Harri Matinlassi

Developing a Master Data Playbook: Extended Master Data Management Across an Organisation

School of Technology and Innovations
Master of Science in Economics and
Business Administration
Industrial Management

Vaasa 2026

UNIVERSITY OF VAASA
School of Technology and Innovations
Author: Harri Matinlassi
Title of the thesis: Developing a Master Data Playbook: Extended Master Data Management Across an Organisation
Degree: Master of Science in Economics and Business Administration
Degree Programme: Industrial Management
Supervisor: Jyri Naarmala
Year: 2026 **Pages:** 86

ABSTRACT:

Master Data Management (MDM) and Data Governance are critical for ensuring operational efficiency, regulatory compliance, and reliable decision-making in today's business environment. However, as organisations have increasingly been adopting new IT systems without efficient enterprise and information architectural blueprints, maintaining data consistency across system boundaries has become increasingly challenging. This thesis investigates the current state of MDM in a global case company, with a particular focus on identifying data governance silos and decentralised document repositories.

To address these challenges, this thesis proposes an MDM playbook as a practical artifact. The playbook is designed to harmonize existing data domains, while serving as a scalable reference for adopting new data domains. This study was conducted using design science research methodology to develop the artifact, thereby bridging the gap between the organisation's current data governance structure and its desired future state. The study combines qualitative insights from semi-structured interviews and document review with a comprehensive literature review to ensure the proposed solution is both theoretically sound and practically relevant for the case company. The interviews were structured around five core themes, considered critical to address MDM maturity, a unified view of data, performance, data lifecycle management, and adaptability to ensure the long-term usefulness of the playbook. Participants were purposefully selected to capture expert insights on the case company's master data practices.

The findings revealed that, despite having a high-level data governance principle, the case company faces challenges related to unclear data ownership, the lack of unified conceptual data models, and the absence of a common data glossary. Consequently, the case company practices remain largely reactive rather than proactive. The study presents theoretically validated data governance and data management frameworks as part of the proposed solution. Additionally, the findings are limited to be utilized within the case company and are not expected to be generalizable to a broader scale.

As a result of the study, the case company is provided with an actionable guideline for managing master data domains across both business and IT landscapes. The playbook serves as a practical reference for data governance by bridging the gap between abstract governance principles and day-to-day operational practices. Ultimately, by integrating data accountability into core business processes, this thesis provides a roadmap for transitioning from system-centric to data-centric organisation, in which data is strategically managed and valued as a critical business asset.

KEYWORDS: Master Data Management, Data Governance, Information Architecture, Data Quality, Data Integrity, Data Modelling

VAASAN YLIOPISTO
Tekniikan ja innovaatiojohtamisen yksikkö
Tekijä: Harri Matinlassi
Tutkielman nimi: Developing a Master Data Playbook: Extended Master Data Management Across an Organisation
Tutkinto: Kauppatieteiden maisteri
Koulutusohjelma: Tuotantotalous
Työn ohjaaja: Jyri Naarmala
Valmistumisvuosi: 2026 **Sivumäärä:** 86

Tiivistelmä:

Master datan hallinta (MDM) ja datanhallinnan ohjaus (Data Governance) ovat kriittisen tärkeitä tekijöitä operatiivisen toiminnan tehokkuuden, säädöstenmukaisuuden ja luotettavan päätöksenteon varmistamiseksi nykypäivän liiketoimintaympäristössä. Organisaatioiden ottaessa yhä enemmän käyttöön uusia IT-järjestelmiä ilman tehokkaita kokonais- ja tietojärjestelmäarkkitehtuurin suunnitelmia, datan yhdenmukaisuuden ylläpitäminen järjestelmärajojen yli on kuitenkin muuttunut entistä haastavammaksi. Tässä pro gradu -tutkielmassa tarkastellaan MDM:n nykytilaa globaalissa kohdeyrityksessä, erityisesti keskittyen datanhallinnan siiloihin ja hajautettuihin dokumenttiarkistoihin.

Näihin haasteisiin vastaamiseksi tutkielmassa esitetään käytännön tuotoksena MDM-pelikirja (playbook). Pelikirja on suunniteltu harmonisoimaan olemassa olevat tietoalueet ja toimimaan skaalautuvana ohjeistuksena uusien tietoalueiden käyttöönotossa. Tutkimus toteutettiin design science -tutkimusmenetelmää hyödyntäen, ja sen tavoitteena oli kehittää artefakti, joka kuroo umpeen organisaation nykyrakenteen ja tavoiteltujen tulosten välisen kuilun. Tutkimus yhdistää puolistrukturoiduista haastatteluista ja dokumenttianalyysista saadut laadulliset havainnot kattavaan kirjallisuuskatsaukseen varmistaakseen, että ehdotettu ratkaisu on sekä teoreettisesti pätevä ja käytännössä relevantti kohdeyritykselle. Haastattelut rakentuivat viiden ydinteeman ympärille, joiden katsottiin olevan ratkaisevia MDM-kyvykkyyden, yhtenäisen tietonäkymän, suorituskyvyn, tiedon elinkaaren ymmärtämiseksi sekä sen varmistamiseksi, että pelikirja on pitkäaikaisesti sovellettavissa ja hyödynnettävissä käytännön työssä. Osallistujat valittiin tarkoituksenmukaisesti, jotta saatiin kattavia asiantuntijanäkemyksiä kohdeyrityksen master dataan liittyvistä toiminnoista. Tutkimuksen tulokset osoittavat, että vaikka ylätason tiedonhallinnan periaatteet ovat olemassa, kohdeyritys kärsii epäselvästä tiedon omistajuudesta, yhtenäisten käsitteellisten tietomallien puutteesta sekä yhteisen sanaston puuttumisesta. Tämän seurauksena kohdeyrityksen tiedonhallinta on pikemminkin reaktiivista kuin ennakoivaa. Tutkimuksessa esitellään teoreettisesti päteviä tiedonhallinnan ja hallintomallin viitekehyksiä osana ehdotettua ratkaisua. Lisäksi tulokset on tarkoitettu hyödynnettäväksi ainoastaan kohdeyrityksessä, eikä niitä ole tarkoitus yleistää laajemmalle organisaatiotasolle. Tutkimuksen lopputuloksena kohdeyrityksellä on käytettävissä käytännönläheinen ohjeistus master data-alueiden hallintaan koko organisaatiossa ja IT-ympäristössä. Pelikirja toimii tiedonhallinnan tukipilarina, joka yhdistää ylätason abstraktit periaatteet ja päivittäisen operatiivisen tekemisen toisiinsa. Liittämällä vastuu tiedosta suoraan osaksi liiketoimintaprosesseja, tämä opinnäytetyö tarjoaa suuntaa muutokselle, jossa painopiste siirtyy järjestelmäkeskeisyydestä kohti tietokeskeistä organisaatiota ja samalla varmistaen, että tietoa hallitaan strategisesti ja sitä arvostetaan organisaation kriittisenä liiketoimintaomaisuutena.

AVAINSANAT: Master datan hallinta, Datanhallinnan ohjaus, Informaatioarkkitehtuuri, Tiedonlaatu, Tiedon eheys, Tiedonmallinnus

Contents

Figures

Tables

1 Introduction

The notion of data can be considered fragile and somewhat obscure, resulting in numerous characterizations. Commonly, data is portrayed as symbols or figures as a depiction of factual information serving as the fundamental material required for daily procedures and sufficient decisions (Cleven & Wortmann, 2010). Schmuck and Rieder (2025, p. 178) state that the digital revolution transforms the ways organizations work, generate value and communicate with their customers. It demands rethinking existing business models, structures and skills. In the centre of it all is data, considered as a growing resource in terms of its value for innovation and management.

According to Earley et al. (2017, p. 17) today, organizations perceive their data as a crucial asset, which gives information about their products, services and customers. Data can assist in innovation and achieving strategic objectives. However, generating value from data does not materialize by itself. It demands strategy, purpose, engagement and coordination, but also strong leadership and effective management.

Earley et al. (2017, p. 17) highlight that data management can be described as a process, which includes the development, implementation and oversight of plans, actions and procedures designed to safeguard, deliver and increase the value of data and information assets during their lifetime. The scope of data management is vast ranging from being able to make sound decisions to generate value from data to technical implementation and efficiency of databases. Thus, effective data management demands a combination of both technical and non-technical capabilities.

Väre (2019, pp. 16–22) describes in their book that master data can be defined as the foundational data that lies at the core of an organization’s operations. It is relatively stable in nature and widely used across the organization. Master data can be viewed as the structures on which everything else in the organization is built, much like the foundations and walls of a building. When new data is developed or changed due to new

business activities, it is essential to ensure that the new data is properly connected to the foundations, that is to the organization's master data.

Master data management (MDM) strives to increase the value of an organization's key data assets by breaking down barriers between different information systems and organizational departments (Vilminko-Heikkinen & Pekkola, 2019, p. 76). Schmuck and Rieder (2025, pp. 177–179) emphasize that today's businesses are experiencing significant digital disruption, which modifies their business models, operations and capabilities all while demanding substantial investments in technology. In this context, master data management has a crucial role in assisting digital transformation by ensuring that data is authentic, dependable and current. Adequate master data management improves decision-making, productivity and competitiveness. It supports organizations in solving data quality challenges, guaranteeing regulatory compliance and enhanced partnership and analytical capabilities in achieving greater success. When data is of high quality, reliable and rational, employees are also more engaged and confident of the information.

1.1 Purpose of the study and intended contribution

In the contemporary business environment, the need for unified, high-quality master data across all systems is critical for strategic decision-making and operational efficiency. Within the case company, there is an existing foundation of well-governed master data domains and established Data Governance Principles. However, within the case company there is a need to broaden the scope of Master Data Management to cover all related IT systems and master data information modelling to ensure consistency.

The aim of this study is to develop a unique artifact that serves two purposes. First, it provides a reference for efficient master data management across IT systems, while incorporating elements from information architecture. Second, it offers a practical guide on how to begin covering identified core business master data under the master data governance framework. To achieve this, existing documentation of the core data

domains are analysed while bridging the gap between relevant literature to identify gaps and areas of improvement.

Based on this analysis, a playbook is developed that outlines the core principles of master data management while utilising best practices already present within the case company. Throughout the development of the playbook, potential challenges are identified and opportunities are considered to avoid any issues or misconceptions. This will require focusing on extending the Master Data Management (MDM) coverage to cover all relevant IT systems and the information flows, as well as developing master data information architecture and the practical implementation of governance model for all relevant business critical master data domains.

1.2 Research questions

The research problem addressed in this study is that the case company lacks a unified view of its core master data domains from an information architecture perspective and across its IT systems. The department's plan in the future is to manage its core master data across all systems and extend these developed practicalities to other business critical data domains, which need to be governed with agreed principles. While the company has currently good data governance principles in place, its IT infrastructure for several core master data domains remain fragmented. Responsibility for managing these domains is distributed across different individuals and systems, which creates inefficiencies and slows down master data management at an enterprise level.

The research consists of three research questions. They are aimed to extend a governed Master Data model beyond established domains like Customer, Product and Finance objects. To achieve this, the case company is required to transition from a project-centric mindset toward a platform-oriented approach.

This transition involves creating a scalable master data management framework, in which data is treated as a shared enterprise asset across all IT boundaries. In addition, the company should shift from reactive data corrections toward proactive data asset management. This study seeks to answer the research questions to justify necessary investments and to support the company's digital transformation initiatives. These include driving the readiness to modern artificial intelligence (AI), machine learning (ML), and analytics, while simultaneously mitigating risks and enabling scalability. The three research questions are presented below.

1. *What are the key challenges and critical information architecture considerations in achieving uniform master data modelling and governance across system boundaries?*

2. *How can the existing Master Data governance model and current master data domains (e.g., Customer and Product) be extended and applied to new master data domains and across different IT systems?*

3. *What kind of functional structure and practical content are required for an effective Master Data Playbook designed to guide Master Data Management (MDM) practices within the organization?*

The first research question aims to define, which master domains are currently governed effectively within the case company and to examine how these best practices can be leveraged for other domains. From an IT landscape perspective, the question also addresses how these domains are connected. Extending the current model requires an approach in which core principles remain constant, while technical attributes and infrastructure can be adapted to accommodate new master data domains. Before a new domain can be incorporated into the new master data management framework, standardisation of core principles needs to be defined, such as data dictionaries and definitions of attributes throughout the organisation. Although some domains have already been

modelled, the overall IT landscape remains highly complex, with several standalone systems and isolated objects.

