particularly if version control and schema validation are already embedded in the development workflow. However, when external partners are involved, especially in heavily regulated industries, DSAs become essential to ensure compliance and manage liability. SLAs, on the other hand, are most relevant when uptime, latency, or availability directly affect business performance or contractual obligations. As Abraham et al. (2019) observe, the decision to deploy a particular governance instrument should be guided by a risk-based approach that considers the nature of the data, the sensitivity of the use case, and the autonomy of the involved parties.

Ultimately, these governance tools are not mutually exclusive but mutually reinforcing. Their integration reflects the evolving maturity of organizations in handling data not just as a technical asset but as a legally protected and operationally critical resource. As Seiner (2020) emphasizes, true data governance is achieved not through singular control points but through a well-orchestrated system of agreements, contracts, and performance guarantees that together create a trusted and efficient data ecosystem.

2.4.4 Previous Implementations and Case Studies

Although Data Contracts and Data Sharing Agreements are relatively recent innovations in the domain of data governance, emerging literature and industry reports provide a growing body of examples that demonstrate both their practical utility and the challenges of implementation. Academic and applied case studies reveal how different organizational contexts shape the deployment of these mechanisms, particularly in environments requiring cross-functional coordination, regulatory compliance, or data product scalability.

In academic literature, much of the early discussion around formal data sharing emerged in the context of collaboration. In a study focused on supply chain ecosystems, Dahlberg and Nokkala (2019) found that companies were hesitant to engage in data exchange without clear contractual boundaries. Their research shows that behavioral change and organizational culture posed greater barriers than technological limitations, underscoring the importance of formal agreements such as DSAs in fostering reliable cooperation between entities.

From the industry perspective, several companies have pioneered the use of Data Contracts as part of their internal data product strategy. At PayPal, the adoption of a data mesh approach has included the introduction of data contracts as a key mechanism for improving reliability and trust in analytical workflows. As described by the PayPal engineering team, contracts are embedded into development pipelines to enforce schema consistency and validate data quality before changes reach production (Perrin, 2022). This integration has reduced the risk of downstream breakages and helped ensure that analytical teams can work with stable, trustworthy datasets.

At Netflix, internal Data Contract-like solutions govern data lineage and expectations between microservices and pipelines. Their approach treats contracts as living documentation that both engineers and analysts can review and validate, supported by observability tools and metadata platforms (Lin et al., 2019). Another frequently cited case is the use of dbt at JetBlue (JetBlue, 2020; dbt Labs, 2020). At JetBlue, according to a case study published by dbt Labs, key progress in their data work was enabled by many facets, such as dbt's version controlling, real-time data capabilities, preventing data quality issues from impacting end users, and improving data transparency and documentation. While it is not specifically named in these case study, all these improvements are central to the core of DCs and DSAs, which provide exactly these benefits. While the term Data Contract might not have been used yet in by Netflix in 2019 or JetBlue in 2020, it is clear in these cases the solutions themselves are part of the same concept.

Despite these successes, various critiques and pitfalls have been observed in both research and practice. A recurring theme is that Data Contracts are only as effective as the enforcement and monitoring mechanisms behind them. Seiner (2020) cautions that overly rigid or top-down governance models often threaten adoption and can lead to governance becoming a checkbox exercise rather than a lived, continuously enforced practice. Seiner argues that embedding 'non-invasive' governance activities into existing workflows, focusing on people and culture, and avoiding treating agreements as static artifacts lend the best results. Another common critique is that without a shared understanding of the business meaning behind data, technically valid datasets may still be functionally inappropriate for their intended use (Fu et al., 2024).

On the DSA side, overly rigid legal agreements can slow down or completely inhibit data sharing by introducing delays and bottlenecks in processes. Khatri and Brown (2010) found that in some organizations, the legal review process for DSAs introduced multi-week lead times, which conflicted with the needs of agile data science teams seeking quick access to third-party data. This raises the important critique that while formal agreements are necessary to ensure compliance and accountability, they must be balanced with operational flexibility to support innovation and responsiveness.

Moreover, as Abraham et al. (2019) point out, the implementation of any data governance mechanism—be it contractual or policy-driven—requires cultural adaptation. Organizations that successfully implement Data Contracts or DSAs tend to foster a culture where data stewardship is shared and embedded across roles. In contrast, organizations that treat data governance as a top-down compliance function often face resistance and limited adoption.

The practical deployment of Data Contracts and Data Sharing Agreements has yielded valuable lessons across both academic and industry contexts. Success factors include early stakeholder engagement, automation of enforcement, integration into development workflows, and ongoing monitoring. Conversely, common pitfalls involve lack of clarity in contract definitions, misalignment between legal and technical teams, and the risk of bureaucratic overhead slowing down innovation. As these mechanisms continue to evolve, a balanced approach that bridges legal safeguards with operational agility appears to be most effective for scalable and trustworthy data sharing.

Taken together, these technological, organizational, and conceptual shifts led to the formalization of Data Contracts and Data Sharing Agreements as a distinct component of modern data architecture (Dehghani, 2022; Segner, 2025; Petersen, 2019). They represent a meeting point of legal and technical enforcement, offering both human-readable commitments and machine-verifiable guarantees.

2.5 Change and Trust in Data Organizations

Change management plays a crucial role in facilitating transitions within data organizations, particularly in the context of adopting data-driven initiatives, improving data governance, and implementing new technologies. Existing literature on this topic identifies several key themes and strategies essential for successful change management in these contexts. Enhancing data-driven decision-making is one of the primary goals of change management, especially during the deployment of Business Intelligence (BI) systems. By addressing challenges associated with data collection, interpretation, and presentation, effective change management is found to improve organizational alignment and decision-making processes (Mahamoud Hasan et al., 2025).

Combining theoretical change management frameworks with practical applications is another significant area of focus. Research underscores the importance of bridging the gap between theoretical models and real-world practice through clear communication, stakeholder engagement, and phased implementation, ensuring the success of change initiatives within data organizations (de Andrade et al., 2016; Buffenoir and Bourdon, 2012). Additionally, the shift from centralized data governance to distributed models (e.g. data meshes from chapter 2.2.2), introduces unique challenges and opportunities. Change management facilitates this transition by promoting adaptability, collaboration, and continuous learning. Agile methodologies have proven particularly effective in fostering resilience and flexibility in these scenarios (Vestues et al., 2022).

Preserving data quality and ensuring system reliability are critical aspects of change management in IT and data environments. Structured change control practices minimize risks and disruptions during transitions, enabling organizations to maintain operational stability while implementing new systems or processes (Mahamoud Hasan et al., 2025). Additionally, in complex and extended organizations, change management is instrumental in integrating data governance practices. By addressing organizational complexities and power dynamics, change management enhances data reliability and organizational control, aligning data practices with broader strategic goals (Buffenoir and Bourdon, 2012).

Best practices for change management in data organizations include comprehensive planning, proactive risk assessment, and fostering a culture of continuous improvement. Frameworks tailored to the specific needs of data governance and sharing help ensure that organizational goals align with technological advancements (Vestues et al., 2022). Change management thus emerges as an indispensable component of data organizations, addressing the challenges of adopting innovative technologies, fostering collaboration, and ensuring the successful implementation of data-driven strategies. These insights provide a theoretical and practical foundation for understanding the interplay between organizational change and data governance.

2.5.1 Organizational Resistance and Change Management

The journey toward data-driven transformation is fraught with challenges that extend beyond technological implementation to encompass the human elements of organizational change. Despite significant investments in data initiatives, many organizations struggle to realize the expected benefits due to resistance deeply rooted in organizational dynamics (Shamaei, 2023; McKinsey, 2022). Change fatigue represents a significant barrier to successful data transformations, particularly in environments where multiple initiatives compete for attention and resources. According to Stouten et al. (2018), organizations frequently underestimate the cumulative impact of concurrent change efforts, leading to employee disengagement and diminished capacity for adopting data-driven practices. This fatigue manifests as passive resistance, characterized by superficial compliance without genuine adoption of new data-centric behaviors and mindsets.

The lack of strategic alignment between data initiatives and organizational objectives constitutes another primary reason for failure. Dethine et al. (2020) observe that data transformations often proceed without clear connections to business value, creating a perception that such initiatives represent technological indulgence rather than strategic necessity. This misalignment engenders skepticism among stakeholders who fail to recognize how data capabilities advance their specific objectives and priorities. The resulting organizational resistance emerges from a perception gap between technical implementation teams and business units regarding the purpose and value of data initiatives.

Successfully navigating these challenges requires structured approaches to change management that address both the technical and human dimensions of data transformation. Kotter's Eight-Step Process for Leading Change provides a comprehensive framework particularly relevant to data initiatives (Kotter, 2012). The model begins with establishing a sense of urgency around data opportunities, followed by building a guiding coalition that spans technical and business perspectives. Creating a strategic vision that clearly articulates how data capabilities will deliver organizational value represents a critical third step, addressing the alignment challenges previously identified. The subsequent phases, communicating the vision, empowering employees, generating short-term wins, consolidating gains, and anchoring new approaches in organizational culture, all collectively transform initial momentum into sustainable change (Kotter, 2012).

The ADKAR model (Awareness, Desire, Knowledge, Ability, Reinforcement) offers a complementary perspective focused on individual transitions through the change process (Hiatt, 2006). This model proves particularly valuable for data initiatives by recognizing that organizational transformation ultimately depends on individual adoption. Beginning with awareness of the need for improved data practices, the model progresses to cultivating desire for participation, developing knowledge of data concepts and tools, building ability through practice and coaching, and reinforcing new behaviors through recognition and measurement. The granular focus of ADKAR enables change leaders to diagnose specific barriers to adoption at the individual level, allowing for targeted interventions rather than generalized change strategies.

The implementation of these frameworks depends critically on effective change leadership distributed throughout the organization. Executive sponsorship emerges as perhaps the most significant predictor of success in data transformations. Leaders serve as visible advocates for data initiatives, connecting data capabilities to strategic priorities and modeling data-driven decision making (Shamsuddin and Abdul Razak, 2023). Without this executive commitment, data initiatives struggle to maintain priority amid competing organizational demands, leading to resource constraints and wavering support.

Complementing executive sponsorship, specialized roles have emerged to facilitate change throughout organizational hierarchies. The data product owner assumes responsibility for ensuring that data assets meet specific business needs, serving as a translator between technical capabilities and functional requirements. This role bridges the technical-business divide through continuous engagement with stakeholders to understand their evolving data needs and priorities. The data champion function operates at the grassroots level, cultivating enthusiasm and demonstrating practical applications of data capabilities within specific business contexts. These individuals typically possess strong domain expertise complemented by data literacy, allowing them to identify high-value use cases that resonate with their colleagues.

Data stewards address the governance aspects of change, establishing and promoting standards for data quality, access, and usage. Their focus on policy development and compliance creates the structural foundation for sustainable data practices. According to Vilminko-Heikkinen and Pekkola (2019), the effectiveness of data stewards depends on their ability to balance governance requirements with practical business needs, avoiding rigid approaches that impede rather than enable data utilization.

The interplay between these various change roles creates a network of support for data initiatives, addressing resistance at multiple organizational levels. Successful organizations achieve what Pappas et al. (2018) describe as ecosystem-level change rather than focusing on any changing actors in isolation. When properly implemented, this distributed change leadership transforms resistance into engagement by demonstrating the tangible value of data capabilities within specific organizational contexts.

2.5.2 Building Trust through Data Contracts and Data Sharing Agreements

According to Forsyth et al. (2011), "*trust is the glue that holds the organization together*" (pp.111). The establishment of trust represents a foundational element for any successful organization, particularly the data organizations. Data contracts and Data Sharing Agreements have emerged as critical mechanisms for cultivating this trust by creating formalized understanding between data producers and consumers.

Data contracts serve as explicit declarations of the properties, qualities, and behaviors that data consumers can expect from data producers. According to Frenté and Ladley (2023), these agreements define critical parameters including data structure, quality standards, update frequency, and access mechanisms. The formalization of these expectations transforms ambiguous data relationships into predictable interactions governed by mutual understanding. This transformation proves particularly valuable in distributed data environments where direct oversight gives way to negotiated agreements between autonomous units.

The trust-building function of data contracts operates through multiple mechanisms. First, these agreements create transparency regarding data lineage and processing, allowing consumers to understand the origins and transformations of the data they utilize (Segner, 2025; Atlan, 2024; dbt Labs, 2025a). Janssen et al. (2020) note that this transparency directly addresses the "black box" concerns that frequently undermine trust in data systems, particularly when analytics drive consequential decisions. Second, contracts establish accountability by clearly delineating responsibilities for data quality and availability (Segner, 2025; Khatun, 2024; Sassoon, 2023). This accountability mitigates the tendency toward blame displacement when data issues arise, creating instead a collaborative framework for problem resolution.

Perhaps most significantly, data contracts reduce the "surprise factor" that frequently erodes trust in data relationships. Unexpected changes to data structures, quality, or availability represent a primary source of friction between data producers and consumers. Data contracts mitigate this friction by establishing notification requirements and change protocols that prevent unilateral modifications with cascading impacts. (Prudhvi, 2025)

Beyond their governance function, data contracts facilitate cultural transformations essential for data-driven organizations (Segner, 2025). The process of negotiating and maintaining these agreements necessitates cross-functional dialogue that breaks down traditional silos (Blohm et al., 2024). Technical teams gain deeper understanding of business requirements, while business units develop appreciation for the complexities of data management (Wider et al., 2023). This ongoing dialogue creates the capacity to understand data challenges and requirements from multiple perspectives. The resulting collaborative mindset represents a significant departure from the territorial approaches that characterized earlier data management practices.

Data sharing agreements extend these principles beyond organizational boundaries to govern external data exchanges. These formalized arrangements address the additional complexities introduced when data flows between legally distinct entities with different priorities and constraints. Research by Allen et al. (2014) identifies several critical dimensions of these agreements: legal frameworks for data rights and responsibilities, technical specifications for secure data exchange, and operational protocols for maintaining data integrity throughout its lifecycle. The comprehensive nature of these agreements establishes the predictability essential for trust in external data relationships.

The transition toward contract-based data governance represents a significant cultural shift for many organizations. Traditional data management approaches emphasized centralized control enforced through technological constraints and policy mandates. The contractual model, by contrast, distributes authority through negotiated agreements that recognize the autonomy of both producers and consumers. This transition from control to collaboration aligns with broader organizational trends toward network structures that emphasize coordination rather than command. This aligns well with what Carmi et al. (2020) describe as "data citizenship", a mindset characterized by shared ownership and collective responsibility for organizational data assets.

Implementing effective data contracts requires balancing formality with flexibility. Overly rigid agreements quickly become obsolete in dynamic environments, while vague commitments fail to create meaningful accountability. Successful organizations develop tiered approaches that establish core principles as non-negotiable standards while allowing context-specific adaptations for individual data relationships. This balanced approach aligns with the perspectives presented by Yaqoob and Thomas (2022), who discusses the challenges and solutions in data governance within the era of Big Data. Yaqoob and Thomas emphasize the necessity for a holistic data governance framework that accommodates the volume, velocity, variety, and veracity of data, without compromising on data quality, security, and compliance. Such a framework requires organizations to establish consistent standards that can adapt to the diverse and dynamic nature of Big Data, ensuring both reliability and responsiveness in data practices.

The trust established through effective data contracts and sharing agreements creates the foundation for organizational data ecosystems characterized by reliable exchange and continuous improvement (Segner, 2025, Sassoon, 2023). When stakeholders develop confidence in data relationships, they increase their willingness to depend on shared data assets for critical decisions and processes. This deepening integration accelerates the organization's overall data maturity by expanding the scope and impact of data utilization (Blohm et al., 2024; Wider et al., 2023). The resulting virtuous cycle transforms data from a contested resource into a collaborative, creative common possibility that advances collective organizational objectives.

3 Methodology

This chapter outlines the methodological choices, including the case context, research design, data collection methods, and analytical strategy employed in this thesis. The chapter is structured as follows: Section 3.1 introduces the case company; Section 3.2 presents the case description; Section 3.3 outlines the research design; Section 3. 4describes data collection; and Section 3.5 details the analytical approach; and finally, section 3.6 discusses ethical considerations.

3.1 Company Background

UPM is a Finnish forest industry company with a long history dating back to the 1870s. The company has undergone significant transformation from a traditional paper manufacturer to a diversified biomaterials-forest industry leader, focusing on renewable alternatives for fossil-based materials. Today, UPM positions itself as a leader in the biomaterials and forest industries, focusing on sustainable solutions and innovation. UPM operates globally, employing approximately 17 000 people in 43 countries, running production facilities in 12 countries, and a sales network extending to most parts of the world. UPM is listed on the Helsinki Stock Exchange, and its headquarters are located in Helsinki, Finland. In 2023, UPM's sales were reported at 10,5 billion euros.

UPM consist of seven business areas, controlling the entire value chain from forests to end products. This vertical integration allows for high resource efficiency, maximizing the value of wood raw materials through various products, such as packaging, construction, and bioplastics. However, each business entity operates independently. Within the scope of this thesis, interviews were targeted at two business areas, UPM Fibres and UPM Energy, as well as the Finance and Controlling (FC) function. The chosen areas and data-specific details pertaining to these areas are presented in the next chapter.

3.2 Case Context

Initial inspiration for this topic stemmed from the work that has been carried out in UPM by an Information Architects' Working Group (IAWG). Within the weekly IAWG meetings, Information Architects from different business areas and functions are invited to participate in discussions on various IA topics, such as project updates, best practice development etc. One of the recurring topics this council came across has to do with the differing data maturity levels and ways of defining relevant roles, terminology, and responsibilities in data organizations.

The IAWG is hosted by the Digital and Data Management Office (DDMO), which operates under the IT function. The mission of DDMO is to assist the entire UPM Group in its goal of being more digitally savvy and data oriented. Many sessions of IAWG are aimed to help business areas to adopt and utilize the Common Data Platform (CDP), which is also part of the responsibilities of DDMO. The CDP is comprised of multiple data sources that are integrated with Snowflake to allow for frictionless data access from many sources to many end users and their goals. While writing this thesis, the researcher was employed by DDMO.

UPM Fibres comprises UPM's pulp and timber operations. The pulp business produces both hardwood and softwood pulp for various end-uses. The timber business produces certified sawn timber for construction, joinery, and furniture industries. UPM Fibres employs around 2 500 people and in 2023 had sales totalling 3,04 billion euros.

Data management in the pulp industry is increasingly shaped by regulatory demands for supply chain traceability and environmental compliance, as well as by digitalization trends in process optimization. A key legislative driver is the EU Deforestation Regulation (EUDR, Regulation (EU) 2023/1115), which requires companies placing forest-risk products, including pulp and paper, on the EU market to demonstrate traceability of raw materials to their source. By the end of 2024, large companies must implement systems that collect and share location-specific data for all wood inputs, with small operators following in 2025 (European Commission, 2023). In response, pulp producers are advancing digital data exchange systems to manage the complexity of tracking wood origin, supplier identity, and batch-level processing. Tools include geolocation systems, satellite monitoring, and Radio Frequency Identification (RFID) tracking that feed into Enterprise Resource Planning (ERP) platforms.

Internally, pulp mills rely on large volumes of industrial process data collected from continuous production systems. This includes sensor data on temperature, chemical usage, energy efficiency, and emissions. These datasets are utilized to support optimization, predictive maintenance, and compliance with environmental regulations. Leading Finnish operators, such as UPM, utilize such data with machinery suppliers and partners in real-time.

What sets the pulp industry apart is the dual need to handle both transactional traceability data for compliance and high-frequency, high-volume sensor data for process control. These are often managed in separate systems with differing governance, which can introduce complexity in aligning data policies, especially when third-party platforms are involved.

UPM Energy is the second-largest electricity producer in Finland, focusing on CO2-free energy generation. The business area operates through hydropower plants and nuclear power plant units through shareholdings. It participates in the Nordic electricity market and emphasizes low-emission electricity generation. In 2023, UPM Energy employed around 80 people and had sales totaling 628 million euros.

Data sharing in the Finnish electricity sector is shaped by both EU-level regulations and national implementations, aimed at ensuring transparency, consumer rights, system interoperability, and market fairness. At the EU level, Regulation (EU) No 543/2013 requires electricity market participants to publish data on generation, load, transmission, and outages through the ENTSO-E Transparency Platform. This supports fair market access and operational coordination. In parallel, Directive (EU) 2019/944 and Implementing Regulation (EU) 2023/1162 guarantee consumers' rights to access and share their metering data in a standardized format across all member states by 2025. These measures are reinforced by the Data Governance Act (2022/868) and Data Act (2023/2854), which provide a legal basis for trustworthy data access across sectors, including energy. To prevent market manipulation, REMIT (Regulation EU 1227/2011) mandates reporting of wholesale energy transactions and inside information, monitored by the EU Agency for the Cooperation of Energy Regulators (ACER). This applies directly in Finland.

At the national level, Finland's key implementation is Datahub, a centralized information exchange platform for the retail electricity market launched in 2022 and operated by Fingrid Datahub Oy. All electricity retailers and distribution system operators are legally required to use it for exchanging contract, metering, and customer data. Datahub ensures encrypted communication, GDPR-compliant data handling, and consumer control over third-party access.

The Finance and Controlling function at UPM supports all business areas through financial planning, reporting, and performance management. While not a business area in itself, this function plays a critical role in data consolidation, governance, and decision-making. It acts as a central node for financial data flows, budget forecasting, and investment analysis.

Within the scope of this thesis, Finance and Controlling provides valuable insight into cross-functional data coordination, especially where data must be harmonized across different systems, business areas, and geographies. This includes integration with ERP systems, business intelligence platforms, and compliance with financial reporting standards. As the whole UPM Group continues to advance its digital capabilities, the Finance and Controlling function is increasingly involved in shaping data architecture and governance frameworks that are utilizable by all areas while simultaneously aligning with regulatory requirements, business needs, and sustainability goals.

Out of all UPM Business areas, these three were selected based on some limitations and specialties. Firstly, given the scope of a Master's thesis, it was clear that not all business areas could be analyzed. Secondly, the business areas were compared in the early stages of this research by their core function, data maturity level, size, complexity and interconnectedness based on preliminary discussions with UPM stakeholders. Thirdly, the preliminarily selected areas and personnel were contacted in order to validate their willingness and capability to participate and represent their business area in this case study.

Given these considerations, the research ultimately focused on three areas — Fibres, Finance and Controlling and Energy — as they represent distinct yet interconnected domains. These areas encompass a broad range of core functions, making them suitable for comparative analysis. Additionally, they each make sufficient use of data and tools, allowing for a critical assessment of their data maturity and background understanding of DCs and DSAs. At the same time, they present identifiable challenges, which makes them relevant and valuable for investigation.

3.3 Research Methods

This study is a qualitative and empirical single-case study, with semi-structured interviews as the primary method of data collection. A single-case study design allows for in-depth exploration of complex phenomena within their real-life context (Stake, 1995). The selected case company provides a relevant environment for studying how data contracts are understood, implemented, and experienced by internal stakeholders.

Given that data contracts remain an emerging topic with limited prior empirical research, a qualitative approach was deemed most suitable to produce a nuanced and context-sensitive understanding. As Eriksson and Kovalainen (2008) argue, qualitative methods are especially well-suited in business research where the objective is to understand how and why certain processes evolve. A qualitative approach suits the exploratory nature of this thesis, which aims to uncover organizational dynamics, challenges, and perceptions surrounding the development and use of Data Contracts and Data Sharing Agreements.

Qualitative research also offers methodological flexibility, which is particularly valuable when studying evolving concepts like data governance and contracts. Braun and Clarke (2006, 2012) emphasize that qualitative methodologies, especially thematic analysis, enable researchers to uncover patterns and meanings in empirical data without pre-imposing rigid theoretical structures. This approach allows insights to emerge more organically from participants' experiences.

To access first-hand insights into organizational data practices, semi-structured interviews were selected as the primary method of data collection. Interviews were conducted with employees from different divisions of the corporation. As the interview approach is semi-structured, the interviews are conducted with predetermined questions in mind, but do not limit the conversation to only focus on them (Wilson, 2014). This approach allows the interviewees to enrich the conversation with more context and their individual knowledge (Given, 2008). This method supports inductive reasoning, ensuring that insights are grounded in the participants' own perspectives rather than predefined categories.

With Stake's (1995) categorization, this thesis follows the logic of an intrinsic single-case study. The research is motivated by the uniqueness of the case organization and its specific approach to managing data contracts, rather than aiming to generate generalizable theory. This intrinsic focus allows for a detailed examination of the organization's internal practices, making it possible to surface context-specific insights that may otherwise remain invisible.

3.4 Data Collection

To understand how data contracts are interpreted and applied within the organization, semi-structured interviews were chosen as the primary data collection method. This method offers a balance between consistency across interviews and flexibility to probe deeper into emerging themes. Semi-structured interviews are particularly useful in exploratory research, as they allow participants to articulate their perspectives in their own terms while remaining focused on core topics (Wilson, 2014; Eriksson and Kovalainen, 2008). The questions and a glossary of relevant terminology were sent to the participants beforehand and are displayed in the tables below.

Table 1 – Interview Questions

Question 1	Describe your position and experience at UPM and elsewhere, if relevant.
Question 2	Describe your business area/function (and its data culture).
Question 3	Which tools and systems do you mostly use when working in your role?
Question 4	Describe any use cases where you have been working with data contracts or data sharing agreements at UPM, or within your other experiences.
Question 5	What has your role been when working with these topics?
Question 6	How have your experiences with DCs and DSAs been? Describe any issues or successes you have had.
Question 7	How do you see DCs and DSAs having an impact on your work in the future?
Question 8	Where would you see DCs and DSAs being useful? Describe current situations where they are in use or situations that you think might emerge.
Question 9	How do you think work within your business area/function could change from a more widespread adoption of DCs and DSAs?
Question 10	How do you think your business area/function could be assisted in the adoption of DCs and DSAs?

Table 2 – Data glossary

Data Product	A data product is a set of data assets that is productized by adding services, interfaces, and standardization, making it modular, reusable, and purpose-driven. It represents one or more data concepts, encapsulates the necessary components for processing and storing data, and provides access through well-defined interfaces governed by a data contract.
Data Concept	A data concept refers to a fundamental idea, entity, or category used to represent and organize data. It provides a logical way to describe data's nature, meaning, and relationship to other data.
Data Asset	A data asset is any digital object or entity made up of data that has value and is logically organized. Data asset always has an owner. It could be a dataset, document, visualization or data service.
Data Contract	A data contract is a document that defines the structure, format, semantics, quality, and principles for exchanging data between a data provider and their consumers. Data contracts specify the expectations of data dependencies and verify given guarantees.
Data Sharing Agreement	A Data Sharing Agreement (DSA) is a mutual agreement between one data provider and one data consumer on data usage for a specific use case in a specific context. A DSA documents what data is being shared and how it can be used.

In total, 10 interviews were conducted. Each interview lasted between 40–60 minutes. Interviews were conducted via Microsoft Teams to accommodate participants across multiple locations. With participants' consent, interviews were recorded and transcribed using Teams' native AI-assisted tools. Transcripts were subsequently reviewed and corrected to ensure accuracy.

Interviewees were selected from three different business areas or functions. These areas differ vastly from each other, which makes for interesting juxtaposition and comparison possibilities. The roles in which interviewees are working range from end-users to higher-up managers, which also allows for drawing similarities and differences based on the views the different roles hold.

Interviewees were given the option to speak in either English or Finnish, depending on their preferences, to ensure that the data collected was as rich and accurate as possible. Direct quotes featured in the Findings-chapter were translated as directly as possible from Finnish to English, while idioms and metaphors were clarified into understandable language rather than rigidly translated. Questions were framed in accessible language, avoiding academic or overly technical jargon. The interview included a core set of themes but allowed room for improvisation and follow-up based on the participant's responses. This flexibility helped capture unanticipated insights. The interviewee details are displayed in the table below.

Table 3 – Interviewees

Alias	Title	Area	years at UPM	interview length (minutes)	language of interview
Interviewee E1	Manager, Strategy and Finance	Energy	13	50	Finnish
Interviewee E2	Data Engineer, Information Architect	Energy	2	49	English
Interviewee E3	Middle-office Manager	Energy	1	48	Finnish
Interviewee FC1	Business Control Manager	Finance and Controlling	8	54	Finnish
Interviewee FC2	Information Architect	Finance and Controlling	2	58	Finnish
Interviewee FC3	Data Solution Manager	Finance and Controlling	1	49	Finnish
Interviewee FC4	Data Solution Manager	Finance and Controlling	6	52	Finnish
Interviewee F1	Data Management and Analytics Manager	Fibres	4	42	Finnish
Interviewee F2	Data Manager	Fibres	19	54	English
Interviewee F3	Data Specialist	Fibres	2	53	Finnish

The interview guide evolved in parallel with the literature review and initial findings, enabling the researcher to iteratively refine the line of questioning and deepen the focus on emerging areas of interest. Interviews began with broad questions about participants' roles and relationship to data and contracts, followed by more specific prompts regarding processes, challenges, and organizational structures surrounding data contracts.

3.5 Data Analysis

Findings from the interviews are analyzed using thematic analysis. Introduced by Braun and Clarke (2006), thematic analysis looks for recurring patterns or themes in the research data. This method was utilized in applicable parts in the making of this thesis. Through thematic analysis, identifying common themes and issues across interviews, the goal is to gather the interviewees' insights on experiences, expectations, and concerns regarding DSAs and DCs. The data analysis for this thesis was based on Braun and Clarke's (2006) six-step framework:

1. Familiarization with the data – Going through all generated interview data, creating transcripts, listening to recordings, making initial observations
2. Generating initial codes: Assigning codes to the data by systematically identifying important features and organizing them
3. Searching for themes: Clustering related codes into broader themes that reflect overarching patterns
4. Reviewing themes: Iteratively refining themes for internal coherence and distinctiveness
5. Defining and naming themes: Formulating clear and precise labels for each theme
6. Producing the report: Integrating the themes into the structure of the findings, supported by direct excerpts from the interview data.