The second research question focuses on moving beyond simple data replication and addresses the structural and strategic challenges caused by a fragmented landscape of system environments. By addressing these challenges, the company can select the most effective integration architecture, such as registry or coexistent systems to ensure scalable and compliant single source of truth strategy. This question also considers the decision between centralized and decentralized data management approaches and highlights the need for global data models that can bridge the gap between systems, such as on-premises applications and cloud-based applications.

The third research question aims to define the functional structure and practical content of a Master Data Playbook that operationalizes the findings of the study. The playbook is intended to serve as a practical guide for implementing and extending Master Data Management practices across the organization. It addresses how governance principles, information architecture, roles, responsibilities, and processes can be documented and communicated in a way that supports adoption, scalability, and long-term sustainability of MDM initiatives.

1.3 Research scope and limitations

This study is designed to address the general gap that has been identified within the case company related to the management of master data across multiple IT systems. The research focuses on defining and designing a Master Data Management (MDM) framework and leveraging this knowledge to create a novel artifact: a Master Data Playbook. The study considers the company’s existing Master Data Governance Framework and principles, which serve as a foundation for the work.

The scope of the study is intentionally limited. It does not include coding, software selection, or implementation of any new technical systems. Instead, the focus is on conceptual and logical-level modelling, rather than physical database design or implementation. The study will leverage existing literature and internal case company documentation to develop a context-specific tool. The findings and the developed artifact are not intended to be directly generalisable on a broader scale.

This study approaches the research problem primarily from a documentation perspective to detect and define the gaps. There is also not a unified view of the principles, frameworks, guidance required to manage master data across different IT systems. Rather than recommending specific systems or software, the study provides guidance on how master data should be governed and managed, and how new master data domains can be implemented into the MDM framework.

1.4 The case company

The case company for this thesis is a leading player in the energy sector, recognized as one of the most prominent and influential enterprises in its home country. Established in the late 1940s, the organization has grown to become one of the world's foremost producers of sustainable aviation fuel and renewable diesel. It boasts a strong legacy of pioneering innovative and sustainable solutions for both road transport and aviation. Initially, the company operated as a local oil refiner to ensure national energy security but has since evolved into a global trailblazer in renewable and circular solutions. A central goal of the company is to contribute towards a healthier and more sustainable planet for generations to come.

The organisation's strategy and future focus are centred on strengthening its market leadership, enhancing cost competitiveness, and advancing technological capabilities in the renewable fuels sector. This involves expanding production capacity, realising the full commercial potential of its facilities, and improving operational performance with an

emphasis on safety and efficiency. The company is prioritising strategic initiatives and cost structure optimisation while preparing for the next phase of growth, focusing on selected development projects.

This study is conducted for the Processes, Data and IT unit of the case company, which operates within the Global Functions unit. The unit's purpose is to support the core processes of the company by providing data platforms for multiple use cases and by operating the core IT solutions and infrastructure of the company. The unit possesses expertise in various areas, including IT solution management, enterprise architecture, data engineering and management, and business process development.

More specifically, this study is carried out for the Business Process Department, which is responsible for process and information architecture as well as data management. The department will be supporting this study, while experts from the whole unit will be involved contributing views and perspectives related to IT and enterprise architecture. Their involvement is essential to ensure that the study also addresses the core principles of building effective data platforms through IT solutions.

1.5 Structure of the study

The thesis is organized into five primary sections. The first chapter is an introduction to the topic and the study by and large. It provides background information about the research subject in question and the motivation behind the study. The chapter also presents the purpose behind the study and the intended contribution and limitations for the study. The second chapter presents the theoretical background behind the study. It explains the basic theories behind master data, master data management, data management frameworks, information architecture and logical data modelling, which are utilised in this research.

The third chapter presents the methodology used in this study. It explains how the study is conducted and outlines the overall research process. The empirical part of the study consists of two main parts. The first part includes acknowledgment of gaps in the company's current master data management practices and building on these gaps to formalise the framework to be followed. The second part consists of semi-structured interviews conducted with key stakeholders within the case company. The interview data is analysed and used to complement and refine the future master data playbook alongside the identified gaps in the current framework.

Chapter four presents the results of the study. The findings will also be later reviewed and evaluated with the case company, and final adjustments to the structure and content of the playbook will be made accordingly. Finally, chapter five presents the conclusions of this study and its findings.

2 Literature Review

This chapter presents the theoretical background of the study. It begins by introducing the concept of master data and then examines master data management (MDM). The chapter outlines key data management principles and challenges, emphasizing the importance of clearly defined roles and responsibilities in MDM. In addition, it describes the main data management processes and reviews different master data management frameworks. Data governance, logical data modelling, and information infrastructure are also addressed as key themes.

2.1 Master Data

According to Väre (2019, p. 16), master data, often described as key, core, or foundational data, is information that is central to business operations. It is generally stable in nature and widely used across the organization. Das and Mishra (2011, p. 131) define master data as the key to business, which is often made up of four classes: people, things, places and concepts. Ibrahim et al. (2021, p. 181) states that master data refers to the core business information of an organization, which is shared across applications, departments and organizations and is considered a critical organizational asset. Master data gives answers to questions such as: What does the business consist of or what does the organization do (services, products) and where does it operate (locations and geolocations) and with whom does it operate (customers, partners, employees, suppliers, and other parties).

Väre (2019, p. 16) describes that master data can be divided into more specific categories. There is structural master data, which usually includes the financial structures and hierarchies of an organization. Some definitions also recognise temporary or time-bound master data, such as pricing information or different product compositions. The effective management of master data is critical for the smooth functioning of an organization's daily functions and for reliable reporting of business activities. Loshin (2008, p. 8)

explains that master data objects can be characterised by their cross-functional utility and stability. They are often organised through complex taxonomies, such as the 'Party' hierarchy. This allows for multidimensional characterisation based on organisational roles for instance supplier, vendor, or customer. According to Earley et al. (2017, p. 171), master data management is the power over master data to facilitate dependable, shared and contextual operations across systems of the most important, current and accurate form of information concerning the critical business activities.

According to Väre (2019, p. 23), the official definition of master data is strongly linked to two main thoughts. First being that master data is critical information for the business and without it, the organization cannot function. Secondly master data is shared across the entire organization, and all different operations use this same data. This is the starting point when an organization considers what constitutes its master data. However, data does not stop being master data even though it is not used by every function. For instance, customer data is still master data even if another department in the organization does not use it. Sometimes too much focus is placed on the second point of the general master data definition even though the first point, criticality of the data, is more significant. Master data can also be distinguished by different layers of shareability.

Loshin (2008, p. 131) complements the identification process of master data as a dual perspective method consisting of top-down and bottom-up processes. The top-down approach is more a strategic review of enterprise data and business process models to identify critical data objects within multiple business areas. The bottom-up approach is more of a technical review of existing enterprise data assets and applications to find data objects already utilised as master data objects. Eventually, the identification process will consist of aspects of both approaches complemented with practical knowledge and subject matter expertise.

Väre (2019, pp. 23–26) defines that the most common master data domains include parties, things and locations, temporal master data, and structural or financial master data.

Temporal master data can be described as an intermediate form between master data and transactional data. Organizations have a lot of data that has a shorter lifecycle than their products, customers, or locations in general. Thus, this type of data is closer to the core operational data of an organization, such as contract data, price lists, and product configurations. Their lifecycles can span from a few days to months, yet they still play a crucial role operationally.

According to Cleven and Wortmann (2010), each industry introduces its own specific master data specifications and characteristics. However, there can be classified three key master data domains: party, thing and location. The party domain comprises business partner master data, including suppliers, employees and customers. The thing domain covers master data concerning company assets, services or products. The location domain encompasses master data associated with regions, sites and places.

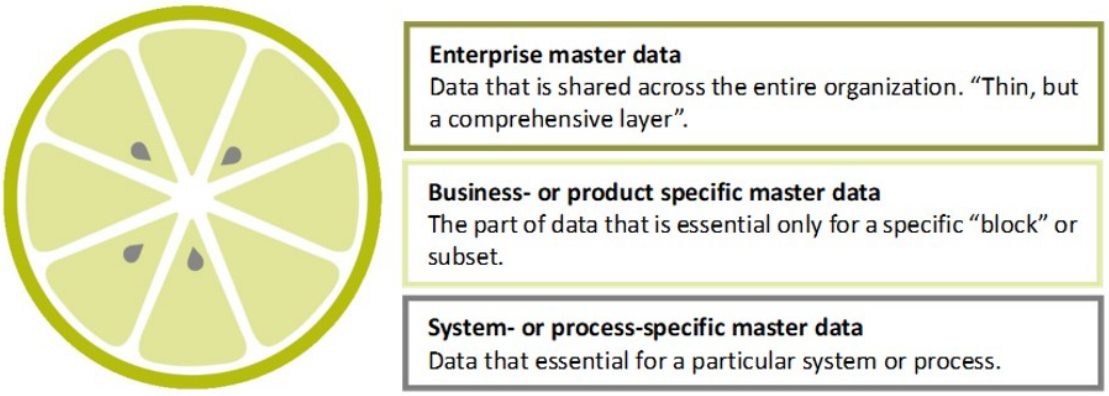

Figure 1. Three layers of master data (Väre, 2019, p. 28).

According to Väre (2019, pp. 28–29), organizations master data often consists of different levels. These levels can be conceptualized as three distinct layers of an orange: a peel layer, a segment layer and a seed layer (Figure 1). The peel layer is data that is shared across the entire organization. It is the part of data that is most essential for uniquely identifying data. If this layer is not governed as part of master data management, master

data is not truly being managed. The next layer consists of segments, which are easy to separate from one another. Only the peel layer links the data of different segments into the same record. These represent the master data of business units, product groups, or different geographical areas. Each segment has its own requirements on master data, defining which attributes it needs. Depending on the organization the last layer can consist of a few seeds or several, meaning that system- or process-specific master data may exist in small amounts or not at all depending on the organization. The most challenging aspect of the seed layer is identifying only those individual data attributes that are critical for the organization's operations.

Hillard (2010, p. 84) describes that one reason for the gained popularity of master data management is the rise of enterprise resource planning (ERP) systems. These systems integrate financial management and business processes into a unified software platform. Master data is the unifying element across divisions and modules describing factors such as identifiers, customer and employee numbers, locations and various static codes. Master data is also often treated as metadata in practice. According to Bourne (2014, p. 383) Metadata is a definition of data attributes. These are typically stored in a centralised data glossary or repository, containing explanations such as field names of databases, tables of where the data fields exist, explanations of field definitions, and lists of applications which reference the database fields.

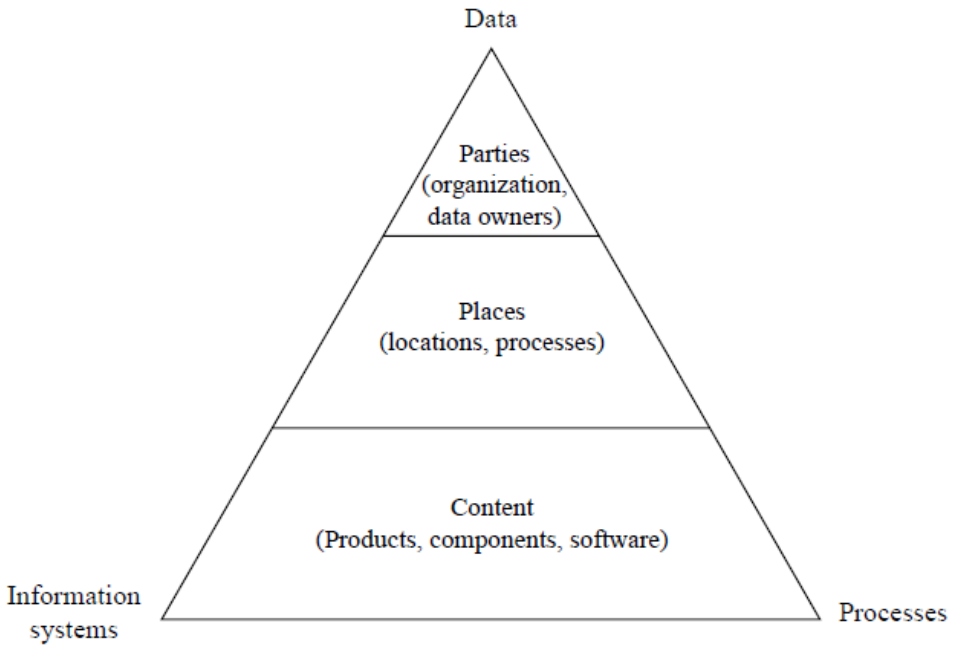

Figure 2. The concept of one master data (Silvola et al., 2011, p. 151).