Coding was done manually using printed-out transcripts and sticky notes, where sentiments were extracted, annotated, and clustered into categories. Initial codes were purposefully quite broad, such as 'using the Common Data Platform (CDP)', 'agility', 'bureaucracy', or 'wild west'. Initial codes evolved through repeated iterations into more refined analytical categories, such as 'roles and responsibilities', 'opposing forces', or 'shared visions'. Some codes, such as the 'CDP', contributed to multiple themes, as it pertained to both balancing central and local governance, and Data Democratization.

While themes were primarily grounded in the empirical data, the analysis was informed by concepts from the literature review, creating dialogue between theory and data. This approach enabled the identification of both expected and unexpected patterns in participants' responses. The figure below presents an example of how the analysis progressed from initial codes to subthemes and finally themes, which inform the structure of Chapter 4 - Findings.

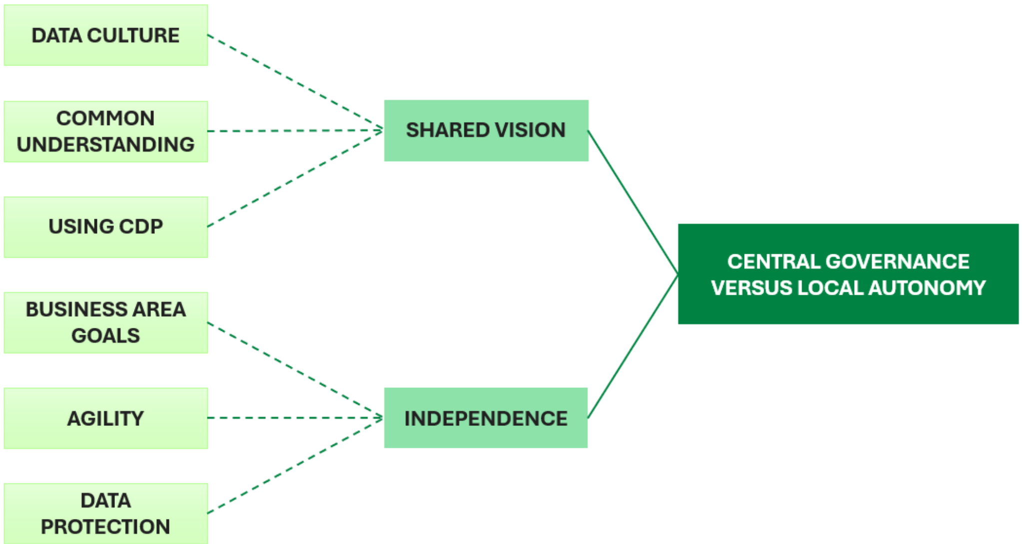

Figure 1 – Example progression of Thematic Analysis

3.6 Ethical Considerations

According to Nii Laryeafio and Ogbewe (2023), the main ethical considerations in qualitative interview data collection are anonymity, privacy and confidentiality, voluntary participation, option to opt out as well as avoiding misuse of acquired interview data. This study aimed to follow all the listed topics as well as possible. All participants were provided with clear information about the study's purpose, proceedings, and their right to withdraw at any time. Their participation being voluntary and the option to opt out were noted in the invitation emails, as well as in one-on-one conversations before and during the interview process. Informed consent was obtained prior to participation, both written consent in the form of email confirmations as well as a verbal agreement in the beginning of each interview. Interviews were recorded and transcribed only after obtaining verbal consent from the interviewees.

Efforts were made to ensure confidentiality, privacy and anonymity by assigning aliases to the interviewees and securely storing all data on encrypted, password-protected devices. Only the researcher has had access to identifiable information, which is to be destroyed following the completion of the study. Care was taken to minimize any potential psychological or social risks to participants, and all interviews were conducted in a respectful, non-intrusive manner. The findings will be presented truthfully and transparently, with appropriate credit given to all contributors.

The researcher was employed by UPM during the period of the study and received financial compensation from the company. This dual role as both employee and researcher presents potential risks for bias or perceived conflict of interest. To mitigate this, the research was designed with transparency, and participants were made aware of the researcher's position within the company. Furthermore, data interpretation was grounded in academic standards, and findings were reviewed with methodological rigor. Efforts were also made to separate the researcher's operational role from the research process. Interview participants were selected outside of the researcher's own function and across different business areas to ensure a range of perspectives. The researcher was committed to providing a realistic and truthful picture of the case study and prioritizing objectivity.

4 Findings

This chapter will go through the findings from the interview data, offering insight into the current state of DCs and DSAs in the organization, as perceived by chosen stakeholders involved in the data contract process. Interviewees are pseudonymized with abbreviations of their home organization, where employees from Energy, Finance and Controlling and Fibres are presented as E, FC and F, respectively. The chapter is structured into three sections, each addressing a core theme identified through the interviews. This chapter starts with generalizable topics, then deepens into more profound organizational challenges, and finally addresses DC and DSA-specific practical applications and paradigms. The narrative builds from foundations to enablers and finally to applications. For each theme, a summary of key insights from the interviews is provided.

First, the overarching governmental challenges are presented. In the interviews, many stakeholders admitted to feeling torn between opposing forces, while doing or trying to improve their work. These Balancing Acts are presented in chapter 4.2. While an important finding in and of themselves, they also lay the groundwork for the following chapters. Chapter 4.3 on Fundamental Challenges relates to organizational and cultural pain points reflected in the interviews. Lastly, in chapter 4.4, the findings specific to Data Sharing are presented.

4.1 System Level Government Tensions

Throughout the empirical material, a recurring theme emerged around the tensions and trade-offs inherent in implementing new data governance practices, particularly those related to data contracts, within a complex, heterogeneous organization. These tensions highlight the need to balance strategic ambitions with operational realities and reflect differing levels of data maturity, organizational readiness, and technological capacity across business areas. This chapter outlines these balancing acts.

4.1.1 Central Governance and Local Autonomy

A recurring theme in the interviews concerns the tension between organizational-level pull toward coherence and standardization and localized-level pull toward autonomy and context-sensitive business cases. On the one hand, centralized governance, harmonization efforts, and shared architectural visions are seen as essential for overcoming silos, reducing redundancy, and fostering interoperability across the enterprise. These aspirations align with broader ambitions to create common ways of working, shared routines, and architectural coherence. Interviewee E3 emphasized the ambition to move towards a more common way of working:

> *Some elements* [of code] *could be important not only locally but also to other business areas, and if written properly, they could be reused across multiple contexts.*

Yet, the very same processes encounter resistance or hesitation at the local level, where business units emphasize the necessity of flexibility, adaptation, and ownership. Strict standardization does not always align with operational realities, and in some cases, synchronizing all systems is neither feasible nor desirable, particularly with varying business models or regulatory conditions across markets. Several interviewees highlighted how data sharing continues to be hindered by siloed practices, despite investments in shared platforms such as the Common Data Platform (CDP). As interviewee FC1 explained:

> *Data sharing still remains a challenge. Even with the CDP, there are silos. Each business operates largely independently, and even within finance, teams work rather autonomously. This means best practices are not necessarily shared effectively, and at times, work is duplicated.*

This lack of cross-functional integration was similarly noted by interviewee FC2, who pointed to fragmented team structures and the absence of interoperability:

> *Each analytics team mostly uses what they have developed within their own silos. There has been no holistic or cross-functional view. Different parts of the organization use their own entity structures, which are not interoperable. And once something has been built in-house, it tends to be protected like one's own child, making collaboration even more difficult.*

These accounts show how organizational-level push toward integration run into localized practices of ownership and self-protection, producing friction rather than coordination. At the same time, participants stressed that local autonomy is not simply a matter of being territorial, but a necessity tied to the specific demands of their markets and regulatory environments. As interviewee E1 emphasized:

If someone says UPM should start doing things in a certain way, we assess whether it works for us. If it does, we adopt it immediately. But it is often hard to make it work, because we have a strong will here [at UPM Energy] *and existing practices that already function well. Synchronizing systems is not always in the interest of the business. For instance, in the electricity market regulation has a huge impact on what we are allowed to do. The obstacles are not about people refusing change, but about legal and market constraints.*

This perspective underscores that governance is not merely a top-down technical matter; it is entangled with sectoral regulation, local market conditions, and the embeddedness of established practices. Standardization may provide efficiency gains, but its suitableness to a given situation must be judged on a case-by-case basis, by the actual value it brings.

This tension between organizational-level and localized-level pull also manifested in the discussion of how new modes of operation should be implemented: through so called ‘pilot projects’, one subject at a time, or organization-widely, equally for every area. Pilot projects were often described as test subjects or “guinea pigs,” offering effective ways to experiment and demonstrate value and change, especially in areas with higher maturity. They are seen as momentum builders and proofs of concept. Yet, while pilots showcase potential, they can also create uncertainty and fragmentation if their purpose and outcomes remain unclear. As participant FC1 recalled of their first experiences with adopting DC- or DSA-style ways of working:

It was honestly quite unclear during the project phase. We were probably among the first guinea pigs when these whole data contract and data sharing elements were introduced. I did not fully understand why this was done or what the idea was supposed to be. It was a rather confusing process, as everyone was learning at the same time.

Others, such as participant F3 below, suggested hybrid models that balance the organizational and local pulls more effectively:

I would prefer a kind of hybrid approach. At first, involve everyone to some extent, and then select one bigger or better-resourced unit to pilot more intensively. That way, it is easier to gain experience while others remain somewhat involved, rather than hearing afterwards that something was done without them.

This reflects a desire for mechanisms that allow experimentation without excluding others, combining the agility of localized pilots with the legitimacy of organizational inclusion.

Pragmatic steps were also described as potential ways to bridge the divide. Incremental harmonization, such as introducing shared naming conventions or centralized visibility into data flows, was seen as enabling coordination without threatening local autonomy. As interviewee FC4 stated:

> [Standardization] *doesn't always have to mean rethinking everything from scratch. Even something as simple as how integration schemas are named can become a routine. It is good that we now have centralized data, and we can see what integration schemas are shared and where they are distributed. Still, I would like to have a clearer overview of 'my data' as a data product owner.*

These small but concrete measures highlight how organizational-level coherence can be advanced without imposing disruptive or rigid uniformity. Together, these accounts suggest that balancing between organizational and localized pull is at the core of governance in complex data environments. Centralization promises efficiency, integration and shared direction, but it risks undermining local realities, ownership and contextual specificities. Conversely, local autonomy allows responsiveness and contextual fit, but without structures of coordination it perpetuates silos and duplication. Navigating this tension requires not only strategic governance design but also pragmatic compromises, such as hybrid pilot approaches and incremental harmonization, that enable coherence while respecting diversity.

4.1.2 Control and Agility

The introduction of formalized Data Contracts embodies a duality of promise and apprehension, showcasing a broader tension between control and agility. On one side, advocates emphasize the role of contracts in clarifying responsibilities, improving documentation, and reducing dependence on individual knowledge. They create transparency regarding data dependencies, enable the enforcement of standards, and contribute to more robust and reliable data pipelines. By setting agreed expectations, contracts also provide a reference point for accountability. As respondent E3 explained:

> *Data contracts bring clarity. And at the same time, they set obligations to those who are responsible. When a certain level has been defined, that we want to maintain, it becomes easier to rely on.*

This view was echoed in accounts highlighting the technical enforcement possible in technical environments in use at UPM, such as dbt. Participant FC3 explained:

> *In dbt, there are technical features that prevent you from breaking the models you have shared. The benefit is that you can rely on the model defined in the data contract, so that when it is updated, your own model updates as well. We need to invest in the technical side and in those dbt data contracts.*

Yet these perceived benefits are accompanied by concerns about rigidity and overhead. Contracts risk being seen as compliance exercises—documents that are rarely consulted unless something goes wrong—rather than as living instruments that actively support collaboration. This fear was captured by participant FC3:

> *It feels like with Data Sharing Agreements, we're trying to include as much info as possible, even when it's not needed. For me it should only state what the data is used for, who requested it, and who gave the approval. I don't think anyone uses or needs them afterwards, unless something has gone wrong. Otherwise, they're just a compliance thing.*

This distinction between proactive value creation and reactive compliance runs through many accounts, raising questions about whether the additional work is justified if the perceived benefits remain intangible in day-to-day practice.

The administrative burden of contracts also generated frustration among teams accustomed to working quickly. As interviewee FC1 recalled:

> *In data projects we often want to move forward quickly. We have ideas about how to implement something but then run into bureaucracy. Either we can't access data or need multiple approvals, which takes time. This slows down our work and increases costs, since we often work with external parties. From an efficiency perspective, it felt like a burden.*

This illustrates the dilemma: while contracts provide safeguards and governance, they may also slow down innovation and create barriers to experimentation if the processes are perceived as too heavy. Participants also pointed to the practical risks of overly restrictive access processes. If obtaining data through formal systems is too slow or uncertain, workarounds emerge. As interviewee E2 stated:

> *It's very hard to share secrets in CDP. I can tell access management that someone needs access, and it just doesn't happen. If I have to wait two weeks and still nothing happens, then I just share it directly over less secure channels. Getting the data fast is the most important.*

Such accounts highlight how excessive control can paradoxically undermine security by incentivizing undesirable practices. This ease-of-access issue (or lack thereof) also raises the question of the broader trade-off between speed and by-the-book-ness. In many business contexts, the timely availability of data, even if imperfect, is seen as more valuable than waiting for complete, fully validated datasets. Participant F1 describes this:

> *It is often better to have something than nothing, so we can react. We can't just stay stuck in endless discussions or have seven people debating how to name a single field. Of course, things can be done very diligently and slowly, but then we may never deliver anything.*

Several participants highlighted how the value of any new guidelines, especially DCs or DSAs, is realized unevenly over time. As respondent F3 summarized, speculating on the effect that adopting DCs and DSAs might have on his everyday work:

> *In the short term it adds work, in the medium term it becomes neutral, and in the long run it saves time.*

This temporal dimension suggests that the tension between control and agility is not static, but shifts as practices mature and as the accumulated benefits of standardization begin to outweigh the initial investments. Nevertheless, interviewees stressed the importance of minimizing unnecessary bureaucracy through templates or simplified procedures for low-risk cases, in order to prevent innovation from stalling.

Balancing between control and agility emerged as a core governance dilemma. Contracts and formal structures provide clarity, accountability, and reliability, but they may also impose rigidity and slow down the speed and responsiveness that data-driven business aims to achieve. Moving too fast or haphazardly can undermine trust in the data, while excessive formalization risks stifling innovation. Navigating this tension requires context-sensitive solutions: lightweight templates for low-risk cases, stronger technical enforcement where dependencies are critical, and a careful assessment of processes to ensure that governance mechanisms enable rather than obstruct organizational agility.

4.1.3 Short-Term Fixes and Sustainable Practices

The legacy of siloed and ad-hoc solutions continues to shape the organization's data landscape. Interviewees described how their data organizations have been historically guided by a culture of quick problem-solving, which provided agility in the moment but now complicates the establishment of more sustainable practices. At the same time, some interviewees emphasized that ad-hoc solutions continue to generate value, and that their organizational culture still relies heavily on them. As F1 put it:

> *Data work is still very ad hoc, and most of the value we create still comes from those kinds of solutions. We are very good at making quick fixes, but now we also need to define how we maintain and best utilize them.*

This account reflects the aspiration to develop interoperable and reusable data products that reduce duplication and inefficiency. The promise of a shared platform lies in its ability to create standardized outputs that benefit multiple projects at once. Interviewee FC2 illustrated this point:

> *Does it make sense to produce the same thing through two different routes? Or should we define a common standardized data product that can be shared across projects? That way, we save enormous amounts of time and resources. The whole idea is not to reinvent the wheel or throw money out the window.*

From this view, sustainable practices are not simply about technical solutions but about embedding a mindset of reusability and efficiency within the organization. However, this shift is complicated by the persistence of legacy infrastructures. Participant E1 framed their current situation, moving from a 15-year-old legacy system to a new shared data platform:

> *Of course, this* [old system] *does not work in the long run, but changing it is not easy either, since all operations depend on it. That is why*

> *the transition to the Common Data Platform has to be carried out piece by piece.*

This dependency illustrates how short-term fixes can gradually become long-term constraints, making transformations to more longstanding solutions slow and risky. As interviewee E2 phrased it:

> *Unfortunately, it takes a very long time to migrate* [the current database] *to CDP. I would have liked to be done by now already, but there's just so many processes that hang to it, and it's very difficult to change a tire on a running car.*

The lack of proper documentation seems to make implementing sustainable, long-term solutions even more difficult. However, even when participants agreed on the need for proper documentation, their faith in getting better practices implemented seemed to be dimmed by time and resource constraints. Many interviewees shared in this frustration and stressed the importance of shared documentation practices, such as FC2:

> *Even if you have the finest solutions, if no one knows about them, someone else may end up reinventing the same thing. But documentation always falls short because there is too little time, and no one is sure whose responsibility it is.*

E3 elaborated in a similar vein:

> *Everyone who knew anything about the system had already left. No one really understood what the code was supposed to do or even where it was. It took me a year to document it while trying to manage daily tasks at the same time. Things would be clearer if everything was documented in a way that makes it easy to find. That would be ideal. But I don't believe in any organization these things are ever perfect, because at some point the effort of maintaining all the documentation becomes too great, unless you hire someone just for that. Do we want to take on that cost?*

These accounts demonstrate how reliance on short-term fixes, without accompanying documentation and governance, perpetuates inefficiencies and risks, and hinders the move toward sustainable practices that could stand the test of time. Many of the existing solutions originated as temporary workarounds that successfully addressed immediate problems but gradually produced fragmented infrastructures, which today prove difficult to integrate or scale.

4.2 Organizational Readiness and Roles

Many participants in this study shared thoughts of unclear roles and responsibilities as well as the prevailing culture around data. Questions of the organizations capability and willingness to act on them were pointed out by on multiple occasions. This chapter goes through these organizational and cultural findings.

4.2.1 Roles, Ownership, and Accountability

A recurring challenge in data-related projects is the unclear definition of roles and responsibilities. Confusion often arises over seemingly basic tasks, such as where documentation should be stored as discussed in the previous chapter, especially when different teams employ different tools and systems. Within the context of Data Contracts, the responsibilities of each actor, including data owners, information architects, and project managers, have often been unclear.

Even the role of the information architect has been subject to interpretation. In some cases, questions emerged as to whether architects should act merely as facilitators or also contribute content-wise.

Several interviewees reported that development teams sometimes failed to prioritize enabling data sharing, even when specifications and contractual agreements were formally in place. These experiences highlight the importance of active engagement from both data owners and project leads to ensure that projects advance according to plan. As Interviewee FC1 reflected on their past experiences:

> *I was the project manager, responsible for the project. I must admit that it remained somewhat unclear what the role and responsibility were in the data contract, for example.*

Several participants emphasized that clearly defined roles, responsibilities, and rights are crucial to avoiding project deadlocks. When authority is unclear, the individual expected to act may lack both the tools and mandate to do so, rendering data contracts theoretical rather than practical instruments for change. In particular, the role of the Head of Data (HoD) was reported as inconsistently interpreted and sometimes ineffectively executed, while information architects (IAs) frequently described feeling isolated and overburdened, working largely alone within the data production chain. As stated by participant F1:

> *One big challenge is that we don't have a common alignment on different levels, and we lack a shared vision in a way that we could get the heads of data involved, and I don't really see that working at the moment. -- Many IAs feel alone with these tasks, which are not small. Even though many people work in the production chain, only a few work with the data in it. In IAWG, we have half-time, full-time, or one-seventh-time people in the group, with no clear mandate or resources to execute things, and that shows.*

A major perceived advantage of data contracts lies in their ability to enforce and clarify data ownership. By requiring that a specific individual assume accountability for a given dataset, contracts help address the longstanding problem of unclear or absent ownership. As noted by participant FC1:

> *What Data Contracts and DSAs can certainly bring is that they force ownership onto someone, as someone must take responsibility, and that person becomes the owner of this data. Even those who have been developing it for a long time may not always be clear about who is accountable, so I see this as a positive opportunity from* [Data Contracts and Data Sharing Agreements].

Clearly documented ownership enables accountability, ensuring that stakeholders know whom to contact, and provides a foundation for more structured and reliable data governance. Additionally, DCs and DSAs clarifying roles can be seen alleviating pre-existing dependencies on individual people. As participant E1 stated, hoping that these tools could include the topic of person-dependency as well:

> *In many parts of our operations the scale is so small, with so few people, that we have maybe just one or two people who know how to do certain things. The person-dependence is huge. I think this should be seen there* [in DCs and DSAs] *as well. This is really critical in an area where the whole business relies on data. It's tough if someone leaves.*

However, despite this potential, many parts of the organization still lack formally assigned data owners, or assigned individuals remain uncertain about their responsibilities, a situation that continues to hamper the full realization of contractual benefits.

4.2.2 Leadership, Culture, and Maturity

Many of the challenges related to roles, ownership, and accountability are underpinned by broader organizational structures and leadership models. Without foundational elements, such as clear ownership, well-defined roles, and repeatable practices, it is difficult to adopt more advanced data practices, including the use of data contracts, particularly in less mature business areas. The decentralized nature of data team leadership, where teams operate independently, often deepens fragmentation and inhibits harmonization.

Interviewees repeatedly emphasized the need for stronger top-level support and strategic vision. At present, they feel there is no comprehensive company-wide mandate for Data Contracts, and data work is not consistently prioritized at the leadership level.

Leadership involvement is particularly important in recognizing the value of less visible work, such as governance, documentation, and system alignment, which is essential to enable high-impact outcomes. Respondent F1 described the challenges of many responsibilities with limited resources:

> *What could go better is prioritization: are we mature enough to focus on the basics so that we are ready for this work? We simply cannot handle it all. We need proper roles and responsibilities, and this is part of maturity: we cannot expect the work to be done unless it is resourced appropriately. It is very difficult to get resources for this invisible, tedious work.*

Additionally, multiple interviewees stated that expectations for adopting standardized practices must be aligned with performance metrics, as it is unreasonable to hold individuals accountable for outcomes they cannot influence. Speculating on the success of DC and DSA implementation, participant E3 shared:

> *It really depends on how these are tied to people's performance bonuses. My predecessors had no pressure to complete documentation, so it was not done. Immediately after my predecessor left, objectives*

were set for my colleagues to produce documentation on certain topics, so these kinds of problems would not happen again. If incentives are aligned, the work gets done.

Interviewees also highlighted the need for accountability structures to support enforcement of shared rules. As FC2 noted:

> *Mandating shared rules should make things easier because everyone is required to follow the standard practice. But then the question is: who is responsible for it? You cannot introduce a standard process if the people whose KPIs and goals are affected by the change do not have the ability to influence it. A person cannot be responsible for things they cannot control.*

A further challenge lies in the mismatch between technological capability and organizational readiness. While tools and platforms increasingly enable structured and scalable data practices, many areas of the organization still lack the foundational elements for effective adoption, including clear roles, defined processes, and basic ownership practices. Even the most capable technologies cannot compensate for insufficient cultural or procedural maturity. As FC1 observed:

> *The basics still need to be put in order. Maybe the DC/DSA can be one thing that helps, as it can bring ownership and clarify roles or at least force the area to consider them. But it may be that the organization is not mature enough yet.*

Cultural differences also play a significant role in shaping data practices. Interviewee FC2 reflected on the current organizational culture, noting:

> *If I had worked like this at my previous company, I would have been fired for spending so much money and time on meetings. That was a culture where efficiency was paramount. Here, you cannot do your work without ten people in a meeting, because otherwise, no one would trust you.*

Finally, strong local leadership can both enable and constrain progress. Participant FC2 highlighted the complex interplay of status, prestige, and informal power:

> *The data team leader is in a position where everyone asks them for customer data—they have the power, the status, the prestige. They could say, 'I don't believe in this, so we'll decide ourselves,' but these are really organizational issues. It's not the data itself; it's the organization, processes, people, and systems.*

4.3 Data Sharing Paradigms

As the main topic of this thesis, DCs and DSAs, focus on easing and better utilizing data sharing, findings on these topics are presented as their own chapter. Many of the points in this chapter echo thoughts from the previous findings, such as balancing opposing forces from chapter 4.1, but in this chapter are viewed through a more data-specific lense.

4.3.1 Data Democratization, Protection and Compliance

A central tension in organizational data practices lies in balancing openness with control. On one hand, broader access to data is widely regarded as a necessity for innovation, value creation, and effective problem-solving. Simple report sharing is often seen as insufficient, as more tangible value is thought to be realized when stakeholders can access data and work with them directly. On the other hand, concerns related to privacy, cybersecurity, and regulatory compliance, especially when data includes personally identifiable information or business-critical insights, require careful governance and limit what data can be shared, with whom, and under which circumstances.

Several participants highlighted the benefits of structured data access frameworks. Respondent FC1 explained:

> *At best, they* [DCs and DSAs] *generate value and, in particular, visibility into which data is being used, how it is being used, and why. This creates a good structure around data sharing, with clear rules about how it can be done.*

Such frameworks, whether formalized as data contracts or otherwise, help establish clarity around usage and responsibilities, supporting both efficiency and accountability. FC4 recalled that in earlier stages, data was simply obtained as needed:

> *I don't know if you would call it the Wild West, but back then, it was done in whatever way seemed best at the time: I needed the data, so I got it however I could.*

These informal practices, while enabling speed and flexibility, risk creating inconsistencies, redundant work, and unclear accountability.

Several participants noted that the introduction of formalized data contracts can reduce the fear of sharing by clarifying levels of access and establishing rules. As respondent F2 stated:

> *Perhaps these data contracts ultimately reduce the fear of sharing. If we achieve a solution that clarifies access levels and establishes rules, we are removing blockers.*

By defining explicit boundaries and responsibilities, contracts create a shared understanding that balances democratization with necessary protection, supporting both compliance and effective collaboration. Participant E2 noted the practical benefits of explicit definitions when it comes to democratizing data:

> *If we have a contract saying we deliver it every day, then we try to do it. If there's even more specification about the format, then we try our best to keep it like that. So, there's an advantage for all parties in knowing that data is appropriate.*

Overall, the findings suggest that organizations must navigate a careful trade-off between enabling access for innovation and maintaining safeguards to protect sensitive or regulated information. Structured agreements, transparency in usage, and clear rules are all critical enablers for achieving this balance.

4.3.2 Trust-Based Sharing and Engineered Control

There is an ongoing debate between the sufficiency of trust and the need for technical control mechanisms in data sharing. Some stakeholders argue that excessive formalization erodes trust and reduces discoverability, hindering data-driven innovation. They advocate for cultures of openness, where data sharing is assumed unless explicitly restricted. Others take the view that trust must be deliberately designed through technical safeguards and governance controls. From this perspective, trust is not the absence of oversight but rather the outcome of transparent and enforceable structures, "the lock on the door" when it comes to protecting sensitive data.

Data sharing practices in organizations are shaped by an ongoing tension between trust-based collaboration and the need for technical control mechanisms. In the literature, some scholars emphasize the sufficiency of trust as a governing principle, arguing that over-formalization can erode relational confidence, reduce discoverability, and ultimately hinder data-driven innovation. In such perspectives, openness is assumed: data is shared freely unless there are explicit reasons to restrict access. Proponents of this approach advocate for a culture where actors rely on interpersonal relationships, familiarity, and shared norms to determine appropriate data use. Participant F2 stated:

> *If the interest is valid, why not give the data if that person is creating value from it? We don't have very clear policies, and I don't want to have very clear policies. It's a matter of trust and good data practices. Despite having cybersecurity measures and access rights management from the platforms, we manage by trust—knowing the people, understanding their interests, and judging whether access is relevant. Of course, conflicts exist, financial data is strictly limited, and personal data is sensitive. But overall, trust guides our decision-making, and I cannot imagine an automated contract replacing this judgment.*

This illustrates a practical reliance on relational trust: stakeholders balance the potential value of data against the need to protect sensitive information, using personal knowledge and contextual understanding rather than rigid formal rules.

However, other perspectives highlight the necessity of deliberately engineered trust. In this view, trust is not the absence of oversight but the product of transparent and enforceable structures safeguarding sensitive data. Technical mechanisms, such as data contracts, formalize and enforce agreed-upon terms of use, providing both clarity and accountability. As interviewee FC2 noted:

> *We have a working version and summary in PowerPoint, but in dbt we start implementing technical contracts. This really ensures that things are 'by the book'—so it's a very good step forward. Trust cannot be assumed, it has to be engineered. That's why we create data sharing*

> *agreements: everyone understands the terms, and finally, the last step is technical enforcement. It's like the final lock on the door.*

This statement highlights a shift from informal, trust-based management to a hybrid approach, where trust is complemented with technical and contractual safeguards. Engineered trust makes accountability, reproducibility, and legal compliance part of the sharing process itself, so that data is handled responsibly.

While much of the debate around trust and control focuses on restricting or enabling access, some interviewees mentioned that the true value of data often lies not in the datasets themselves but in the contextual knowledge and analytical work that accompanies them. This was particularly evident in the Energy area, where much of the underlying market data is already publicly available. As participant E2 explained:

> *Market information is mostly public, so there's sort of no risk at all. If something leaks, well, everybody knows already. It's more about the shaping of it, how do you format it and present it? Other business areas don't have detailed knowledge of the Energy sector, so they want a good representation of the costs, how to run their plants, what the energy costs will be. But they don't want to gather thousands of inputs and put it all together. Most of the data is public anyway, so in that regard, I don't think any contract would mean anything there.*

This raises the point that for some domains, governance frameworks cannot be limited to controlling access. Instead, the competitive advantage arises from the capacity to curate, analyze, and contextualize widely available data so that it becomes meaningful for decision-making. In these cases, data contracts and sharing agreements may play a smaller role in protecting the data points themselves, and a larger role with the processes of analysis, development, validation, and delivery.

The contrast between trust as an informal social norm versus trust as an engineered mechanism illustrates governance dilemma. While a more laid-back approach to trust supports flexibility, rapid innovation, and collaboration, technical control provides assurance, enforceability, and protection against misuse. Organizations navigating this tension should seek to balance cultural and technical measures, tailoring their governance frameworks to the qualities of the data, the stakes of potential misuse, and the strategic value of openness. When balancing these two opposing forces, it is key to recognize that the competitive advantage of data does not always come from the raw data points themselves but from the analyzing, shaping, and contextualizing it into knowledge.