Silvola et al. (2011, pp. 151–152) define three primary concepts that form a single master data concept: data, processes, information systems (Figure 2). Together these create a practical governance framework for implementing and managing master data management. The data concept comprises models, attributes and explanations. One master data is corrected, rationalized and consolidated into a single system of record (SOR). High data quality is achieved over time through sustained data quality management (DQM). The process concept includes clearly defined data ownership, as well as procedures for maintaining the data. Information systems concept comprises applications and technologies that enable automated data integration and distribution.

According to Silvola et al. (2011, p. 152) these three concepts together enable a unified master data system and bring together three core master data entities: parties, places and content. Parties represent the most critical entity, with organizational roles, responsibilities, and the network of data owners. Places refer to operational environments and processes where data is used. Content includes the actual master data objects.

2.2 Master Data Management

Väre (2019, p. 38) defines master data management as practices and methods, whose sole purpose is to ensure the accuracy and appropriateness of master data within an organization. This ensures that master data becomes the objective of the entire organization rather than the responsibility of individual specialists, and that the data is managed prudently. A robust MDM system supports informed business decisions, efficient operations and regulatory compliance with internal policies as well as external regulations (Schmuck & Georgescu, 2025, p. 575). The main objectives of master data management are to ensure that there is one authoritative record for each real-world entity, data is created once with reuse in the organization, data is updated only once and ensured that it is updated everywhere it is needed and that data is made available when and where it is needed.

According to Väre (2019, pp. 38–41), the main general objectives of master data management must be aligned inside an organization to its business goals and needs. The objectives are pursued through six different domains, which together form organization-wide master data management. The first domain is the background drivers of master data management such as architectures and technologies, followed by the framework of master data management and the data governance model. The development path of MDM includes the vision, goals and strategy, while the mechanics include data governance processes and the data lifecycle. Finally, the master data quality is measured and improved.

Hikmawati et al. (2021, p. 90) define master data management (MDM) as an approach for preserving, incorporating and integrating master data to achieve consistency across information systems. The main purpose of MDM is to manage master data so that it remains dependable, authentic, present, and contextual to support various business units and applications. According to Das and Mishra (2011, p. 131), MDM can also be described as the technologies, means, and mechanisms necessary to establish and sustain clean, rational and adequate master data. It complements business knowledge and

serves as a reliable source of dimensional data. It is important to emphasize that MDM should not be regarded solely as a technical solution, but rather as a strategic asset that forms an essential part of organizational management (Schmuck & Rieder, 2025, p. 177).

> "Master data management (MDM) is an application-independent process which describes, owns and manages core business data entities. It ensures the consistency and accuracy of these data by providing a single set of guidelines for their management and thereby creates a common view of key company data, which may or may not be held in a common data source" (Smith & McKeen, 2008, pp. 65–66)

According to Cleven and Wortmann (2010), the definition above clearly underscores that MDM extends beyond technological choices to also include organizational support and well-defined processes. Existing MDM research has identified five key components when launching an MDM initiative: master data structure, master data systems architecture, master data governance, master data processes, and master data quality. Vilminko-Heikkinen and Pekkola (2017) state that master data management seeks to resolve data quality challenges by emphasizing business operations, data quality, and integration of information systems. It aims to overcome issues relating to data dissolution, isolated systems, conflicting operations and architectural complexity. According to Väre (2019, pp. 47–50), effective master data management creates business value by enabling new and improved data-driven business opportunities and increasing operational efficiency.

A study by Zhao et al. (2020, p. 1408) finds that usually, master data management is established upon existing systems. Through these it retrieves current data and employs advanced technologies and practices to support up-to-date, authentic and automated circulation, analysis and validation across the organization. MDM has various possibilities including enterprise-wide data mining, subsystem data collection, data distribution, establishment of consolidation master data structures, integrated information management and support for multi-user data management.

Hillard (2010, pp. 147–148) notes that while the definition of master and reference data is simple, it becomes more difficult to define them within the systems. The first task when initiating MDM should be to define reference and master data used by each system and define a CRUD (Create-Read-Update-Delete) matrix, which represents the roles and permissions of different systems in the creation of master data. In an ideal landscape of systems each master data item would have only one point of creation, update, and deletion, which would reduce the issues of synchronization across the landscape. In a real-world situation, however, this is seldom the case since in a multi-system environment there is a need to add and update entries or attributes during the process. Reading, by contrast, is not considered a problem.

According to Silvola et al. (2011, p. 149) master data management comprises best practices in data management, and coordinates core stakeholders, participants and customers. It operates as a workflow-oriented process through which business units and information systems collaborate to unify, sanitize, distribute and safeguard shared information through the organization. The goal of MDM is to deliver timely and accurate data to support business and growth, while ensuring that key business data is defined and standardized and systematically managed as it evolves.

Silvola et al. (2011, p. 149) specify that master data management is commonly categorized into two areas: operational MDM and analytical MDM. Operational MDM focuses on incorporating operational systems, such as ERP, CRM and supply-chain management to support consistent data flows. Analytical MDM, on the other hand, is associated with warehousing (DW), including customer data integration and managing financial performance. Combined, these areas form organizational master data management. Recent studies also indicate according to Adapa (2025, pp. 99–101) that 78 percent of organizations struggle with data consistency maintenance across their IT system landscape. On the other hand, organisations who establish an MDM solution are able to produce substantial improvements on data reconciliation efforts and data management costs. Eventually, organisations with mature MDM practices can outperform their counterparts by

94 percent improvement in data accuracy and by achieving 40 percent better improvement meeting the regulatory compliance requirements. Current trends also indicate that cloud-based MDM solutions are dominant in the market, reporting almost half shorter implementation deployment times compared to on-premises solutions.

2.2.1 Data management principles

Earley et al. (2017, p. 19) have identified five guiding principles for effective data management. Organizations should view their data and information as a vital asset, while also managing it cautiously to guarantee its protection, sufficiency, purity, understanding and proper usage. Data management should be overseen by both business data stewards and data management professionals. They also note that data management is both a formal business operation supported by a set of connected disciplines. Finally, data management is also an evolving professional field as a part of information technology.

Earley et al. (2017, pp. 21–23) underline that data is an asset with distinctive characteristics that shape how it should be managed. Even though data shares similarities to other assets, it differs from for instance financial and physical assets through its use. Data is not similarly consumed when it is used. Nevertheless, data has value and therefore should be considered economically. Organizations seeking to improve data-driven decision-making must develop appropriate methods to assess and quantify the value of their data. The concept of viewing data as a product can be guided by four principles (Hubert Ofner et al., 2013, p. 475; Wang & Strong, 1996): identifying and understanding users' data requirements, treating data as the result of successful business process, recognizing data has its own lifecycle and assign a dedicated professional to oversee the data processes and outcomes.

According to Earley et al. (2017, pp. 21–23), effective data management requires managing data quality. Making sure that data is fit for its intended use is a key objective of data management. This involves metadata, which means that there is data about the

asset. Data management also requires a plan and an enterprise-wide mindset as it draws on a range of skills and knowledge across the organization. Furthermore, data management is also lifecycle management and demands risk management and leadership engagement.

Hannila et al. (2022, p. 30) further highlight the importance of rethinking organisation's data management principles. To be a data-driven organisation requires changes in the business culture and restructuring of the organisation. Employees' assumptions and experience can be complemented with data analysis, but organisations should not blindly trust that IT technology is the sole solution but rather complement it with people, processes and technologies.

2.2.2 Data management challenges

Cleven and Wortmann (2010) describe that despite increased awareness and development, organizations still face challenges in managing master data. One key factor is that swift advances in information processing and digital storage have led from data scarcity to exponential growth of data volumes. According to Schmuck and Georgescu (2025, p. 577), operational businesses continue to face challenges in implementing master data management, making the assessment of MDM maturity a critical factor in defining appropriate actions for future development.

The success of master data management is assessed through enhanced transparency, quality, conformity, operational and decision-making competence and confidence and fulfilment (Schmuck & Rieder, 2025, p. 179). When master data is inadequately managed, deficiencies and weaknesses arise, which can result in flawed stakeholder decisions (Hikmawati et al., 2021, p. 90). Cleven and Wortmann (2010) also note in their study that weak master data management can lead to significant issues such as operational deficiencies, poor decisions and waste in resources and time. As data volumes and the

number of systems grow, issues such as duplicates, missing attributes and inconsistencies become more common.

According to Silvola et al. (2011, p. 160) the main challenges related to master data concern ambiguous definitions and low data quality. Process-related challenges include insufficient data ownership, inconsistent data management processes and the absence of data quality efforts. Furthermore, integration across multiple applications remains a central challenge in building a holistic master data environment. Many of the main challenges regarding data stem from poor definitions of master data and data model. Organizations struggle to establish common definitions and enterprise-wide models. Additional difficulties can arise from unclear data ownership and limited employee commitment towards these issues. The use of multiple systems to manage master data makes integration also costly and uncertain.

According to Haug and Stentoft Arlbjørn (2011, p. 288) organizations struggle to achieve high data quality due to several critical barriers. By increasing the knowledge of these barriers that require attention, organizations can improve their ability to attain higher data quality. Their findings indicate that unclear responsibilities of master data management have the greatest impact on quality. The study also confirms that most organizations acknowledge that weak master data quality leads to serious adverse effects.

In another study Haug et al. (2013) examine a manufacturing organization and its master data and determine 12 master data quality barriers. These include the absence of clearly assigned responsibilities for certain forms of master data, insufficient definition of roles regarding data creation, usage and upkeep. The authors further highlight the shortcomings in organizational processes, insufficient attention to data quality and data quality measurement. Further barriers include the absence of incentives or sanctions linked to data quality, inadequate training for data users, and missing formal data quality policies and procedures. In addition, managers may fail to highlight the value of data quality, and

organizations may lack appropriate IT support. Finally, there may be limited opportunities for user input in existing systems and poor overall system usability.

Haug et al. (2013, p. 237) study further argues that insufficient data quality has negative consequences, including reduced customer satisfaction, higher costs, poor decision-making, low performance and personnel fulfilment. Moreover, low data quality undermines trust in data and can result in resistance to new data initiatives. By identifying data quality barriers, the authors aim to increase organizational knowledge on data quality issues and that these should be addressed not only as technical but organizational challenges.

In their study, Ibrahim et al. (2021, p. 185) identify 19 causes that negatively affect master data and its quality including data governance, information systems, data quality policy and standard, data quality assessment, integration, continuous improvement, teamwork, data quality vision and strategy, understanding of the systems and data quality, data architecture management, personnel competency, top management support, business driver, legislation, information security management, training, change management, customer focus, and data supplier management. The identified causes can be classified into five categories: organizational, managerial, stakeholder, technological, and external.

When organizations, systems and applications operate in silos, information can become isolated. This fragmentation produces false, deficient and duplicate data, weakening analytics and increasing incorrect business decisions (Fatehali, 2011 in Vilminko-Heikkinen & Pekkola, 2017). Haneem et al. (2019, p. 25) examined the adoption of MDM in local government context and identified that data quality and data governance are the primary drivers of adoption. In addition, complexity, management endorsement, technological proficiency and population demand were shown to have a crucial impact on MDM adoption.

In the context of information management Inmon et al. (2015, p. 63) expresses that corporations struggle to manage unstructured information, which is estimated to account

for over 80 percent of all data, mostly in a textual form. Textual information is often the most informative form of information and holds the most potential for business gains. Unstructured information can be further split into two main categories – repetitive and nonrepetitive. However, corporate decisions are done based on structured data primarily due to its' nature to be easily automated and organised into databases for systematic analysis.