5 Discussion

This chapter discusses the findings of the case study in their context, in relation to the existing literature reviewed, and offers insight for future practices for the case company. Firstly, this chapter analyzes the findings in relation to the case organization: how the different features of studied roles or business areas define the way topics are seen, which findings are applicable and generalizable across the whole organization, and which themes could be true for an even wider context outside of this study. Secondly, the chapter focuses on key literature domains against which the empirical findings are reflected. In the third part, the case is reviewed with the research questions in mind. Finally, practical recommendations for the case company based on the identified challenges and improvement opportunities are presented.

5.1 Findings in the Case Context

Across all business areas, several shared sentiments were expressed. On positive notes, interviewees highlighted the potential of Data Contracts (DCs) and Data Sharing Agreements (DSAs) to establish shared ways of working, increase standardization, improve data quality, and strengthen transparency and documentation across the whole UPM Group. They were seen as tools that could enhance reliability and promote collaboration across organizational boundaries. Importantly, they were also described as mechanisms that enforce ownership, which has been a persistent challenge in many areas. These benefits were considered particularly significant in large, cross-functional or complex domains and cases where coordination has been difficult.

At the same time, several challenges were consistently noted. Large organizations face inherent difficulties of leadership, prioritization, and competing demands. Many interviewees pointed out that people are often simultaneously under pressure to move quickly yet weighed down by slow processes. Problems related to lacking data ownership or documentation, difficulties in efficient collaboration, and siloed practices were recurring themes across all roles and business areas. Another major concern was resourcing: while there is a strong will to use data, there seems to be less willingness to invest in the governance structures needed to manage it sustainably, or even the understanding of why this would be important. This problem with resourcing applies for both the lack of employees in the data organizations, but also the existing employees' time being stretched thin over many topics at the moment.

When comparing the different business areas, distinct characteristics can be observed. In Fibres, the self-described lower confidence in data utilization and maturity may be due to the pulp business being traditionally less data-driven than areas such as Finance and Controlling or Energy, which are inherently numerical domains. While there is belief in the benefits of DCs and DSAs, there are also concerns that the maturity of data practices and data product thinking is not where it needs to be yet. Introducing new processes too quickly is feared to worsen already heavy work loads of Fibres' data organization.

In Energy, data is highly valued and forms the basis of all operations. Constant availability of data is critical to both their core business as well as legal and compliance reasons. However, the area struggles with outdated legacy solutions and long-standing documentation challenges. Interviewees expressed hope in DCs' and DSAs' potential to clarify responsibilities, improve data quality and speed, and even create efficiency gains that could constitute competitive advantage. At the same time, there were concerns about bureaucracy and slow implementation. A unique feature in Energy's core business is the requirement that a significantly larger portion of the data must be more publicly available than in other areas, raising questions about how this openness is to be managed in practice – especially when solutions are wished to be implemented affect the whole organization.

In Finance and Controlling (FC), the entire function is built around processing data for both its own needs and for those of other business areas. Data work therefore takes on more of a service role, with FC producing and distributing data for others to use. Within FC, however, significant differences in capabilities and maturity were noted between units and even individuals. DCs and DSAs were expected to formalize cross-departmental collaboration, establish clear rules, and enforce ownership. They were seen as a way to reduce duplicated effort and make it possible to reuse solutions across different areas rather than reinventing them each time.

However, concerns were raised about the complexity and workload associated with implementing DCs and DSAs, as well as the lack of visibility into how other areas define or use their data. This lack of transparency makes it difficult for FC to provide suitable data services and often leads to situations where the same data distributed by FC must be processed separately in multiple areas, likely creating another spot of duplicated work that could be eliminated through improved coordination. Interviewees emphasized that greater efficiency would be achieved if FC could provide the kind of standardized, ready-to-use data for all users that their function currently strives to provide. However, this goal has not been realized, given the fragmented practices outside of FC as well as internal maturity issues.

Within the analysis, the role of organizational position also emerged as an interesting point of analysis. The more people work across departmental boundaries in their daily tasks, the more positively they tend to see DCs and DSAs. Information Architects (IAs), in particular, recognized the potential benefits very clearly and often advocated for them. At the same time, they stressed that making these agreements work requires organizational-level commitment and clear role definitions, which are not always in place. Limited resources also mean that IA work can be deprioritized. Current 'Head of Data'-roles were described as ambiguous and not useful in practice, which further complicates IA work. While IAs have a strong understanding of the strengths and weaknesses of their areas, they often lack the mandate to steer data work in a unified direction.

In regard to how the participants themselves worked in the data pipeline, most interviewees held the role of data producers with only a few being end-users as well. For producers, DCs were seen as helpful in clarifying how data is to be used and why, providing much-needed structure and rules. At the same time, data producers emphasized that drafting contracts can be time-consuming and costly, involving lengthy meetings and at times, external consultants. In the short term, this was seen as reducing efficiency, as many employees already operate at full capacity and cannot easily absorb additional tasks. While some interviewees believed that implementing DCs and DSAs would ultimately reduce workload, the immediate concern was that the current burden leaves little room for added responsibilities.

For data users, DCs and DSA's were viewed as a way to improve data quality and reliability, both of which are essential for decision-making. They could also help to understand data dependencies and prevent failures. However, some speculated that less technologically savvy end-users might not see much benefit in fancy new solutions, if existing solutions work "well enough." For end-users, quick access to data was seen as more important than perfection. Concerns were also raised that discovery and development could potentially become harder if DCs and DSAs introduce too much control, slowing down innovation.

5.2 Findings in Relation to Earlier Research

One of the most prominent overarching observations is that several of the identified problems, especially those related to opposing organizational tensions, unclear roles and responsibilities, as well as conflicting priorities, can be traced back to a set of foundational issues. These challenges seem to be interconnected in that they all at their core deal with the case organization's overall data maturity and more specifically its governance and utilization capabilities.

The lack of clearly defined data ownership and role responsibilities rose as a central and recurring problem. This foundational issue underlies many of the tensions highlighted in the empirical material. For example, efforts to assign and enforce data ownership through mechanisms such as data contracts are often met with the reality that data owners either do not exist or, if identified, do not fully understand their role or the data they are expected to manage. In this context, data contracts are perceived as a possibility for good because they force someone to assume responsibility. This is something that has remained neglected thus far.

The absence of clear ownership also manifests in difficulties related to documentation. Although comprehensive documentation is broadly recognized as essential, it is often deprioritized due to time constraints and, again, a lack of assigned responsibility. As a result, knowledge remains siloed and dependent on specific individuals. This finding resonates with earlier research that traces the origins of silos to poor forward planning during growth phases, a lack of shared understanding between providers and consumers, non-integrated technologies, and overly narrow team-centric perspectives (Sadiq and Indulska, 2017; Skinner, 2022). As Günther et al. (2017) emphasize, these silo-inducing factors tend to reinforce one another over time, creating durable barriers to enterprise-wide data integration. In such environments, even the most sophisticated technical capabilities cannot be fully utilized.

The lack of clear responsibilities hinders the effective adoption of new operating models such as data contracts, creates confusion around responsibilities, and slows down processes. Furthermore, the unclear expectations placed on roles such as project managers or information architects, and the limited participation of those nominally responsible for data (namely HoDs), reflect a broader lack of commitment. Even upper-level roles like Heads of Data often fail to meet expectations, reiterating the structural and cultural nature of this challenge. These findings support the view of Khatri and Brown (2010), who stress that political challenges, including interdepartmental power struggles, unclear resource allocation, and mismatched performance incentives, tend to complicate the establishment of coherent governance structures. Abraham et al. (2019) also state that the balance between open flows of information and managing risk has become increasingly complex in today's data-intensive environment. Together, these issues point to a common root cause: the lack of clear and accountable data ownership, as well as insufficient role definitions overall.

Closely tied to the ownership problem are cultural factors and the organization's readiness for change. Many of the tensions observed relate directly to how the organization approaches change and data governance broadly. One key tension arises between the perceived need for formal contracts and standardization on the one hand, and the fear of increased bureaucracy and inflexibility on the other. Data contracts, for instance, can be regarded as overly theoretical, compliance-driven instruments that may hinder innovation rather than enable it. This perception can be due to resistance to change or a limited understanding of the potential benefits that new practices could offer, or most likely, both.

As discussed in the literature review, Günther et al. (2017) and Ransbotham et al. (2019) state that when employees see direct, tangible benefits from data sharing, and when incentives are aligned with collaborative behaviors, participation is more likely to increase. However, this case proposes that even when employees see the benefits, the absence of organizational prioritization leads to recurring under-resourcing of governance work. While the literature focuses on aligning incentives at the individual level, this case study suggests that resource allocation at the structural level is equally, if not more, crucial.

The organization appears to struggle with balancing agility with precision. A cultural preference for speed and "good enough" solutions is often prioritized over accuracy and comprehensive documentation. As long as existing ways of working seem to produce acceptable results, there is little incentive to adopt more stringent rules or more structured data practices. This tendency reflects a broader organizational tension: while accelerating the pace of decision-making is a common aspiration in data-driven environments (Huyn et al., 2019), it can lead to compromises in data quality and visibility. Wang and Strong's (1996) widely cited framework shows that quality is multi-dimensional, and all relevant dimensions (intrinsic, contextual, representational, and accessibility) shape whether data is trusted and used. Additionally, Sadiq and Indulska (2017) emphasize that quality assurance must be embedded throughout the data lifecycle, not treated as an afterthought.

While more advanced areas may be capable of serving as proof-of-concept examples, other parts of the organization require broader support to adopt new practices effectively. Furthermore, the prevalence of ad hoc technical solutions, often described as "band-aid fixes", highlights an organizational culture that has historically worked on local improvisation, at the expense of long-term maintainability, sustainability, and integration.

The underlying issue here is the variation in organizational readiness and cultural orientation toward data. Some areas of the company display more advanced data literacy and invest accordingly, while others report that their organization lacks basic foundations. These disparities prevent the organization from developing a shared contextual understanding of data, which Sadiq and Indulska (2017) identify as essential for trust and usability. The perceived inadequacy of strong leadership alignment and a shared strategic vision further complicates the problem. These disparities prevent the organization from effectively utilizing available technological opportunities. The perceived inadequacy of strong leadership alignment and a shared strategic vision further complicate the problem.

Another fundamental challenge lies in the fragmentation of operations and the lack of standardization. Organizational silos and the absence of shared practices emerged as recurring themes in the empirical material. While flexibility at the local level is often valued, it is also seen to result in structural gaps, duplicated work, and inconsistencies in how data is managed and interpreted. Decentralized leadership among data teams encourages custom ways of working, making harmonization efforts significantly more difficult. These findings align well with Shubladze's (2023) observation that democratization efforts are often framed around dismantling data barriers in order to make information broadly accessible.

The need for joint action, common standards, and clear guidelines becomes increasingly top-of-mind as digitalization and data governance mature. The continued reliance on incompatible, locally developed and at times even outdated solutions contributes to fragmented systems that interfere with effective data sharing. Tensions also arise between aspirations for open data access and democratization, as well as concerns about data protection, privacy, and regulatory compliance. This echoes earlier research pointing to the challenge of balancing openness with security (Abraham et al., 2019). As mentioned in the literature review, a global player such as the UPM Group must be aware of multiple different legislatures (e.g. the GDPR in Europe, or the LGPD in Brazil), in order to stay compliant but also allow for data democratization where possible (European Union, 2016; ANPD, 2018).

Addressing these concerns requires clear and standardized processes, as well as solid rules for responsible data use. The organization's siloed structure and the limited visibility into how other units define and manage data make it difficult to collaborate effectively and to share data in an effective and scalable way. The case study clearly illustrates a need for harmonized practices, including the structural clarity that Data Contracts and Data Sharing Agreements could provide.

These findings echo a challenge broadly recognized in the literature. Balancing standardization, accuracy and control with flexibility and agility is not an easy task. The naissance of the OCDS stemmed from this exact issue, where they found differing data formats greatly hindering cross-publisher analysis (Open Contracting Data Standard, 2024). The UK Governments implementation of the OCDS ran into the same issue, where the desire for consistent structures clash with the pre-existing realities and differing contexts (Crown Commercial Service, 2020). The literature has recognized this balancing challenge between agility and governance on multiple occasions, for example by causing delays or overruling practical business needs (Khatri and Brown, 2010; Vilminko-Heikkinen and Pekkola, 2019)

Lastly, many of the organization's challenges stem from how resources are allocated and how work is prioritized. Resource constraints, namely in terms of time, continuous education and staffing, were frequently cited as barriers to progress. Documentation work, for example, is under-resourced time and again despite its recognized importance. Likewise, the introduction of formal agreements such as Data Contracts is perceived as an additional workload, the benefits of which are not always seen as clear or even "worth it" in the short term. This hesitation points not only to a lack of resources, but also a lack of compelling incentives. Quieter work related to data governance is frequently relegated to the background in favor of more immediate or visible demands. As a result, long-term investments in foundational capabilities have been repeatedly postponed, accumulating more and more technical debt as time has passed.

The root cause here lies in the uneven prioritization and the chronic under-resourcing of critical, though less visible, data governance tasks. Activities such as documentation, model development, and implementation of shared standards are often deemed tedious and unglamorous and therefore fail to receive the support they would need to best be implemented. This may cause development teams to not prioritize enabling data sharing even when contractual frameworks are in place.

Taken together, these findings point to a shared underlying issue: the immaturity and underdevelopment of the organization's coherent data governance capabilities and appreciation of data culture. This immaturity is observable across multiple dimensions. On the individual level, role ambiguity, at-times low data literacy, and challenges in assigning responsibility prevent progress. At the process level, ad hoc workflows, missing standards, and the generally slow processes make it difficult to implement better practices. From the high-level Group perspective, while individual areas strive for improvements and achieve some levels of progress, the lack of ways to implement best practices to other areas reduces both resource allocation and the overall value of these initiatives, since they remain too fragmented to enable cross-functional collaboration.

From a technological standpoint, although the case company utilizes modern tools quite well in some areas, their full Group-level potential seems to remain unrealized due to process and people-related bottlenecks. On the organizational level, fragmentation, a lack of top-down support, and subpar leadership commitment to harmonized data governance curb improvement.

The key challenge here lies in developing the organization's data governance maturity. Effective data governance requires not only technical solutions, but also sustained managerial attention, coherent structures, and cultural transformation. As Díaz et al. (2018) and Carillo (2017) suggest, developing a true data culture requires extending analytics capabilities from technical experts to managerial roles and cultivating data-driven practices at both grassroots and executive levels. Without these elements firmly in place, the promise of DCs, DSAs and any similar initiatives will remain unfulfilled.

5.3 Addressing the Research Problem and Aims

This study set out to explore the value that Data Contracts and Data Sharing Agreements can provide, as well as the conditions under which this value can be effectively realized. The findings suggest that while the theoretical and practical benefits of these tools are widely acknowledged, their successful implementation is highly dependent on the organization's broader data governance maturity and its readiness for cultural and structural change.