Another study by Baghi et al. (2013) validates that data quality management is a core element of information management and is essential across organizations. Low data quality negatively affects business operations, raises costs and reduces stakeholder contentment. Pricing data errors for instance causes incorrect invoices, while inaccurate customer data can result in delivery failures and complaints. Inadequate accounting data increases the need for manual work. Approximately 75% of organizations report costs related to errors in master data and thus, enhancing data quality has become a key priority.

Haug et al. (2023, pp. 1903–1904) explore the key barriers preventing organizations from achieving high master data quality. Evidence from industry questionnaires also highlights the severity of these challenges with 95% of organizations reporting negative impacts from low data quality and 84% of CEOs expressing concern over their data quality. Moreover, even with technological progress in data management, the economic impact of data quality problems appears to grow. Wang and Strong (1996, p. 5) also note in their study that deficiencies in data quality have significant social and economic consequences. In contrast, high-quality data is accurate, fit for purpose, clearly represented and easily available to those who use it.

According to Earley et al. (2017, pp. 23–25), unlike other assets, data is easy to copy and transport, but difficult to restore once it's lost or destroyed. Data is also vulnerable to theft without being physically gone. In addition, the valuation of data assets is particularly challenging due to their contextual and commonly temporal nature. Low data

quality is therefore expensive for organizations making it vital to work closely with data consumers to meet the needs and attributes of high-quality data.

Another study by Merwe et al. (2024, pp. 570–571) state that various practices have been implemented around the world to solve master data management issues and to improve organizational supply chain performance. These practices aim to overcome the key barriers related to master data integration to enable proper MDM. A common solution is the standardized data quality frameworks such as ISO 8000. It is the international data quality standard that offers internationally accepted principles and practices for managing master data. In addition, new technologies such as machine learning and artificial intelligence have reshaped master data management. By enabling systems to learn and adapt, AI technologies can automate data cleansing, enrichment and classification, while recognizing anomalies that manual approaches may miss. These capabilities not only improve data accuracy but also support predictive analytics, enabling businesses to forecast demand, optimize inventory, and improve supplier management.

2.3 Data governance

Abraham et al. (2019) describe data governance as the authority, control and accountability that guide data management across the organization. Its objective is to enhance the value of data while reducing risks and costs related to its use. By defining a cross-functional governance framework, data governance considers data as a strategic asset. It sets clear decision rights and responsibilities, establishing approaches and standards, and ensures compliance. Data governance helps organizations address issues such as poor data quality and incompleteness. The concept of data governance appeared as a synonym for structured data management in the early 2000s, succeeding Enterprise Information Management (EIM) with a focus on data warehouses and data marts (Hannila et al., 2022, p. 32).

According to Väre (2019, p. 59), the starting point for a data governance strategy is always the organization’s business strategy. To develop this strategy successfully, the

organization must understand its business strategy well. Ideally, the people who are involved in defining business strategies also participate in the development of the data strategy. The objective of data governance strategy is to clarify how data supports achieving the organization's strategic business goals. The business strategy defines the objectives and requirements for data. The data governance strategy defines how data and data management are developed and advanced to achieve business goals.

Hikmawati et al. (2021, p. 91) describe data governance as the policies and procedures set of how data is managed within an organization. It is a structured process that guarantees data assets are well-managed and used effectively. Data governance brings various gains such as data standardization, sufficient policy development and defined stakeholder responsibilities. It enables alignment between data management and business preferences, supporting compliance and risk mitigation.

According to Brous et al. (2016, pp. 121–123), data governance is guided by four core principles: organization, alignment, compliance and common understanding (Figure 3). These principles address the organization of data management, adjusting data with business objectives, regulatory and policy compliance and shared understanding of data across the organization. Alignment includes securing data that meets quality standards and supports business needs through definitions, supervision and imposing data policies both at internal and external levels. Effective data governance, however, requires that the organization understands the meaning of its data and its significance to business operations.

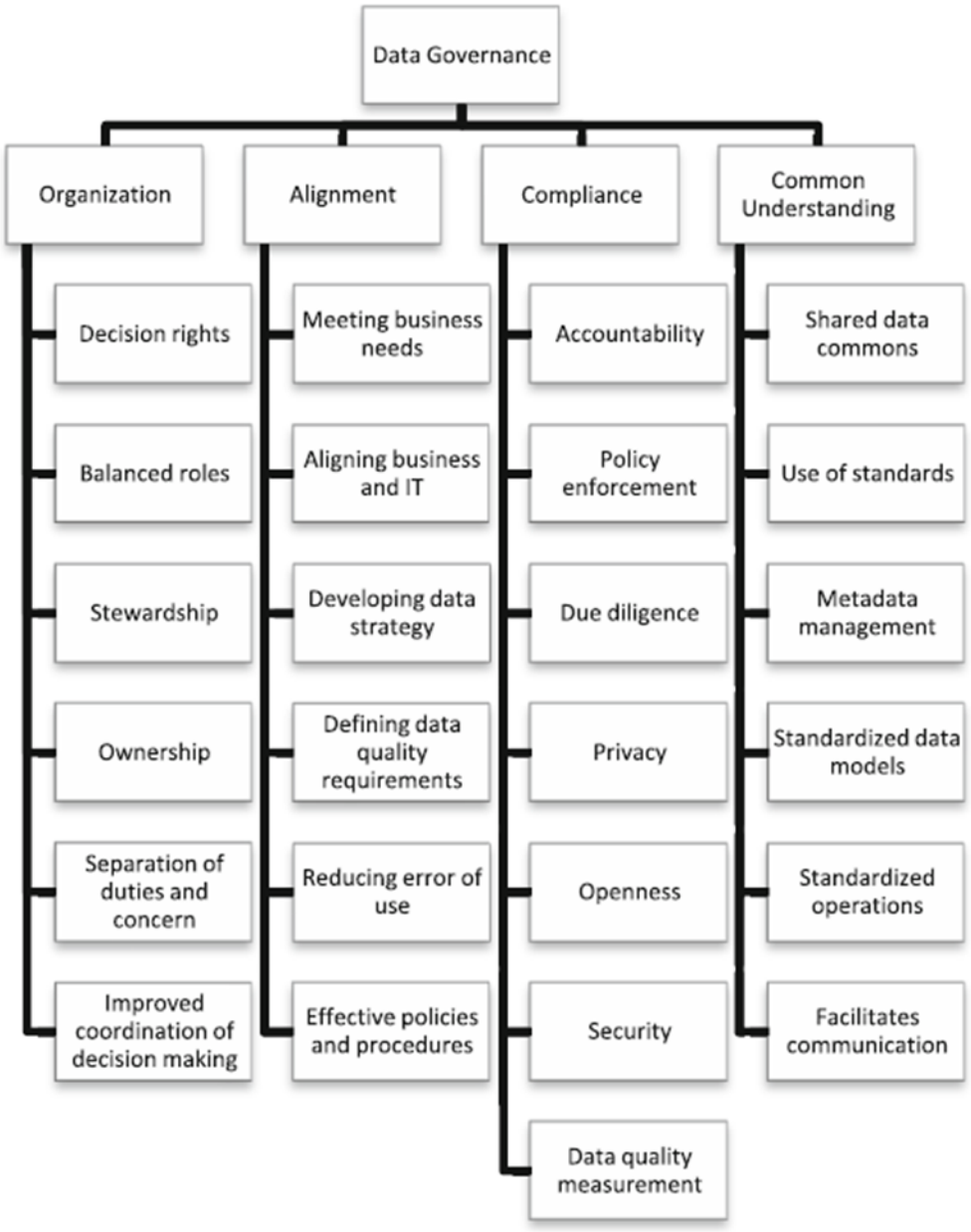

Figure 3. List of key concepts and principles of data governance (Brous et al., 2016, p. 122).

Weber et al. (2009, p. 2) find that data governance clarifies roles and responsibilities for data-related decisions. It sets enterprise-wide standards for data quality management and ensures alignment with strategy and legal requirements. They argue that effective data governance must be designed to an organization's specific context. Another study by Gregory (2011, p. 230) highlights that a lack of sufficient data governance increases

organizations' risk of failing to follow corporate governance and compliance and exposes themselves to substantial risk. Moreover, the absence of proper data governance prevents organizations from fully maximizing the value of their data assets.

According to Hikmawati et al. (2021, p. 94) through data governance, master data management supports improved data management by guiding organizations to adjust the roles and responsibilities of business and IT stakeholders. Data governance outlines the necessary roles, operations and obligations to achieve successful MDM implementation. When these are adequately defined, data governance can improve data quality by ensuring that master data is well maintained.

A study by Guerreiro et al. (2025, p. 190645) finds that data governance is particularly important when organizations strive to enhance the quality of data, protection and compliance. However, studies highlight persistent issues related to implementation, maturity and usefulness of existing frameworks. Commonly, data governance is viewed as a collection of actions, operations and systems applied to secure data integrity, protection and utility. The effects of data governance are substantial, such as better decisions, operational performance, strengthened compliance and risk control. By enabling organizations to manage data as a strategic resource, data governance reinforces business goals and fosters innovation, strengthens the data management maturity, and increases the value and reliability of data assets. These capabilities allow organizations to become more flexible and quicker to respond to emerging possibilities to gain competitive advantage.

The study results by Guerreiro et al. (2025) demonstrate that structured data governance is increasingly crucial for improving organizational efficiency and meeting regulatory obligations. The research stresses the value of adaptive governance frameworks that incorporate structural, procedural and relational systems that cooperate to strengthen data management practices. Furthermore, the study underlines the need to assess data

governance maturity levels as a means of detecting improvement areas and for monitoring development.

Osakwe (2026, p. 19) presents several metrics to measure the success of governance effectiveness. While metrics tracking compliance provides a baseline for governing bodies, it is recommended to implement a variety of metrics to measure the overall effectiveness of data governance. The measurement framework should integrate data quality dimensions, such as accuracy and consistency for high-value data to ensure data reliability over time. Stakeholder satisfaction levels measuring effectiveness from user perspective accompanied with surveys, focus groups and feedback ensures that stakeholders understand the objectives, reasoning and governance as an enabler rather than hindering work. Risk metrics, such as tracking security incidents and access violations assures governance prioritisation and value over expenses. Business value metrics justify governance activities by aligning them with broader organisational goals, such as enhanced decision-making and the achievement of competitive advantages.

2.3.1 Roles and responsibilities

Data management typically involves structures and roles encompassing multiple organizations and individuals. Every organization differs in their priorities and requirements and therefore employs various approaches to organizing data management and defining individual roles and responsibilities (Earley et al., 2017, p. 29). Identifying a key business owner for every data element remains one of the main challenges in any MDM initiative and unclear ownership may lead to poorly defined processes that hinder data maintenance (Smith & McKeen, 2008, pp. 64–65).

Vilminko-Heikkinen and Pekkola (2019, pp. 76–87) examine the evolution of data ownership, responsibilities and roles during master data management development. Their study looks at how these changes unfold in development projects and influence the outcome of the project. The findings demonstrate that data ownership has a significant

impact on development outcomes and how roles and responsibilities grow. Unclear ownership and roles often lead to conflicts. Organizational challenges are the main barriers for achieving high-quality master data, in particular ambiguous roles and responsibilities, the absence of data management policies and procedures and limited leadership support. These challenges emphasize the importance of clear ownership, roles and responsibilities, which are commonly addressed through data governance. Data governance seeks to define policies and processes that enhance the maintenance and development of master data. Dahlberg and Nokkala (2015, pp. 30–31) argue that the body of corporate governance should guide and monitor IT by setting objectives and establishing accountabilities for the use of IT and data through Evaluate, Direct, and Monitor (EDM) cycle. The corporate governance body ensures preparation against uncertainties and secures business continuity within the context of data management, supported with information architecture and information risk management.

Vilminko-Heikkinen and Pekkola (2017, 2019) highlight that effective management is essential to achieve success in any change initiative. The study identifies imprecise data ownership as the primary obstacle to prosperous master data management, significantly influencing decision-making, investments and liability amidst the development phase. Roles and responsibilities associated with MDM change and develop sluggishly and inconsistently, commonly prevailing to be unclear when data crosses multiple operations and organizational groups.

Vilminko-Heikkinen and Pekkola (2019) also state that MDM change does not result from technology alone but from evolution of social factors, such as governance, power and people and material elements together, such as structures and information. A recurring confusion regarding ownership of data and information technology caused business groups to consider data quality as an IT problem as opposed to a common concern for the organization. Their study highlighted MDM as an organization objective rather than a technical quick fix, emphasizing that clear data ownership is fundamental to enhancing data quality and achieving viable governance. Hillard (2010, p. 199) states that

information and data users should be guided by the governance strategy and by the description of different user groups involved. This implies understanding how information is used throughout the organization, across all levels and business areas.

According to Väre (2019, p. 174), assigning responsibilities concretely across different roles can naturally be challenging. It is particularly important to ensure that no area of master data management is left without ownership, but also responsibility should not be assigned to multiple individuals. If several people are responsible for the same task, there is a significant risk that the task will not get done or that conflicts arise over how it should be conducted. Clear role assignments improve communication, execution of tasks and oversight.

Table 1. Allocation of responsibilities in master data management (Väre, 2019. p. 176).

Data Management Area	Sponsor	Data Owner	Data Expert	Data Steward
Strategy and objectives	A	R	C	I
Data governance model	A	R	C	I
Data management processes	I	A	R	C
Data definition	I	A	R	I
Data quality	I	A	C	R
Architecture and technologies	I	A	C	I

Väre (2019, p. 176) proposes a master data management responsibility allocation table, where it is possible to see how responsibilities can be distributed across different roles. In Table 1 the author uses the RACI-model, which is commonly applied and a well-proven method for defining different responsibility roles. The RACI model distinguishes four levels of responsibility. “A” refers to Accountable, which has the highest level of responsibility. Accountable is responsible for ensuring that tasks are completed on time and as agreed. “R” stands for Responsibility, meaning the one who performs the work. Responsible one carries our task according to what has been agreed. “C” means Consulted and

is the role that provides advice but is not responsible for completing the task. “I” stands for Informed and refers to individuals who must be notified about the decisions or results. The best outcomes for different tasks are achieved when the Accountable and Responsible roles are assigned to those who are best positioned to influence the result.

2.3.2 Processes and standards

According to Väre (2019, pp. 89–90), master data management processes must ensure high data quality. These processes include the creation of new master data, the maintenance and modification of existing data, and the deletion or archiving outdated data. In addition, effective master data requires well-defined processes for transferring data between systems, migrating data from old systems to new ones, and combining master data with other data for reporting and analytics. These objectives mean that master data is handled manually as little as possible, and that its availability across different systems and business operations is ensured as efficiently as possible. Even though the above refers to master data management processes as their own processes, it is important to remember that each of them is carried out as a part of a business process. This means that MDM processes are not separate, independent processes, but different perspectives on existing business processes.

According to Hikmawati et al. (2021, p. 91) organizational process is the most crucial part of master data management. They argue that a successful MDM process begins with identifying the organizational needs, the relevant master data and the proper operating system, followed by MDM governance and defining the process maintenance and data standards. Furthermore, organizations should describe a plan for enhancing data quality in the future, design the MDM architecture and ensure proper training and communication for relevant stakeholders. Lastly a clear roadmap or strategy should be developed for master data management and define the relevant attributes.

Table 2. Data standard and procedural guidelines (Earley et al., 2017, p. 49).

Area	Description
Data modelling and architecture standards	Standards covering data naming conventions, definition standards and standard domains and abbreviations
Business and technical metadata standards	Definition of standard business and technical metadata to be captured, maintained, and integrated
Data management guidelines and procedures	General guidelines and operational procedures for managing data
Metadata integration and usage procedures	Procedures for integrating metadata and governing its use
Database and continuity standards	Standards for database recovery and business continuity, performance, data retention, and external data acquisition
Data security standards	Standards and procedures related to data security
Match, merge, and data cleansing standards	Rules and procedures for data matching, merging, and cleansing
Business intelligence standards	Standards and procedures for business intelligence development and use
Enterprise content management standards	Standards and procedures for enterprise content management, including enterprise taxonomies, legal discovery support, document and e-mail retention, electronic signatures, report formatting, and report distribution methods

Earley et al. (2017, p. 48) define that data standards and guidelines consist of naming, setting requirements, data modelling, constructing a database as well as architecture and operation specifications for every data management action (Table 2). These can differ considerably both within a single organization and between different organizations. However, data standards should generally be developed by data management specialists and assessed, accepted and implemented by a data governance council. Data standards also need to be clearly introduced, overseen, imposed and reassessed regularly.

Väre (2019, pp. 92–95) outlines that master data is created and maintained within a business process. When aiming to ensure the quality of master data, the focus is put on the business processes in which the data is created or changed. From these processes, all activities that are related to the data are collected and formed into one. From this it is possible to see how data is generated or transformed as the process evolves. In developing a process, effective collaboration is essential. The focus is on developing the process while also taking quality requirements into account and not on creating separate master data processes. If a process is made too complicated or slow in the name of quality confirmation, individuals executing the process will develop their own informal ways of working. Therefore, it is encouraged to follow the basic principles of Lean thinking when developing processes. Emphasis on people's skills and expertise is highlighted in Reif et al. (2022) study as the core element of successful business process management. In practice, commitment to developing processes should be managed through enabling an environment where process flows are designed through clear process descriptions, manageability via user-friendly tools and visible management support.

2.4 Data Management frameworks

As has been established, according to Earley et al. (2017, pp. 173–177), master data management refers to the mechanism of defining and sustaining the ways master data is generated, combined, preserved and utilized across an organization. A key objective of master data management is to determine or develop a single "golden" source for each objective, such as product, location, individual or organization. A designated "system or record" can serve as the ultimate source of information for a particular item. That said, a single system can unintentionally generate duplicate records of the same item. Therefore, various methods can be applied to determine as accurately as possible the most correct information about the entity.

Earley et al. (2017, pp. 33–43; 173–177) find that master data management can be applied using various tools, such as data integration tools (for example ETL), data cleansing

tools, operational data stores (ODS) that act as master data hubs or dedicated MDM systems. Data management consists of several connected actions, with their own purpose, tasks and obligations. Therefore, a dedicated framework is useful for understanding data management as a whole and for seeing how different parts connect to each other. The authors outline several data management frameworks, including the Strategic Alignment Model, the Amsterdam Information Model, the DAMA DMBOK framework, the DMBOK Pyramid, and the evolved DAMA data management framework.

According to Earley et al. (2017, pp. 33–43), the Strategic Alignment Model focuses on the key factors that guide data management. At its core is the link between data and knowledge. Knowledge is mostly related to business strategy and how data is used in daily operations. Data on the other hand is linked to information technology and the functions reinforcing the access and management of data. Around this are four main areas of strategic decision-making: business strategy, IT strategy, organizational structure and processes, and IT infrastructure and processes. The Amsterdam Information Model looks at how businesses and IT consider strategy. It is also described as the 9-cell model. The model includes a middle layer concentrating on construction and tactical issues, such as planning and architecture. In addition, it highlights the importance of communication and sharing information.

Earley et al. (2017, pp. 33–43) highlight that the DAMA-DMBOK Framework explains the different knowledge areas that together form data management. There are various visualisations of DAMA's Data Management Framework. The DAMA wheel (Figure 4) defines the knowledge categories. Data governance is at the core because it ensures consistency and balance between all data management activities. All other knowledge categories circle around the wheel. All of them are important for effective data management, but organizations may apply them at various situations relative to their needs.

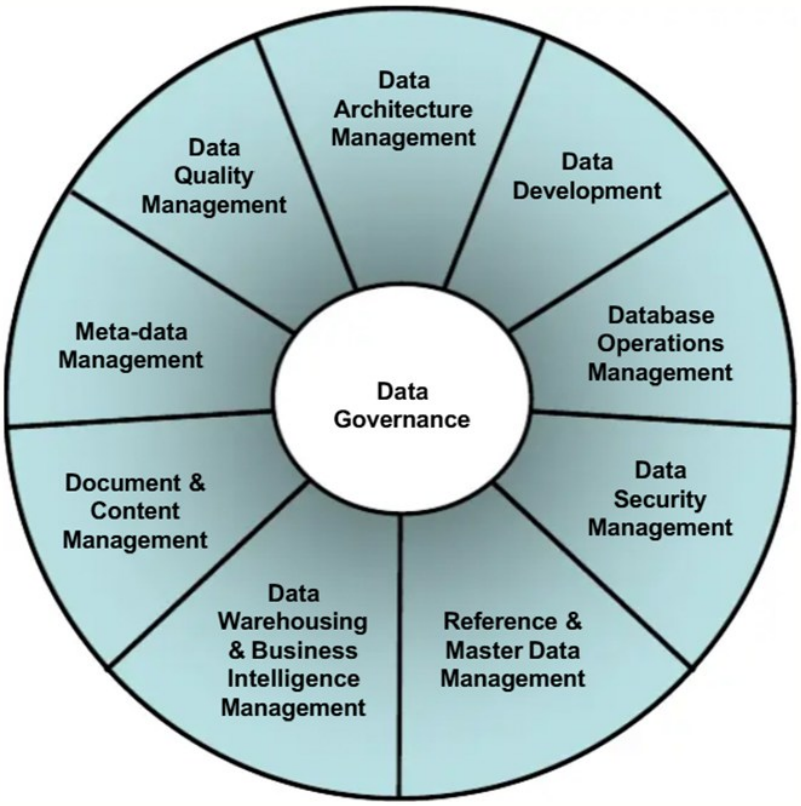

Figure 4. DAMA wheel - data management functions (Earley et al., 2017, p. 36).

The study by Schmuck and Rieder (2025, pp. 177–178) explains that master data management plays an important role in enabling data-driven transformation. Applying MDM is a strategic decision for an organization. It involves both technical and organizational actions to ensure that master data is gathered, managed and used in a consistent and standardized way. Their study introduces a framework, developed using the Design Science Research methodology, which shows how master data management, together with proper change management, supports organizational prosperity.

According to Cleven and Wortmann (2010), it is also important to keep in mind that there is no single approach that works for all MDM implementations. Individual goals call for different MDM approaches and thus, outline four distinct strategies for detecting important data entities and for on-going observation and managing master data quality. These strategies are data-driven, process-driven, problem-oriented and solution-oriented strategies.

2.5 Information infrastructure and logical data modelling

A study by Otto (2012, p. 337) finds that designing and sustaining a master data architecture is considered one of the main tasks of master data management. Developing the architecture is a complex task with many aspects. It includes considering the different needs of stakeholders, handling various technical options and satisfying the demands of multiple types of master data. Using architectural design patterns can help manage this complexity more successfully.

According to Väre (2019, p. 59), for the benefits of master data management to truly materialize, data must be easily accessible, usable and shareable across the organization. Succeeding in this requires a carefully chosen model for the architecture between the systems. There are many different architectural models, each suited to different types of organizations and needs. From a master data management perspective, the choice of architectural model is strongly influenced by the number and diversity of systems used by the organization.

2.5.1 Information architecture

Kotusev et al. (2022, p. 432) state that information architecture (IA) is commonly viewed as a holistic master plan for an organization’s data assets and an important part of enterprise architecture. However, research results indicate that IA should not be understood as a single, all-encompassing plan for knowledge. Instead, it consists of a wide range of loosely connected tools and practices that help organizations handle information (Figure 5).