With regard to the first research question — What value do Data Contracts and Data Sharing Agreements provide? — the interview findings illustrate that these tools serve multiple important functions. At their core, Data Contracts help clarify ownership and accountability, which are otherwise often absent or ambiguous. They provide a formalized structure that enforces responsibility for specific data assets, reducing the risk of neglect or misuse. This clarity supports not only more reliable data quality but also facilitates traceability and transparency in data flows. Moreover, Data Contracts can function as coordination tools that bridge organizational silos by technically defining expectations and interfaces between data producers and consumers. In doing so, they support more efficient collaboration, reduce misunderstandings and mishandlings, and improve trust in shared data.

Similarly, Data Sharing Agreements offer a framework for managing risks related to privacy, security, and compliance, while enabling broader use and democratization of data. These arrangements establish shared rules and standards that allow data to be reused across business areas and systems without compromising regulatory or ethical obligations. Both instruments therefore contribute to clarity, cooperation, coordination, standardization, process refinement, risk management, and more effective utilization of organizational data resources.

However, the second research question — How can this value be realized? — deals with more complex dynamics. The realization of value from Data Contracts and Data Sharing Agreements is not automatic, but contingent on a variety of organizational and contextual factors. Foremost among these is the presence of clear roles and responsibilities. Without established data ownership and well-defined governance structures, the introduction of formal contracts may lead to "useless" additional work in the form of superficial compliance, without enacting any meaningful change. The value of these instruments also depends on whether they are perceived as helpful enablers or as bureaucratic burdens. In this case study, some interviewees expressed concern that DCs and DSAs could be too rigid or theoretical, particularly if introduced without adequate communication, training, or alignment with day-to-day practices.

Realizing the value of Data Contracts and Data Sharing Agreements therefore requires a carefully thought-out implementation plan that accounts for the organization's overall maturity and the current imbalances between different business areas' maturity levels. Leadership support plays a critical role in ensuring that governance efforts are prioritized and resourced appropriately. Without sustained managerial commitment, the "invisible" but essential work of governance risks being deprioritized in favor of short-term operational pressures.

Furthermore, the standardization brought by Data Contracts and Data Sharing Agreements must be balanced with the need for contextual flexibility. Harmonization efforts are most successful when they are not mandated as one-size-fits-all obligations, but as frameworks that offer guidance while allowing adaptation to local conditions. Finally, the implementation of these tools must be supported by solid technical infrastructure and interoperable systems, but the technology alone cannot compensate for deficiencies in governance processes or organizational culture.

Data Contracts and Data Sharing Agreements provide significant potential value in terms of clarifying responsibilities, reducing risk, enhancing collaboration, and enabling more strategic use of data. However, this value can only be fully realized when utilized within a broader organizational effort that addresses foundational governance gaps, supports cultural change, and prioritizes long-term capacity building over short-term advancements and hype.

5.4 Managerial and Practical Implications

Based on the relevant literature, interviews, and analysis, a number of managerial recommendations and practical implications can be drawn.

Clarifying roles, rights and responsibilities

Clear accountability structures must be established both for data ownership and for higher-level data leadership. Since the organization aspires to implement more shared practices, leadership with an adequate mandate should be made responsible for steering both discussion and practice in this direction. These leadership roles should also include built-in incentives for “invisible” but essential work, such as documentation, adherence to shared goals, and other foundational tasks that have often been undervalued. Sustained commitment from upper-level leadership is required to ensure that foundational governance work receives the visibility, prioritization, and resources it needs.

Additionally, access to any company data must be systematically engineered through transparent controls and clearly articulated rules, rather than left to assumptions of goodwill. If an employee in a certain role needs access to relevant data, it has to be granted based on the rights and responsibilities defined to the role in question. Concerns that such role-based access models could hinder data democratization can be mitigated by prioritizing openness and usability in governance design. For example, applying lightweight approval processes for low-risk datasets, providing standardized templates, or maintaining clear metadata and discovery tools can aid in enabling control and accessibility simultaneously.

Prioritizing groundwork

The principle of "garbage in, garbage out" proves its validity quite well in this case study. Without a reliable data foundation, advanced initiatives such as artificial intelligence will underperform. These advancements necessitate starting with foundational work, such as clarifying ownership, improving documentation practices, and raising awareness about the benefits of structured data governance.

Hype around new technologies will not materialize if data remains in unsuitable formats, such as fragmented data lakes, unstructured sources, or datasets without metadata. Fixing this issue requires more appreciation and resourcing for the "boring" but essential work of documentation, quality assurance, metadata management and infrastructure maintenance. Groundwork must be prioritized before more complex solutions can be adopted.

Investing in Education

Raising data literacy across the organization is critical. This applies not only to those working directly with data, but also to decision-makers who may have previously relied on intuition or old ways of working rather than data. Targeted training can encourage a shift away from "gut feeling" decision-making toward more data-driven practices. Investing in awareness and training is essential, as many of the concerns considered in this study stem from limited understanding. Education and training are essential in reducing resistance and fostering a shared understanding of why better ways of working matter.

Making use of differences between business areas

This study highlights a paradox across business areas. Finance and Controlling and Energy are naturally data-driven and thus more advanced in certain practices, but they remain burdened by legacy systems and outdated processes that no longer fully serve their needs. Fibres, by contrast, reports lower overall data maturity but is more flexible, allowing it to adopt newer practices from the outset.

A cross-organizational approach could turn these differences into an advantage. Areas weighed down by legacy systems can share lessons on what has not worked, while areas with fewer constraints can experiment with new methods and avoid repeating earlier mistakes. Likewise, less mature areas can benefit from the experience and structures already developed in more data-intensive units. Recognizing and harnessing both perspectives, the experience of long-established functions and the flexibility of newer ones, could accelerate the streamlining of practices across the organization. If successful, such an approach would not only strengthen internal efficiency but also provide the group with a competitive edge.

Reassessing tools and costs

Rather than striving for perfect execution, maximum speed, and the deployment of shiny new technologies, the organization would benefit from focusing on foundational improvements, such as clarifying data ownership and identifying responsible contact points. Given the high costs of licenses and new platforms, the organization should evaluate which functionalities are truly necessary and which tools are underutilized. Rather than continuously acquiring new tools to manage, it may be more effective to dedicate time and resources to strengthening foundations and ensuring existing systems are used to their full potential. Only after these basics are solid, additional investments in expensive tools are justified.

6 Conclusions

This thesis set out to find the value associated with Data Contracts and Data Sharing Agreements in a case company. Through semi-structured interviews and thematic analysis, the findings suggest that the value in DCs and DSAs requires thorough groundwork, improvements in governance and a stronger data culture in order to be realized.

As a whole, the participants of this Thesis were optimistic yet realistic about the promise of DCs and DSAs. Many saw how they could be transformative in data work, and even where they saw challenges, were able to suggest remedial measures. These visions, fears and corrective ideas were well in line with established literature. However, while most participants felt they could easily name the possible remedies, not as many were optimistic about the case company's capability or willingness to invest in these improvements. This results in participants feeling less inclined to strive for these improvements, as their efforts might be in vain.

The suggestions presented in this Thesis reflect how the participants themselves saw potential risks mitigated and are in line with both industry and academic literature. Advice and guidance to build new capabilities slowly and meticulously, focusing on groundwork and reassessing current situations are commonplace, but much easier said than done.

6.1 Limitations

As the first limitation to the value of this research comes from its nature as a Master's Thesis study. The limited questionnaire and small sample size of participants is in line with the depth of a Master's Thesis but means decisive conclusions cannot be drawn from this study alone, even with the findings of this study aligning well with previous work on the field. Additionally, as a single-case study, this Thesis is not generalizable across other cases, industries or companies. While this Thesis covers some cross-comparison between the selected UPM Business Areas and functions, the generalizability would have been greater if the scope had allowed for comparison across company or industry borders.

While it can be very fruitful to look into the first steps of implementing a new process and how change moves people and systems, it does not offer the visibility into actual practice once the dust has settled. Similarly, while the people interviewed for this thesis provide their experiences in the implementation stage and change management around it, their thoughts cannot be thought to represent a fully mature state, where their hopes and fears might not ever realize. Understanding the fears stakeholders have can help in avoiding these identified pitfalls, but others might emerge more prominently where the stakeholders' vision did not see them emerging. Likewise, not all the hopes and improvements named by these interviewees necessarily come true, and they might in these early stages be unable to identify perks that come along at a more mature point in the process.

As the last limitation, the researcher's position within the company might have affected some participants' comfortability in sharing their frustrations, even though efforts were made to mitigate this risk.

6.2 Suggestions for Future Research

As Data Contracts and Data Sharing Agreements are relative newcomers in the data world, not much academical work has been made with them. However, out of these two, Data Contracts remain even less formally studied. The gap in the research is evident here: more academical studies should be conducted into the origins, current use cases as well as potential gains and losses related to employing Data Contracts.

Conducting comparative studies between this thesis and other similar case studies on DCs and DSAs could mitigate the low generalizability of a study of this limited scope. Additionally, a more longitudinal research approach could follow and more concretely capture the process of implementing DCs and DSAs, asses their practicality, and see how well these initial expectations compare to post-implementation experiences. While the industry reports provide valuable insights and case examples, they do not hold the same academical rigor and value as peer-reviewed, academically sound research papers.

References

Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. International Journal of Information Management, 49, 424–438. https://doi.org/10.1016/J.IJINFOMGT.2019.07.008

Achanta, A. (2023). Data Democratization: Empowering Non-Technical Users with Self-Service BI Tools and Techniques to Access and Analyze Data Without Heavy Reliance on IT Teams. International Journal of Computer Trends and Technology, 71(8). https://doi.org/https://doi.org/10.14445/22312803/IJCTT-V71I8P106

Ahlawat, P., Borgman, J., Eden, S., Huels, S., Iandiorio, J., Kumar, A., & Zakahi, P. (2023). A New Architecture to Manage Data Costs and Complexity. https://www.bcg.com/publications/2023/new-data-architectures-can-help-manage-data-costs-and-complexity

Alhassan, I., Sammon, D., & Daly, M. (2016). Data governance activities: an analysis of the literature. Journal of Decision Systems, 25(s1), 64–75. https://doi.org/10.1080/12460125.2016.1187397;WGROUP:STRING:PUBLICATION

Allen, C., des Jardins, T. R., Heider, A., Lyman, K. A., McWilliams, L., Rein, A. L., Schachter, A. A., Singh, R., Sorondo, B., Topper, J., & Turske, S. A. (2014). Data governance and data sharing agreements for community-wide health information exchange: lessons from the beacon communities. EGEMS (Washington, DC), 2(1), 5. https://doi.org/10.13063/2327-9214.1057

Alliance for Integrity. (2018). COMPLIANCE AND DIGITALISATION - HOW TECHNOLOGY CAN FOSTER TRANSPARENCY IN AFRICAN COUNTRIES.

Antal, C., Cioara, T., Anghel, I., Antal, M., & Salomie, I. (2021). Distributed Ledger Technology Review and Decentralized Applications Development Guidelines. Future Internet 2021, Vol. 13, Page 62, 13(3), 62. https://doi.org/10.3390/FI13030062

Araújo Machado, I., Costa, C., & Santos, M. Y. (2022). Advancing Data Architectures with Data Mesh Implementations. Lecture Notes in Business Information Processing, 452, 10–18. https://doi.org/10.1007/978-3-031-07481-3_2

Atlan. (2024). Data Contracts 101: Importance, Validations & Best Practices. https://atlan.com/data-contracts/

Batistič, S., & van der Laken, P. (2019). History, Evolution and Future of Big Data and Analytics: A Bibliometric Analysis of Its Relationship to Performance in Organizations. British Journal of Management, 30(2), 229–251. https://doi.org/10.1111/1467-8551.12340

Bellin, E., Fletcher, D. D., Geberer, N., Islam, S., & Srivastava, N. (2010). Democratizing information creation from health care data for quality improvement, research, and education-The montefiore medical center experience. Academic Medicine, 85(8), 1362–1368. https://doi.org/10.1097/ACM.0B013E3181DF0F3B,

Bhardwaj, R., & Datta, D. (2020). Consensus Algorithm. Studies in Big Data, 71, 91–107. https://doi.org/10.1007/978-3-030-38677-1_5

Blinowski, G., Ojdowska, A., & Przybylek, A. (2022). Monolithic vs. Microservice Architecture: A Performance and Scalability Evaluation. IEEE Access, 10, 20357–20374. https://doi.org/10.1109/ACCESS.2022.3152803

Blohm, I., Wortmann, F., Legner, C., & Köbler, F. (2024). Data products, data mesh, and data fabric: New paradigm(s) for data and analytics? Business and Information Systems Engineering, 66(5), 643–652. https://doi.org/10.1007/S12599-024-00876-5/TABLES/3

Bode, J., Kuhl, N., Kreuzberger, D., & Holtmann, C. (2023). Towards Avoiding the Data Mess: Industry Insights from Data Mesh Implementations. IEEE Access, 12, 95402–95416. https://doi.org/10.1109/ACCESS.2024.3417291

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101. https://doi.org/10.1191/1478088706QP063OA

Braun, V., & Clarke, V. (2012). Thematic analysis. In APA handbook of research methods in psychology, Vol. 2: Research designs: Quantitative, qualitative, neuropsychological, and biological (Vol. 2, pp. 57–71). American Psychological Association. https://www.researchgate.net/publication/269930410_Thematic_analysis

ANPD: Brazilian Data Protection Law LGPD. (2018). Retrieved September 17, 2025, from https://falabr.cgu.gov.br/

Broek, T. van den, & Veenstra, A. F. van. (2015). Modes of Governance in Inter-Organizational Data Collaborations. 0–12. https://doi.org/10.18151/7217509

Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating decision-making in data management activities: A systematic review of data governance principles. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 9820 LNCS, 115–125. https://doi.org/10.1007/978-3-319-44421-5_9

Buffenoir, E., & Bourdon, I. (2012). Reconciling complex organizations and data management: the Panopticon paradigm. Proceedings - Pacific Asia Conference on Information Systems, PACIS 2013. https://arxiv.org/abs/1210.6800v2

Carillo, K. D. A. (2017). Let's stop trying to be "sexy" – preparing managers for the (big) data-driven business era. Business Process Management Journal, 23(3), 598–622. https://doi.org/10.1108/BPMJ-09-2016-0188/FULL/XML

Carmi, E., Yates, S. J., Lockley, E., & Pawluczuk, A. (2020). Data citizenship: Rethinking data literacy in the age of disinformation, misinformation, and malinformation. Internet Policy Review, 9(2), 1–22. https://doi.org/10.14763/2020.2.1481

Caserta, J., Dubois, J.-B., Roggendorf, M., Roth, M., & Srinidhi, N. (2023). Demystifying data mesh. McKinsey . https://www.mckinsey.com/capabilities/quantumblack/our-insights/demystifying-data-mesh

Chauhan, K., Ghawate, M., Joshi, S., & Kawade, S. (2024). Tracing the Impact of GDPR on Global Data Privacy. International Journal of Science and Research (IJSR), 13(9), 406–410. https://doi.org/10.21275/SR24906110537

Cisco. (2023). Global AI Readiness Index—Intentions Outpacing Abilities. https://www.cisco.com/c/dam/m/en_us/solutions/ai/readiness-index/documents/cisco-global-ai-readiness-index.pdf

Cockburn, A. (2005). The Hexagonal (Ports & Adapters) Architecture. https://alistair.cockburn.us/hexagonal-architecture

Columbus, L. (2018, January 7). 83% Of Enterprise Workloads Will Be In The Cloud By 2020. Forbes. https://www.forbes.com/sites/louiscolumbus/2018/01/07/83-of-enterprise-workloads-will-be-in-the-cloud-by-2020/

Costantino, G., Martinelli, F., Matteucci, I., & Petrocchi, M. (2017). Analysis of data sharing agreements. ICISSP 2017 - Proceedings of the 3rd International Conference on Information Systems Security and Privacy, 2017-January, 167–178. https://doi.org/10.5220/0006207501670178

Crown Commercial Service. (n.d.). Open Contracting Data Standard Guide to Implementation on Contracts Finder. Retrieved May 22, 2025, from https://data.gov.uk/publisher/crown-commercial-service

Cutler, E. (2024). Evolution of Data Governance When Implementing Data Democratization.

Dahlberg, T., & Nokkala, T. (2019). Willingness to share supply chain data in an ecosystem governed platform - An interview study. 32nd Bled EConference Humanizing Technology for a Sustainable Society, BLED 2019 - Conference Proceedings, 619–638. https://doi.org/10.18690/978-961-286-280-0.33

Data mesh - core ideas and benefits. (2021). Futurice Technology. https://www.futurice.com/fi/blog/when-should-organizations-consider-data-mesh

Data protection adequacy for non-EU countries. (n.d.). European Comission. Retrieved December 20, 2024, from https://commission.europa.eu/law/law-topic/data-protection/international-dimension-data-protection/adequacy-decisions_en

Data protection under GDPR . (2016). European Union. Retrieved May 14, 2025, from https://europa.eu/youreurope/business/dealing-with-customers/data-protection/data-protection-gdpr/

Davenport, T. H., & Kudyba, S. (2016). Designing and Developing Analytics-Based Data Products. MIT Sloan Management Review. https://sloanreview.mit.edu/article/designing-and-developing-analytics-based-data-products/

dbt Labs Developer Hub. (2025a). Model contracts. Dbt Labs. https://docs.getdbt.com/docs/mesh/govern/model-contracts

dbt Labs Developer Hub. (2025b). contract . Dbt Labs. https://docs.getdbt.com/reference/resource-configs/contract

dbt Labs. (2020). JetBlue eliminates data engineering bottlenecks with dbt. Retrieved September 17, 2025, from https://www.getdbt.com/case-studies/jetblue

de Andrade, P. R. M., Albuquerque, A. B., Teófilo, W. D., & da Silva, F. A. (2016). Change management: Implementation and benefits of the change control in the information technology enviroment. International Journal of Advanced Information Technology, 6(1), 23–33. https://doi.org/10.5121/ijait.2016.6102

de Lauretis, L. (2019). From monolithic architecture to microservices architecture. Proceedings - 2019 IEEE 30th International Symposium on Software Reliability Engineering Workshops, ISSREW 2019, 93–96. https://doi.org/10.1109/ISSREW.2019.00050

Dehghani, Z. (2022). Data Mesh: Delivering Data-Driven Value at Scale. O'Reilly Media, 2013–2015. https://www.oreilly.com/library/view/data-mesh/9781492092384/

Dehghani, Z. (2019). How to Move Beyond a Monolithic Data Lake to a Distributed Data Mesh. https://martinfowler.com/articles/data-monolith-to-mesh.html

Deshpande, M. (2023). Mastering Data Product Management: Paving the way for a Data-Driven Enterprise . Journal of Artificial Intelligence, Machine Learning and Data Science, 2023(3), 348–353. https://doi.org/10.51219/JAIMLD/Mahesh-deshpande/105

Dethine, B., Enjolras, M., & Monticolo, D. (2020). Digitalization and SMEs' export management: Impacts on resources and capabilities. Technology Innovation Management Review, 10(4), 18–34. https://doi.org/10.22215/TIMREVIEW/1344

Díaz, A., Rowshankish, K., & Saleh, T. (2018). Why data culture matters. McKinsey Quarterly.

Dong, X. L., & Rekatsinas, T. (2018). Data Integration and Machine Learning: A Natural Synergy. 11(12), 2094–2097. https://doi.org/10.14778/3229863.3229876

Duggineni. (2023). Impact of controls on data integrity and information systems. Science and Technology 2023, 13(2): 29-35. https://doi.org/10.5923/j.scit.20231302.04

Eastwood, B. (2021). The case for building a data-sharing culture in your company. MIT Sloan. https://mitsloan.mit.edu/ideas-made-to-matter/case-building-a-data-sharing-culture-your-company

Ekiye, A. E., Ahmadi, H., & Iwendi, C. (2023). CIA in data security. In ICACTCE23 - International Conference on Advances in Communication Technology and Computer Engineering . https://ub-ir.bolton.ac.uk/esploro/outputs/conferencePaper/CIA-in-data-security/999411908841

Endapally, S. (2023). Who Is Responsible for an Organization's FAIR Data? TWDI. https://tdwi.org/articles/2023/01/17/diq-all-who-is-responsible-for-fair-data.aspx

Eriksson, P., & Kovalainen, A. (2008). Qualitative Methods in Business Research. Qualitative Methods in Business Research. https://doi.org/10.4135/9780857028044

Fassnacht, M., Benz, C., Leimstoll, J., & Satzger, G. (2023). Is Your Organization Ready to Share? A Framework of Beneficial Conditions for Data Sharing. Proceedings of the 44th International Conference on Information Systems (ICIS). https://www.researchgate.net/publication/374422448_Is_Your_Organization_Ready_to_Share_A_Framework_of_Beneficial_Conditions_for_Data_Sharing

Foote, K. D. (2023, March 30). Developing a Strategy to Democratize Data. Dataversity. https://www.dataversity.net/articles/developing-a-strategy-to-democratize-data/

Frenté, S., & Ladley, J. (2023). Data Governance Framework: 5 Key Elements. Dataversity. https://www.dataversity.net/5-key-elements-of-a-data-governance-framework/

Fu, Q., Nicholson, G. L., & Easton, J. M. (2024). Understanding data quality in a data-driven industry context: Insights from the fundamentals. Journal of Industrial Information Integration, 42, 100729. https://doi.org/10.1016/J.JII.2024.100729

Gade. (2021). Data Analytics: Data Democratization and Self-Service Analytics Platforms Empowering Everyone with Data. MZ Computing Journal.

Gao, T., Dontcheva, M., Adar, E., Liu, Z., & Karahalios, K. (2015). Datatone: Managing ambiguity in natural language interfaces for data visualization. UIST 2015 - Proceedings of the 28th Annual ACM Symposium on User Interface Software and Technology, 489–500. https://doi.org/10.1145/2807442.2807478/SUPPL_FILE/P489.MP4

Given, L. (2008). The SAGE Encyclopedia of Qualitative Research Methods. In The SAGE Encyclopedia of Qualitative Research Methods. SAGE Publications, Inc. https://doi.org/10.4135/9781412963909

Goswami, P., Faujdar, N., Debnath, S., Khan, A. K., & Singh, G. (2024). Investigation on storage level data integrity strategies in cloud computing: classification, security obstructions, challenges and vulnerability. Journal of Cloud Computing, 13(1), 1–23. https://doi.org/10.1186/S13677-024-00605-Z/TABLES/8

Grande, D., Machado, J., Petzold, B., & Roth, M. (2020, July 31). Reducing data costs without sacrificing growth. McKinsey. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/reducing-data-costs-without-jeopardizing-growth

Grossman, R. L., & Siegel, K. P. (2014). Organizational Models for Big Data and Analytics. Journal of Organization Design, 3(1), 20–25. https://doi.org/10.7146/jod.9799

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

Gupte, A. (2023, June 14). What Good Data Product Managers Do — And Why You Probably Need One. Monte Carlo. https://www.montecarlodata.com/blog-what-good-data-product-managers-do-and-why-you-probably-need-one/

Haneem, F., Ali, R., Kama, N., & Basri, S. (2017). Resolving data duplication, inaccuracy and inconsistency issues using Master Data Management. International Conference on Research and Innovation in Information Systems, ICRIIS. https://doi.org/10.1109/ICRIIS.2017.8002453

Hasan, R., & Legner, C. (2023). Understanding data products: motivations, definition, and categories. 31st European Conference on Information Systems (ECIS).

Hassani, H., Huang, X., & MacFeely, S. (2025). Mapping the evolution of data governance scientific research. Data & Policy, 7, e51. https://doi.org/10.1017/DAP.2025.10014

Hiatt, J. (2006). ADKAR: A model for change in business, government, and our community. . Prosci Learning Center Publications.

Hosseini, A., Emami, H., Sadat, Y., & Paydar, S. (2023). Integrated personal health record (PHR) security: requirements and mechanisms. BMC Medical Informatics and Decision Making, 23(1), 1–7. https://doi.org/10.1186/S12911-023-02225-0/TABLES/2

How, M. (2020). The Role of the Data Contract. The Modern Data Warehouse in Azure, 163–180. https://doi.org/10.1007/978-1-4842-5823-1_6

Howard, H., Alder, F., Ashton, E., Chamayou, A., Clebsch, S., Costa, M., Delignat-Lavaud, A., Fournet, C., Jeffery, A., Kerner, M., Kounelis, F., Kuppe, M. A., Maffre, J., Russinovich, M., & Wintersteiger, C. M. (2023). Confidential Consortium Framework: Secure Multiparty Applications with Confidentiality, Integrity, and High Availability. Proceedings of the VLDB Endowment, 17(2), 225–240. https://doi.org/10.14778/3626292.3626304

Huang, G., He, J., Chi, C. H., Zhou, W., & Zhang, Y. (2015). A Data as a Product Model for Future Consumption of Big Stream Data in Clouds. Proceedings - 2015 IEEE International Conference on Services Computing, SCC 2015, 256–263. https://doi.org/10.1109/SCC.2015.43

Huylebrouck, D. (2019). Missing Link. 153–166. https://doi.org/10.1007/978-3-030-04037-6_9

Hyun, Y., Kamioka, T., & Hosoya, R. (2020). Improving Agility Using Big Data Analytics: The Role of Democratization Culture. Pacific Asia Journal of the Association for Information Systems, 12(2), 2. https://doi.org/10.17705/1pais.12202

Irwin, L. (2018, April 5). How NIST's Cybersecurity Framework Protects the CIA Triad. IT Governance USA. https://www.itgovernanceusa.com/blog/how-nist-can-protect-the-cia-triad-including-the-often-overlooked-i-integrity

Jacobsen, A., Azevedo, R. de M., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., … Schultes, E. (2020). FAIR Principles: Interpretations and Implementation Considerations. Data Intelligence, 2(1–2), 10–29. https://doi.org/10.1162/DINT_R_00024

Jahnke, N., & Otto, B. (2023). Data Catalogs in the Enterprise: Applications and Integration. Datenbank-Spektrum 2023 23:2, 23(2), 89–96. https://doi.org/10.1007/S13222-023-00445-2

Janssen, M., Brous, P., Estevez, E., Barbosa, L. S., & Janowski, T. (2020). Data governance: Organizing data for trustworthy Artificial Intelligence. Government Information Quarterly, 37(3), 101493. https://doi.org/10.1016/J.GIQ.2020.101493

JetBlue. (2020). How JetBlue became a data-driven airline using dbt. https://www.youtube.com/watch?v=MlQGHR5bvRI

Datacontract.Com. Data Contract Specification – Data contracts bring data providers and data consumers together. Retrieved December 10, 2024, from https://datacontract.com/

Ravva, Karthik. (2025). Comparative Analysis of Centralized vs. Decentralized Governance Models for AI-BI in Multi-Cloud Enterprises. Journal of Computer Science and Technology Studies, 7(6), 873–879. https://doi.org/10.32996/JCSTS.2025.7.103

Khatri, V., & Brown, C. v. (2010). Designing data governance. Communications of the ACM, 53(1), 148–152. https://doi.org/10.1145/1629175.1629210

Khatun, J. (2024). Understanding Data Contracts. https://doi.org/10.36227/techrxiv.171779368.80821952/v1

Klosowski, T. (2021, September 6). The State of Consumer Data Privacy Laws in the US (And Why It Matters) | Wirecutter. The New York Times Wirecutter. https://www.nytimes.com/wirecutter/blog/state-of-privacy-laws-in-us/

Kostic, N. (2024). Service Level Agreement: Definition, Example, Best Practices. https://phoenixnap.com/blog/service-level-agreement

Kotter, J. P. (2006). Leading Change Why Transformation Efforts Fail Leaders who successfully transform businesses do eight things right (and they do them in the right order). www.hbrreprints.org

Lebens, M., J Finnegan, R., C Sorsen, S., & Shah, J. (2021). Rise of the Citizen Developer. Muma Business Review, 5, 101–111. https://doi.org/10.28945/4885

Lefebvre, H., Legner, C., & Fadler, M. (2021). Data democratization: toward a deeper understanding. Forty-Second International Conference on Information Systems. https://www.researchgate.net/profile/Christine-Legner-2/publication/354906721_Data_democratization_toward_a_deeper_understanding/links/61539051fd7b3d1215599a07/Data-democratization-toward-a-deeper-understanding.pdf

Lewis, J., & Fowler, M. (2014, March). Microservices. https://www.martinfowler.com/articles/microservices.html

Lin, D., Lingappa, G., & Aswani, J. (2019, March 25). Building and Scaling Data Lineage at Netflix to Improve Data Infrastructure Reliability, and Efficiency. Netflix Tech Blog. https://netflixtechblog.com/building-and-scaling-data-lineage-at-netflix-to-improve-data-infrastructure-reliability-and-1a52526a7977

Liu, K., Desai, H., Kagal, L., & Kantarcioglu, M. (2018). Enforceable Data Sharing Agreements Using Smart Contracts. https://arxiv.org/pdf/1804.10645

Loukiala, A., Joutsenlahti, J. P., Raatikainen, M., Mikkonen, T., & Lehtonen, T. (2021). Migrating from a Centralized Data Warehouse to a Decentralized Data Platform Architecture. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13126 LNCS, 36–48. https://doi.org/10.1007/978-3-030-91452-3_3

Loukides, M. (2011, September 15). The evolution of data products . O'Reilly Radar. https://www.oreilly.com/radar/evolution-of-data-products/

Luis, B., Mei, D., & Elly, B. (2024). Data Governance Strategies for Ensuring Compliance with Data Protection Regulations: Discussing effective data governance strategies for managing data protection and compliance in business analytics. https://www.researchgate.net/publication/386409954_Data_Governance_Strategies_for_Ensuring_Compliance_with_Data_Protection_Regulations_Discussing_effective_data_governance_strategies_for_managing_data_protection_and_compliance_in_business_analytics

Lundgren, B., & Möller, N. (2017). Defining Information Security. Science and Engineering Ethics, 25(2), 419. https://doi.org/10.1007/S11948-017-9992-1

Mahamoud Hasan, M., Aparisi-Torrijo, S., & González-Ladrón-de-Guevara, F. (2025). Change management and organizational performance: current key trends. Cogent Business & Management, 12(1), 2478447. https://doi.org/10.1080/23311975.2025.2478447

Makhecha, J., & Harker, D. (2024). Information Control & the Need-to-Know principle. Deloitte UK. https://www.deloitte.com/uk/en/Industries/financial-services/blogs/information-control-and-the-need-to-know-principle.html

Matteucci, I., Petrocchi, M., Sbodio, M. L., & Wiegand, L. (2012). A Design Phase for Data Sharing Agreements. Lecture Notes in Computer Science (Including Sub-series Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7122 LNCS, 25–41. https://doi.org/10.1007/978-3-642-28879-1_3

McKinsey. (2022) The data-driven enterprise of 2025. https://www.mckinsey.com/capabilities/mckinsey-digital/our-insights/the-data-driven-enterprise-of-2025

Meierhofer, J., Stadelmann, T., & Cieliebak, M. (2019). Data Products. In Applied Data Science: Lessons Learned for the Data-Driven Business (pp. 47–61). Springer, Cham. https://doi.org/10.1007/978-3-030-11821-1_4

Meyer, B. (1992). Applying "Design by Contract."