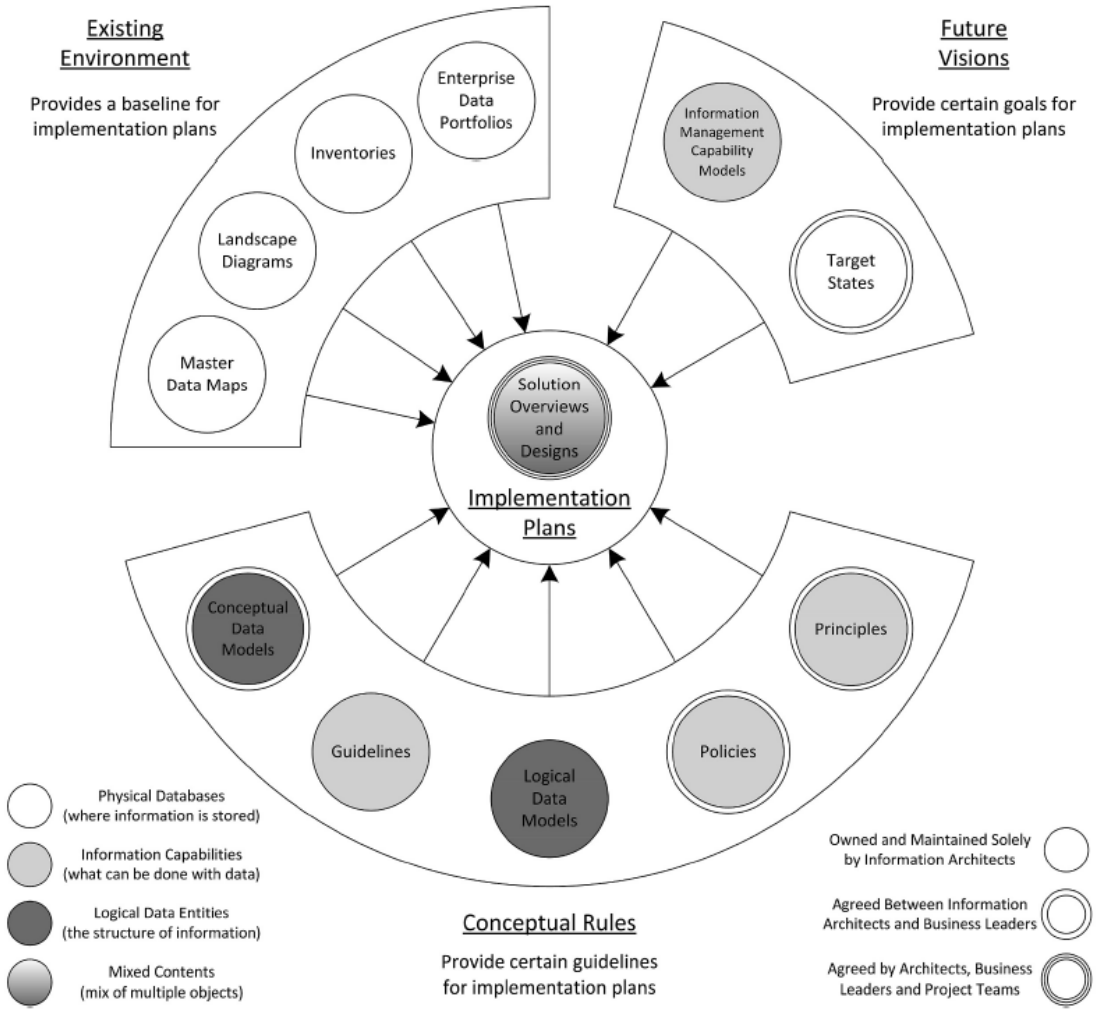

Figure 5. An integrated view of Information Architecture (Kotusev et al., 2022, p. 447).

Moreover, Ding et al. (2025, p. 3) define information infrastructure as the process of simplifying and organising information for users. Beyond simple data organization, it serves as a strategic method for consolidating and integrating information spaces such as channels and platforms. This creates a structured framework necessary for users to find, interpret, and manage information effectively to support informed decisions-making. According to Pansara (2023, pp. 58–59) a pivotal part of master data management is the governance and planning of the master data architecture. Through the design of roles, access, duplication, and data movement the architecture helps to ensure data quality. By definition master data architecture comprises the data application

architecture and a conceptual master data model. The architecture can be further distinguished between source and target systems and the data flows connecting them.

Hillard (2010, pp. 198–199) describes that the development of information architecture should begin with a clearly defined objective aimed at improving the use of information as a strategic asset. Over time, it should evolve into a central and continuous part of business strategy. This responsibility is typically shared between the chief data officer (CDO), who acts as the owner of information assets and the chief information officer, who owns technology assets. A well-designed information architecture comprises three domains that are essential to addressing business challenges and are often illustrated as shown in Figure 6.

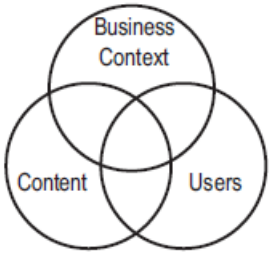

Figure 6. Domains of Information Architecture (Hillard, 2010, p. 199).

According to Hillard (2010, pp. 199–202), the business context forms the basis for all solution development. It encompasses the objectives and operations that guide the organization. The creation of information architecture business context should therefore be closely linked to information governance analysis and aligned with executive leadership's strategic direction. Furthermore, information architecture should begin with an introduction that clearly explains the objectives and provides a description of what the initiative is about. It should highlight strong executive sponsorship and express the concrete issues the organization previously faced, such as poor data quality or ineffective data use. The architecture should then describe the business context in which it operates, the key user groups and the major domains involved. This section explains who uses the information, why they need it and what types of information are most important to the

organization. Next one should explain the organization's information layers and enterprise metadata. A comprehensive system inventory should also be included.

Ding et al. (2025, pp. 138–139) highlight that information infrastructure is an evolving discipline and organizations often struggle to clearly link it to value creation, which makes the active promotion of information architecture crucial. More importantly, information infrastructure plays a critical role in mitigating information overload while ensuring effective use of multiple systems, particularly as digital systems rise together with user expectations. IA is also tied to business outcomes; it provides a framework necessary to translate complex systems architecture into measurable strategic value, thereby fostering strong relationships with organizational stakeholders.

According to Hillard (2010, p. 202), master data should be documented by listing the organization's core data entities and linking them to the systems responsible for creating, updating, or deleting them. This should address any issues related to naming or structures and describe a clear path for resolving these issues. Information governance including roles and responsibilities should be outlined as well as the identification of the organization's key data sets and data set metrics. Information flows and users are important to document to show how data moves through the organization, from source systems to users and business processes. Information users should be grouped and explained how each group uses data and what methods they prefer to access information. The architecture should describe priority standards for high-value data sets and priority investments of where most value is gained from improving information capabilities. Finally, data quality measures should be included and generated automatically wherever possible.

2.5.2 Data modelling

According to Allen and Cervo (2015, pp. 1–33), data modelling is a method for analysing data requirements and designing data structures to support them. A data model consists

of data descriptions and diagrams that describe the needs and design choices (Earley et al., 2017, p. 71). Data models can be categorised into three different types: conceptual, logical, and physical, which can be used to structure and describe business concepts, data entities and elements and their relationships. A key challenge of adopting an existing MDM concept is the impossibility of one universal data model being able to accommodate the diverse needs of every organization. Data within each organization is unique not only in its substance but also in its design. The substance makes data valuable; however, the structural design is closely linked with business requirements.

Earley et al., 2017, pp. 93–101) state that conceptual and logical data modelling largely focus on requirement analysis, whereas physical data modelling is a design-oriented task. The practice of data modelling is multifaceted, encompassing collaboration between individuals and technological systems, without compromising the soundness and care of the data. Well-designed data models capture exactly data needs and clearly communicate quality solution designs.

According to Allen and Cervo (2015, pp. 165–167), the three data models, conceptual, logical, and physical, play a key role in communicating, confirming, and applying an option that accurately reflects business requirements. Too often, organizations underestimate the importance of data models and treat them as secondary artifacts. In reality, data models are critical and should be developed and overseen by modeling experts and data architects. Moreover, the conceptual and logical data models play an important role in verifying business concepts, knowledge and needs. By doing so, they support the development of physical data models and that the implementation is more maintainable and flexible. These models provide a link between business concepts and terminology, as well as their physical location across various data within the organization.

A study by Simsion et al. (2012, p. 151) outlines that data modeling in the context of database development has traditionally been a descriptive activity, in which the reality is captured in a conceptual model and subsequently developed into a logical structure.

However, data modeling is also viewed as a design activity involving flexible needs, creative choices and multiple solutions. According to Kotusev et al. (2022, pp. 438–439), conceptual data models (CDMs) are high-level data models representing the structures and relationships between core data objects crucial for the whole organization. CDMs are timeless and typically focus on core concepts, such as customers, products and orders and allows standardizing the data consistency across different IT systems. Data models can be generalized through industry specific standards that allow for faster implementation and cost saving, while still requiring customization (Inmon & Lindstedt, 2015, p. 188).

According to Earley at al. (2017, pp. 93–94), a conceptual data model presents a high-level visual overview of a business-relevant subject area. It includes essential business entities relevant to a particular domain or activity, describing each of them and their relationships. Through this, the model describes the fundamental business semantics, such as the vocabulary of key concepts and the subject areas represented often correspond to data used in business operations or application functions. Furthermore, conceptual data models are independent of technological and usage contexts, including specific systems or data storage solutions.

According to Earley at al. (2017, pp. 96–101), a logical data model on the other hand provides a comprehensive description of data needs and the data quality rules, typically developed to protect application needs. Logical data models are also not tied to any particular technology or implementation. They commonly go further as opposed to conceptual data models by introducing attributes to defined entities. Thus, the use of organizational naming standards is encouraged to ensure consistent naming of logical data objects. A physical data model addresses how data needs and business rules are implemented in practice by considering technological limitations, application needs, performance demands and modeling conventions.

According to Spruit and Pietzka (2015, pp. 1071–1074), a data model also defines which data is classified as master data, how it is organized, which systems consume it, and where the data resides. In an advanced master data management maturity model, the data model is continuously maintained according to a clear and well-communicated plan that describes the roles, responsibilities and maintenance intervals. All data sources and their system usage are documented, duplicate data is removed, data logic supports scalability and unnecessary systems are replaced.

Srikant (2006, pp. 13–16) states that a logical data model (LDM) is visual depiction of business concepts and the relationships between them, independent of the underlying database technology. Creating an LDM that reflects business requirements is vital, as it supports the effective storage and accessibility of data for the creation of information and knowledge products and reinforces data as a valuable enterprise asset. The logical data model impacts data integration approaches, reference data, metadata, repositories, and enterprise systems such as enterprise resource planning (ERP) and HR systems. Furthermore, logical data models form the base for high-quality data.

According to Srikant (2006, pp. 13–16), creating a logical data model starts with the identification of enterprise data warehouse (EDW) requirements. These, combined with insights from subject matter experts (SMEs) are used to produce a conceptual model that captures the key business entities and their relationships. This is followed by detailed attribute modelling derived from available enterprise application data. The model is then analysed and amended if needed before being transformed into a physical data model (PDM) and subsequently into a semantic data model (SDM). Various challenges can arise when developing a logical data model, including the absence of a shared view for the data warehouse, organizational complexity, the need for current models, and attempting to design a single all-inclusive model.

Väre (2019, p. 71) finds that the purpose of a data model is to describe what different elements master data consists of. It illustrates how these elements relate to each other

and to other business data. Models are a conceptual way of describing how things are connected. Therefore, data modelling should start with identifying the most important concepts within the organization, such as a domain or an entity. For master data concepts refer to a word or expression that has a clear meaning for the organization's operations. The most important concepts usually represent real-world entities that exist and can be concretely defined. Once key master data concepts and their definitions have been identified, it is possible to begin creating the actual data model. It is important to note that a data model cannot be built in isolation from the business and requires extensive involvement from business experts.

According to Väre (2019, pp. 73–75), the highest level of data model is a contextual model. This level describes the organization's master data domains and is often used to portray the areas and scope of an organization's master data management. The next level of a data model is the most significant one in master data management and is the conceptual model. It describes all the concepts within the organization and how they relate to each other. In large organizations, however, the conceptual model can expand quickly, which is why conceptual models are often divided into smaller diagrams. Typically, the structure of the contextual model is used to create separate conceptual models for each master data domain. It is also good practice to include information in the model about how each concept is identified. This ensures that the concept represents a single real-world entity and that it is correctly defined and unique. For instance, when creating a conceptual model, it is discovered that two concepts are identified using the same attributes, it is useful to verify whether they are two different names for the same thing.

According to Väre (2019, pp. 73–75), the most detailed level of the data model is the logical level. This level does not need to be defined immediately in master data management, nor does it need to be created for all master data domains. The logical data model primarily guides the development of technical systems and their integration. The model is thus taken to the logical level when it is needed to support the development of information systems. Defining the logical model is also the most challenging process, as it also

requires a thorough understanding of database logic. The decisions made on the logical level affect also the flexibility of hierarchies and classifications. The model can tie the organization into a stiff structure that is difficult to change once it is implemented, or it can enable adaptability and flexibility to business changes. Thereby, defining the logical level requires an experienced data modeler who understands the business implications of these decisions.

3 Data & Methodology

This section presents the research methodology and describes how the study was conducted. First, the selected research methods and the underlying theories are introduced. This is followed by a detailed description of the research process and execution. According to Helo et al. (2019), there are two types of research problems in the field of Industrial Management, which to some extent limit the choices of available methods. A nomothetical research problem aims to describe the current situation of conditions, whereas a normative research problem seeks to define how the conditions should be organised in the future.

In collaboration with the case company, it was agreed that the study should result in the development of a practical artifact to manage master data across IT systems. The artifact is also used to detect, define, and govern master data elements, which are defined as company specific critical elements. To ensure business continuity and efficient master data management, the research problem of this study is primarily normative, as it defines the desirable conditions that should exist in the future. However, the study also incorporates elements of nomothetical research approach, as the current state was examined through interviews and document reviews.

This study follows a qualitative study research design. In the context of normative research according to Helo et al. (2019) the aim is to determine how processes should be structured in the future. This research was conducted for a case company that requires a clearly defined process to extend Master Data Management (MDM) practices to new business-critical data objects and IT systems. To address this need, the study develops a practical artifact: A Master Data Playbook. This artifact identifies the specific attributes and perspectives required to govern business-critical master data. The study also incorporates nomothetical elements by analysing the current state of governed master data domains, ensuring that case company-specific best practices are integrated into the final solution. The qualitative data supporting this research was gathered through semi-structured interview and analysis internal documents.

3.1 Design science research

According to Vom Brocke (2020, pp. 8–9), Design Science Research (DSR) can be defined as a problem centered paradigm that aims to improve human knowledge through the development of novel artifacts. It strives to give answers to how and why things can and should be designed or organized. According to Cleven and Wortmann (2010, pp. 4–5), the Design Science Research Approach (DSR) focuses on the development of effective solutions for practical problems, i.e. on accomplishing utility. Usually, DSR projects begin with an opportunity or a problem found in an actual application environment, which makes the development of a solution relevant to a certain group of stakeholders.

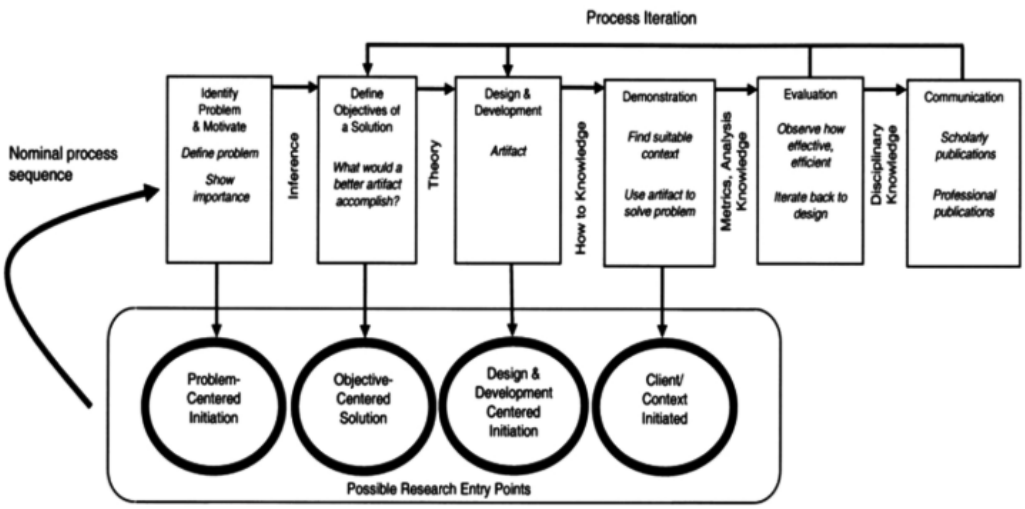

Figure 7. DSR methodology process model (Peffers et al., 2007, p. 54).

According to Peffers et al. (2007, p. 49), design science research method is characterized as an approach to address identified problems through creating an artifact, which includes evaluating designs but also the communication of results (Figure 7). Artifacts may take various forms and include constructs, methods, models, and instructions. Furthermore, the results can also include design knowledge, which offers better understanding of why artifacts improve or disrupt the application contexts.

In their study Otto et al. (2012, p. 397) adopt a DSR approach to advance and estimate a functional reference model regarding master data quality management (MDQM). Reference models are commonly understood as information models that can be applied across multiple situations (Becker et al. 2004; Fettke & Loos 2007 in Otto et al., 2012, p. 397). They aim to provide commonly applicable solution patterns for existing issues. Epistemologically, MDQM reference model constitutes an artifact and an outcome of design-oriented research.

This study aims to create a novel artifact with relevant items, such as data model, information architecture, and data governance practicalities to address the master data management issues across different IT systems in the case company. According to Peffers et al. (2007, pp. 49–50), design science research investigates the gap between the business environment and the academic knowledge base. The research process typically begins with identifying a need through environmental analysis or existing literature. DSR aims to extend the current knowledge base with new frameworks, models, or tools using methods like interviews iteratively developing them.

The business and operational environment of the case company in this study is highly complex. Furthermore, the infrastructure and system architecture are fragmented. Various units across the organisation utilise data for localised purposes, consequently, the organisation suffers from silos where the same data objects are conceptualized differently and driven by different conflicting objectives. From an architectural perspective, there is a lack of visibility into how data objects function with end-to-end processes and how data flows through the value chain.

This complexity is further affected by standalone applications that currently lack a formal data governance framework. This fragmentation forms the basis of the research problem and the objective to create a unique “playbook” to guide the organization in building a unified view of different data domains across all IT systems. The knowledge base for this study is derived from established theories, previous research, and internal

documentation and are further supplemented by interviews with key stakeholders within the company. By combining these elements, a novel item can be designed.

3.2 Interviews

In this study, semi-structured interviews were conducted as the primary data collection method. Often referred to as thematic interviews, this approach is one of the most widely used research methods, as it allows the researchers to focus on essential themes and topics necessary to address research questions (Vilkka, 2025). Following the approach outlined by Saunders et al. (2007, pp. 312, 315–316) an interview guide was developed to provide structure while allowing for open dialogue (Appendix 1).

To capture a diverse range of viewpoints and to ensure comprehensive understanding of the case company's master data activities, participants were selected purposefully to reflect a wide range of professional roles, industry backgrounds, work experience, and academic qualifications. The interviews involved various subject-matter experts from different areas of expertise, such as information architecture, master data management, data ownership, concept owners, and ownership and head of data management and group business processes. In total, eight interviews were conducted over a three-week period in February and March 2026. The interview sessions were held either face-to-face or via Google Meet.

The interview structure was built around five core themes:

1. Data Governance and Maturity Assessment: This theme focused on leveraging the existing maturity to build a scaling foundation
2. Data Modelling, Information Architecture, and integrations: This theme focused on defining the unified logical reference model across IT systems
3. Data Quality and Performance: Focus on this theme was on defining the data quality standards and definition of success for the playbook

4. Data Lifecycle and System Roles: Theme focus on mapping the “Single source of truth” and master data lifecycle
5. Playbook Utilization, Scalability, and Adaptation: This theme focused on ensuring the playbook serves as a practical tool to be utilized

The interviews were conducted in both Finnish and English. To support thorough preparation and enable more in-depth discussions, the interview themes were shared with participants in advance. With the participants’ consent, the interviews were recorded and automatically transcribed. The transcripts were carefully reviewed and validated against written notes to ensure accuracy. In total, the data collection process yielded approximately 420 minutes of recorded material and approximately 40 pages of summarised transcripts. Transcribing the interviews immediately after each session enhanced the reliability and consistency of the findings.

Although master data management has been widely studied and various tools have been developed to support it, each organisation has its own nature and way of working and culture. These organisational characteristics may either act as enablers or barriers in effective implementation. Consequently, the interview structure was intentionally designed to cover a broad range of themes from multiple areas. No single interviewee was expected to possess deep knowledge across all areas, but nevertheless, each perspective was considered valuable in contributing to the study.

4 Results

This chapter presents the results of the empirical part of this study, derived from semi-structured interviews and an internal document review. The analysis follows an abductive content analysis approach, where empirical observations are compared against the theoretical framework of successful Master Data Management (MDM). The chapter is structured to first provide a comprehensive view of the case company current state of master data management, followed by a gap analysis against the academic literature, and finally defining the design requirements for the Master Data playbook.

4.1 Current state of Master Data Management in the Case Company

The empirical evidence gathered through interviews and documents reviews revealed a divided landscape within the organisation's data management practices. On one hand, the company possesses a matured and disciplined core environment focused on its primary ERP system, but this stability is contrasted by a significant degree of fragmentation and administrative obscurity once data moves beyond these established boundaries. This section analyses the current state by exploring the strengths of the internal ecosystems, the underlying governance vacuum and the systemic challenges posed by an architectural strategy.

A primary finding of this study is that Master Data Management (MDM) operates with a high degree of precision when considering the core ERP environment. Interviewees consistently highlighted that for core domains – specifically for Finance, Customer, Supplier, and Material data the organisation has achieved a good level of process maturity – supported by well-defined process maps, clear RACI matrices, and comprehensive standard operating procedures. Success is largely attributed to the rigid, rules-based constraints of the ERP's Master Data Governance module, which enforces strict validation and clear workflow triggers. In these specific areas, the roles of Master Data Managers are not

only well-defined but also actively operationalised, providing the organisation with a reliable foundation for its primary business transactions.

> *"It can be stated with confidence that the Master Data Management within core ERP is exceptionally rigorous; the underlying processes and governance frameworks are clearly defined and strictly maintained. We also did a lot of development." (Interviewee 2)*

This can also be viewed as operational excellence, which demonstrates that when technical constraints are aligned with clear process ownership, the resulting data quality remains exceptionally high. Despite these rigorous frameworks, interview findings indicate that operational responsibilities are concentrated within a small group of employees. This may lead to heavy workload threatening the established standard ways of working and potentially decrease the data quality and broader risks for the case company.

> *"We need a data governance framework that accommodates master data elements. The company must understand the different data types: Master Data, Transactional data and Reference data. If our team were responsible for all these data points the list would be extremely long. This would mean either expanding the team resources or building tools for quality control, such as reporting that compares different data sources."* (*Interviewee 8*)

Despite the technically rigorous ERP system, the analysis identifies a critical governance vacuum as soon as data crosses departmental or systems boundaries. While the process of data entry is often well-documented, the accountability for the data's long-term accuracy remains alarmingly ambiguous. A recurring concern among stakeholders is the lack of a global data governance framework that would provide an end-to-end responsibility across the company. This ambiguity is deteriorated by high employee turnover and frequent internal restructuring, which has led to a reliance on tacit knowledge. When errors occur such as incorrect tax parameters or inconsistent product specifications,

employees are in the loop of finding the right person rather than following a standardised escalation path. This suggests that the company's current MDM success is localised rather than global, leaving wider organisational reporting and strategic data use at risk.

> *"We are good at creating the data but then the rest is forgotten… long-term accountability is often compromised due to responsible individuals changing roles in rapid cycles eventually breaking the cycle of ownership and knowledge." (Interviewee 6)*

The company has defined a Data Governance Principle. The principle establishes the data governance responsibilities for all case company employees, who interact with the data. The purpose is to set expectations for maintaining data quality, confidentiality, integrity and availability of different data sets such as reference, master and transactional data. It describes a high-level concept of taxonomy, roles and responsibilities, data quality, risks, access management, information security, backup and archiving procedures, and definitions of transactional, master and reference data and calculated entities and data sets.