Moses, J. (2023). MICROSERVICES, ETHICS, AND INNOVATION: SOLV-INGTHE PUZZLE OF SCALABLE ARCHITECTURE. https://www.researchgate.net/profile/Joshua-Moses-10/publication/393465571_MICROSERVICES_ETHICS_AND_INNOVATION_SOLVING_THE_PUZZLE_OF_SCALABLE_ARCHITECTURE/links/686bf59b07b3253fd1ccd16f/MICROSERVICES-ETHICS-AND-INNOVATION-SOLVING-THE-PUZZLE-OF-SCALABLE-ARCHITECTURE.pdf

Mucci, T., & Stryker, C. (2024, February 23). What Is Data as a Product (DaaP)? IBM. https://www.ibm.com/think/topics/data-as-a-product

Naganathan, Venkatesh (2018). Comparative Analysis of Big Data, Big Data Analytics: Challenges and Trends. International Research Journal of Engineering and Technology (IRJET), 05(05).

Nicolazzo, S., Nocera, A., & Pedrycz, W. (2024). SERVICE LEVEL AGREEMENTS AND SECURITY SLA: A COMPREHENSIVE SURVEY.

Nielsen, O. B., Persson, J. S., & Madsen, S. (2019). Why governing data is difficult: Findings from Danish local government. IFIP Advances in Information and

Communication Technology, 533, 15–29. https://doi.org/10.1007/978-3-030-04315-5_2

Nii Laryeafio, M., & Ogbewe, O. C. (2023). Ethical consideration dilemma: systematic review of ethics in qualitative data collection through interviews. Journal of Ethics in Entrepreneurship and Technology, 3(2), 94–110. https://doi.org/10.1108/JEET-09-2022-0014

Oliveira, B., Duarte, A., & Oliveira, Ó. (2024). Towards a Data Catalog for Data Analytics. Procedia Computer Science, 237, 691–700. https://doi.org/10.1016/J.PROCS.2024.05.155

Open Contracting Data Standard 1.1.5 documentation. (n.d.). Open Contracting Data Standard. Retrieved May 22, 2025, from https://standard.open-contracting.org/latest/en/

Osarenren, P. (2024). Understanding Schema Drift -- Causes, Impact & Solutions. https://www.acceldata.io/blog/schema-drift

Otto, B. (2011). Data governance. Business and Information Systems Engineering, 3(4), 241–244. https://doi.org/10.1007/S12599-011-0162-8

Pal, S. (2024). How Data Mesh Helps Your Organization Be Data-Driven and Achieve Velocity. TWDI. https://tdwi.org/articles/2024/02/14/arch-all-how-data-mesh-helps-your-organization-be-data-driven-and-achieve-velocity-3.aspx

Pappas, I. O., Mikalef, P., Giannakos, M. N., Krogstie, J., & Lekakos, G. (2018). Big data and business analytics ecosystems: paving the way towards digital transformation and sustainable societies. Information Systems and E-Business Management, 16(3), 479–491. https://doi.org/10.1007/S10257-018-0377-Z/FIGURES/1

Parker, D. B. (1998). Fighting Computer Crime: A New Framework for Protecting Information. In Computer Security Handbook. John Wiley & Sons. https://books.google.com/books/about/Fighting_Computer_Crime.html?hl=pl&id=k_u_QgAACAAJ

Paulus, E., & Fauzan, M. A. (2018). A Framework to Ensure Data Integrity and Safety. Data Science: Journal of Computing and Applied Informatics, 2(1), 1–11. https://doi.org/10.32734/JOCAI.V2.I1-90

Perrin, J.-G. (2022). The next generation of Data Platforms is the Data Mesh. The PayPal Technology Blog. https://medium.com/paypal-tech/the-next-generation-of-data-platforms-is-the-data-mesh-b7df4b825522

Petersen, C. (2019). User-focused data sharing agreements: a foundation for the genomic future. JAMIA Open, 2(4), 402–406. https://doi.org/10.1093/JAMIAOPEN/OOZ043

Pierce, E. M. (2005). What's in your Information Product Inventory? . In R. Y. , Wang, E. M. Pierce, S. E. Madnick, & C. W. Fisher (Eds.), Information Quality (1st ed., pp. 99–114). Routledge. https://www.taylorfrancis.com/chapters/edit/10.4324/9781315703480-10/information-product-inventory-elizabeth-pierce

PIPC Personal Information Protection Commission of South Korea. (2023, September 15). Amended Personal Information Protection Act (PIPA) and its Enforcement Decree Become Effective. PIPC, Korea, GPA, 2025 GPA, GPA Seoul, 2025 GPA Seoul, AI, Data, Privacy, GPA 서울, Global Privacy Assembly. https://pipc.go.kr/eng/user/ltn/new/noticeDetail.do?bbsId=BBSMSTR_000000000%20001&nttId=2331

Prudhvi. (2025). Data contracts in the wild: An approach for redefining trust and accountability in modern data pipelines. World Journal of Advanced Research and Reviews, 26(3), 393–399. https://doi.org/10.30574/WJARR.2025.26.3.2166

Ransbotham, S., Khodabandeh, S., Fehling, R., LaFountain, B., & Kiron, D. (2019). Winning With AI. MIT Sloan Management Review. https://sloanreview.mit.edu/projects/winning-with-ai/

Redman, T. C. (2013a). Are You Data Driven? Take a Hard Look in the Mirror. Harvard Business Review. https://hbr.org/2013/07/are-you-data-driven-take-a-har

Redman, T. C. (2013b). Become More Data-Driven by Breaking These Bad Habits. Harvard Business Review. https://hbr.org/2013/08/becoming-data-driven-breaking

Richter Gade, I., Penagos, N., & Brown, S. (2020). Open Contracting Reforms in Colombia. In R. Bajpai & C. B. Myers (Eds.), Enhancing Government Effectiveness and Transparency : The Fight Against Corruption (Vol. 1 of 2). World Bank Group. https://thedocs.worldbank.org/en/doc/406451611783796514-0090022021/original/OpenContractingReformsinColombia.pdf

Rokis, K., & Kirikova, M. (2022). Challenges of Low-Code/No-Code Software Development: A Literature Review. Lecture Notes in Business Information Processing, 462 LNBIP, 3–17. https://doi.org/10.1007/978-3-031-16947-2_1

Sadiq, S., & Indulska, M. (2017). Open data: Quality over quantity. International Journal of Information Management, 37(3), 150–154. https://doi.org/10.1016/J.IJINFOMGT.2017.01.003

Sallam, R. (2019). How Augmented Analytics Will Transform Your Organization. A Gartner Trend Insight Report. Retrieved September 18, 2025, from https://emt.gartnerweb.com/ngw/globalassets/en/doc/documents/441420-how-augmented-analytics-will-transform-your-organization.pdf

Saltzer, J. H., & Schroeder, M. D. (1975). The Protection of Information in Computer Systems. Proceedings of the IEEE, 63(9), 1278–1308.

Samarasinghe, S., & Lokuge, S. (2023). Data Democratization: Empowering Employees for Data-Driven Innovation. In Data-Driven Approaches for Effective Managerial Decision Making (pp. 155–183). IGI Global Scientific Publishing. https://doi.org/10.4018/978-1-6684-7568-3.CH008

Samonas, S., & Coss, D. (2014). THE CIA STRIKES BACK: REDEFINING CONFIDENTIALITY, INTEGRITY AND AVAILABILITY IN SECURITY. Journal of Information System Security, 10(3), 21–45. www.jissec.org

Sassoon, Y. (2023, November 22). What is, and what isn't, a data contract. https://datacreation.substack.com/p/what-is-and-what-isnt-a-data-contract

Scaria, E., Berghmans, A., Pont, M., & et al. (2018). Study on data sharing between companies in Europe : final report. https://doi.org/10.2759/354943

Segner, M. (2025). Data Contracts: How They Work, Importance, & Best Practices. Monte Carlo. https://www.montecarlodata.com/blog-data-contracts-explained/

Seiner, R. (2020, July 23). The Non-Invasive Data Governance Approach. Alation. https://www.alation.com/blog/data-governance-non-invasive-approach/

Shaikh, R., & Sasikumar, M. (2015). Data Classification for Achieving Security in Cloud Computing. Procedia Computer Science, 45(C), 493–498. https://doi.org/10.1016/J.PROCS.2015.03.087

Shamsuddin, F., & Abdul Razak, A. Z. (2023). Development of a Model for Data-Driven Decision Making: Critical Skills for School Leaders. Malaysian Journal of Social Sciences and Humanities (MJSSH), 8(12). https://doi.org/10.47405/MJSSH.V8I12.2614

Shubladze, S. (2023, September 14). Unlocking The Power Of Data: Why Democratization Matters. Forbes. https://www.forbes.com/councils/forbestechcouncil/2023/09/14/unlocking-the-power-of-data-why-democratization-matters/

Skinner, M. (2022). What are data silos and how can you eliminate them? https://business.adobe.com/blog/basics/what-are-data-silos-and-how-can-you-eliminate-them

Snodgrass, E., & Soon, W. (2019). API practices and paradigms: Exploring the protocological parameters of APIs as key facilitators of sociotechnical forms of exchange. First Monday. https://doi.org/10.5210/FM.V24I2.9553

Stake, R. (1995). The art of case study. SAGE Publications. https://uk.sagepub.com/en-gb/eur/the-art-of-case-study-research/book4954

Thuermer, G., Walker, J., & Simperl, E. (2019). Data Sharing Toolkit: Lessons learned, resources and recommendations for sharing data. https://www.researchgate.net/publication/337635502_Data_Sharing_Toolkit_Lessons_learned_resources_and_recommendations_for_sharing_data

Uren, V., & Edwards, J. S. (2023). Technology readiness and the organizational journey towards AI adoption: An empirical study. International Journal of Information Management, 68, 102588. https://doi.org/10.1016/J.IJINFOMGT.2022.102588

Vestues, K., Hanssen, G. K., Mikalsen, M., Buan, T. A., & Conboy, K. (2022). Agile data management in NAV: A Case Study. Lecture Notes in Business Information Processing, 445 LNBIP, 220–235. https://doi.org/10.1007/978-3-031-08169-9_14

Vilminko-Heikkinen, R., & Pekkola, S. (2019). Changes in roles, responsibilities and ownership in organizing master data management. International Journal of Information Management, 47, 76–87. https://doi.org/10.1016/J.IJINFOMGT.2018.12.017

Wang, R. Y., Lee, Y. W., Pipino, L. L., & Strong, D. M. (1998). Manage Your Information as a Product. MIT Sloan Management Review, 39(4), 95–105.

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. Journal of Management Information Systems, 12(4), 5–34. https://doi.org/10.1080/07421222.1996.11518099;CTYPE:STRING:JOURNAL

Webber, K. L. (2020). Can You Trust Your Eyes? Learn How to Minimize Misinterpretation of Data Reports and Visualization. https://www.scup.org/wp-content/uploads/2023/07/SCUP_PHE_Can-You-Trust-202307-3278da-EDU.pdf

Weber, K., Otto, B., & Österle, H. (2009). Ones size does not fit all -A contingency approach to data governance. Journal of Data and Information Quality, 1(1). https://doi.org/10.1145/1515693.1515696

Weill, P., & Ross, J. W. (2004). It Governance on One Page. SSRN Electronic Journal. https://doi.org/10.2139/SSRN.664612

Wende, K., & Otto, B. (2007). A Contingency Approach to Data Governance. https://www.researchgate.net/publication/44939108_A_Contingency_Approach_to_Data_Governance

Wider, A., Verma, S., & Akhtar, A. (2023). Decentralized Data Governance as Part of a Data Mesh Platform: Concepts and Approaches. Proceedings - 2023 IEEE International Conference on Web Services, ICWS 2023, 746–754. https://doi.org/10.1109/ICWS60048.2023.00101

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J. W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. Scientific Data, 3, 160018. https://doi.org/10.1038/SDATA.2016.18

Wilson, C. (2014). Semi-Structured Interviews. Interview Techniques for UX Practitioners, 23–41. https://doi.org/10.1016/B978-0-12-410393-1.00002-8

Woods, H. B., & Pinfield, S. (2022). Incentivising research data sharing: a scoping review. Wellcome Open Research, 6, 355. https://doi.org/10.12688/WELLCOME-OPENRES.17286.2

Yan, Q. (2023). The Importance of Chinese Walls in Banks and Management of Information Flows. Frontiers in Business, Economics and Management, 8(3), 210–214. https://doi.org/10.54097/FBEM.V8I3.7871

Yaqoob, F., & Thomas, J. (2022). Data Governance in the Era of Big Data: Challenges and Solutions. https://doi.org/10.5281/zenodo.8415833

Zeng, J., & Glaister, K. W. (2017). Value creation from big data: Looking inside the black box. Strategic Organization, 16(2), 105–140. https://doi.org/10.1177/1476127017697510/ASSET/0CA85FCC-8354-4A56-8298-797323B1324A/ASSETS/IMAGES/LARGE/10.1177_1476127017697510-FIG1.JPG

Zhang, N., Bahsoon, R., Tziritas, N., & Theodoropoulos, G. (2022). Explainable Human-in-the-loop Dynamic Data-Driven Digital Twins. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 13984 LNCS, 233–243. https://doi.org/10.1007/978-3-031-52670-1_23

Zolnowski, A., Christiansen, T., & Gudat, J. (2016). Business Model Transformation Patterns of Data-Driven Innovations. Twenty-Fourth European Conference on Information Systems (ECIS). Retrieved September 17, 2025, from https://www.researchgate.net/publication/304492522_Business_Model_Transformation_Patterns_of_Data-Driven_Innovations

Zubair, M., Sabzevari, M., Khatri, V., Tarkoma, S., & Hätönen, K. (2024). Access control for trusted data sharing. Eurasip Journal on Information Security, 2024(1). https://doi.org/10.1186/S13635-024-00178-Z

List of Tables

List of Figures