The organisation's strategic move towards best practices on IT architecture has unintentionally fostered a landscape of disconnected data silos. While the official understanding and policy designates the primary ERP systems as the "single source of truth", the reality of the operations are far more complex. Different business units have developed their own localised versions of the truth based on their immediate operational needs. For example, sales departments may view CRM as their primary authority, while production units rely on specialized platforms used for operative execution. This fragmentation becomes most visible in the integration layer, where data quality is described as blurry. Certain specialized domains, such as enterprise asset management and raw materials, have largely been left to operate in isolation, suffering from slow update cycles and poor data health. This architectural vacuum underscores the need for a unified playbook that can provide a harmonised reference on how to organise master data across different IT

systems and establish an enterprise wide standard for information modelling and data lifecycle management.

> *” To my knowledge we only have a unified definition for customers, but not for a product or are these concepts widely known. We should have for product and for raw materials… regarding the data models some of them are done, but we are missing the logical level and then also the reviews and validation haven’t been done.” (Interviewee 1)*

The current state of the case company documentation and what was raised by the interviewees in the discussions is mixed. Many of the interviewees feel that some key elements are there, but overall governance driving force is missing in some sense due to lack of resources, time and allocation. For some domains there are actual driving forces, for example for materials and there is a good concept built around the domain. A similar observation can be made from the finance function perspective, which is often the most stable element within the organisation. In most cases, business and particularly finance already have well-defined reporting structures and hierarchical frameworks in place. Finance functions are responsible for monitoring the profits and costs, as well as identifying the need for changes when needed. Governance structures within finance are typically well established and integrations between systems and organisations tend to operate reliably.

One of the most pressing and imminent issues raised in the discussions is that the company does not have a unified data glossary or documentation, which would describe the basic information. The typical element is that the ones who are working with the information usually know who to contact and what is needed. Unifying understanding of attributes are often found through repeatable discussions together with stakeholders. However, even the master data framework lacks the overall view of technical integration and data model pictures. The documentation is scattered and it becomes hard to update whenever there is change needed since it might be stored in several places. The most

well-kept are the end user instructions for requestors on how to request new master data records or change existing master data.

> *"While there is no official data glossary, most data definitions can be found from end user instructions and from core ERP system data model design documentation." (Interviewee 3)*

> *"I don't think there exists any data glossary, but I have added descriptions for all the necessary fields in the end user instructions, also some higher-level definitions exist such as what a cost centre means."(Interviewee 4)*

For some of the domains there is uncertainty regarding who owns the data and what is the governance model. Without clear responsibilities and accountabilities, the operational management usually falls under one team or person. Different teams and units have their own needs and can also be called as data consumers, requesting new attributes to be developed without having a unified view of the domain.

How the master data is considered across different teams differs based on requirements, ranging from financial reporting, operational purposes, and for tax compliance reasons. Consequently, a unified view of master data is missing. For example, because a single customer entity includes attributes which are controlled and dictated by regulatory standards or separate teams, centralised data management becomes difficult. For example, a customer has taxation attributes, which are governed by the tax team, while payment methods are owned by accounts payable - details that may remain unclear for the procurement department. This leads to a concern where any updates, new attribute configurations, or rule changes may have downstream impacts that remain vaguely understood or entirely undetected.

> *"In our current environment, we are close to a solid master data model regarding customers, specifically within the Order-to-Cash (O2C) cycle. Looking at the*

> *Customer entity and our business model as a whole, a customer and a supplier can be the same partner or entity. Yet, the concept of a 'Business Partner', which would solve this, doesn't really exist in our systems. Effectively, only a third of the model is well-executed; the rest isn't. This isn't anyone's fault, nor is it 'wrong' as such. It's simply that our current as-is description of the process and information architecture is so incomplete that a better model isn't even possible yet. We lack the necessary people and resources. I've introduced information architecture to the company myself, and the positive takeaway is that both the business and process sides are now finally discussing how data should flow and what priorities matter in this concept."(Interviewee 1)*

The document analysis revealed that while the case company possesses formalised process definitions, which include established Master Data Manager's roles within those frameworks, the overall documentation related to master data is highly fragmented. Centralised repositories or "single source of truth" for MDM instructions are non-existent. Furthermore, this fragmented documentation was also identified as a significant obstacle for data management. Respondents noted that information is scattered across IT systems, local folders and the tacit knowledge of subject matter experts.

> *"Currently, finding the correct procedure for data maintenance requires knowing exactly who to ask, or asking many people before finding the correct person, as there is not a central place to verify the standard process." (Interviewee 1)*

The lack of centralized repository leads to inconsistent practices across different business units and IT systems, increasing the risk of silos. Overall, the key challenges include the lack of a centralized driving force to harmonise efforts across organisational silos. The architectural emptiness of information modelling, and the fact that "single source of truth" designations argue with obsolete and localised systems and operational realities. In addition, from an information architecture perspective, a key challenge is the absence

of conceptual, logical and physical data models, resulting in the lack of system independent definitions and data content.

4.2 Gap analysis: Reality versus Literature

The synthesis of empirical findings against established academic literature reveals a disconnect between the organisation's operational reality and the strategic ideals of Master Data Management. While the company has successfully implemented localized controls, it lacks overall structural integrity described in the Data Management Body of Knowledge (DAMA-DMBOK). This gap analysis focuses on three critical areas: the absence of a centralised governance framework, the formal information architecture, and the discrepancy between documented processes and the actual data ownership.

In the academic literature, a successful MDM initiative is predicted on a centralized governance framework that orchestrates data across all business units. However, the case company currently operates in a state of fragmented excellence. The interview results overall present the same conclusion; the company lacks a framework of controlling the overall data management framework. For some of the areas data ownership is missing and in other domains it is working well. According to Earley et al. (2017, p. 68) emphasis on ownership is important and according to this playbook should cover a relatively clear role matrix. It becomes a painful and heavy burden, if the roles have not been decided to operate efficiently.

The interviews confirm that while individual domains like Finance or Materials have high maturity, there is no global driving force or steering committee to harmonize these efforts. According to Earley et al. (2017, Chapter 3) there is natural separation of duties between Data Governance (oversight) and Data Management (execution), where governance ensures data is managed by exercising authority and decision-making and data management performs the actual activities. Furthermore, without formal role matrixes, these boundaries blurs.

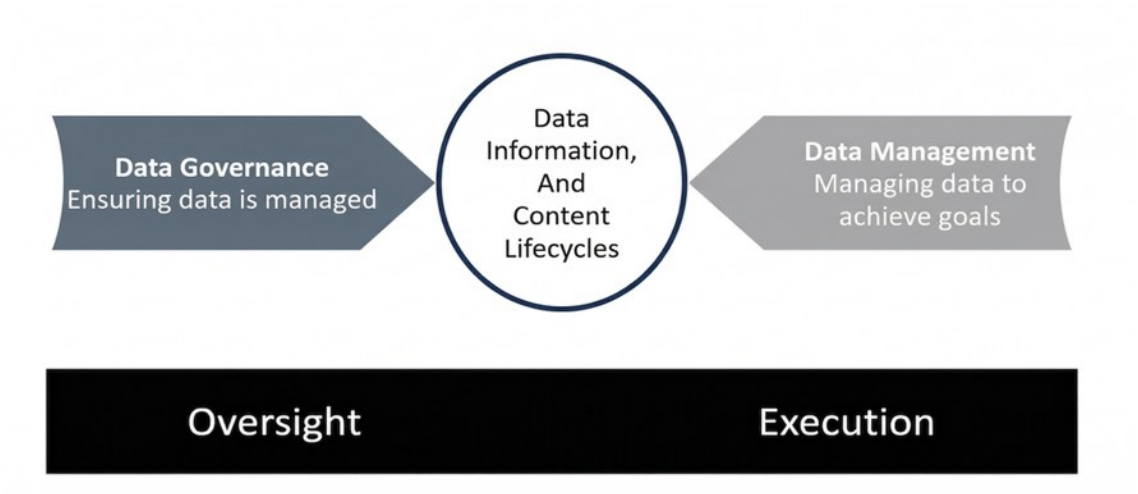

Figure 8. Data Governance and Data Management (Earley et al., 2017, p. 72).

The empirical phase of this study revealed that the case company's Data Governance Principle serves as the principal framework intended to direct the organisation's overall data governance. Table 3 was created to depict what the case company data governance principles are versus what the academic literature is instructing to have to be an organised and successful data governance community. This comparative approach facilitates the development of data governance within each individual category. In addition, the table represents a gap analysis for each category, highlighting missing or partially covered components relative to the DAMA-DMBOK framework. Identifying these elements for data governance principles reduces the risk of structural deficiencies in the overall data governance strategy, while further improves critical elements such as data quality, accountability, efficiency and data asset valuation. Furthermore, the table focuses solely on the categories identified as missing or partially covered in the case company's data governance principle, excluding data governance areas that are well-defined are left out as they are well defined within the principle.

Table 3. Case company data governance principles versus DAMA-DMBOK (Earley et al., 2017, Chapter 3).

Category	DAMA-DMBOK requirements	In company principle	Gap/Missing component
Strategy	Integration with Enterprise Business Strategy	Partial	Lacks a formal DG roadmap
Organisation	Formal Bodies: Steering committee, DG Council & Stewardship	Partial	Mentions Process & IT Committee and owners but lacks a formal Data stewardship program and dedicated DG Council
Metadata	Management of Metadata	Missing	There is no mentioning of business/data glossary to ensure common definitions across the company
Valuation	Calculating the value of data as an asset	Missing	Defines data as an "asset" but provides no method for data asset valuation or measuring the impact
Lifecycle	Data lifecycle management	Partial	Covers "CRUD", archive and retention, missing disposal of data
Change Management	Addressing the cultural shift and data literacy	Missing	Awareness, process to improve data literacy and Organisational change management (OCM)
Metrics	Measuring DG program effectiveness and KPIs	Missing	Scorecards for DG effectiveness such as policy compliance rates, stewardship activity
Issue Management	Formalised escalation path for data issues and conflicts	Missing	Lacks a documented escalation path for issues to resolve conflicts
Regulatory	Proactive assessment of global data regulations	Partial	Focused only on GDPR, lacks broader framework of specific regulations

Analysis reveals that the case company document is a robust operational principle for data quality, including role and data definitions, risks and access management. However, compared to existing literature it is currently missing a layer of strategic guidance and goals. To be aligned with established academic literature the company should also consider and expand into Data Stewardship as the human layer, Metadata management as the context layer and Data valuation as the financial layer of data governance (Earley et al., 2017, pp. 68–95).

A critical finding from the gap analysis is the imbalance between process modelling and information architecture. The company has invested in documenting how operative processes flow i.e. process modelling, but it lacks a formal representation of what data truly represents. Academic theory emphasizes that mature Master Data Management requires three layers of data modelling to represent data: conceptual, logical, and physical (Earley et al., 2017, Chapter 5).

In the case company, these three layers are currently incomplete or disconnected:

- Conceptual modelling: There is no shared understanding of key data entities in the case company. The organisation lacks a shared conceptual understanding of each data domain including items such as data glossary that would define key entities such as “Customer” or “Product” consistently across the organisation.

- Logical modelling: The relationships between data attributes and business rules are not documented within the case company. This results in different teams assuming different definitions for the same attributes. This reinforces the siloed understanding identified in this current state analysis.

- Physical Modelling: While physical schemas exist within separate databases, they are not mapped to a higher-level enterprise architecture. The data is managed within the silos of specific IT systems and by SMEs rather than as global assets.

> "*We should go towards conceptual level of understanding rather than system level of understanding of each of the master data domains… Focus on data content not the systems. We should understand and share the knowledge of what a data set is and the information related to it."* (*Interviewee 8*)

This architectural imbalance directly justifies the need for the MDM playbook to include a dedicated section for information architecture. Without any specific guideline for information architecture, the company will continue to struggle with data consistency issues across the system landscape and integrations or even when adopting new data domains under the MDM framework. By incorporating information architectural principles into the playbook, the company can transition from system-centric view to a domain-centric model. Furthermore, this would enable data domains to integrate with business models and ensure that data contents remain as the priority while adopting a single source of truth regardless of the IT infrastructure.

Literature on information governance Hillard (2010) identifies data ownership as the cornerstone of any sustainable business and senior leadership should be accountable for data accuracy with direct reporting responsibility. Furthermore, data ownership requires a budget and incentives not only for risk mitigation but also for using data in a creative way to drive business growth. Within the case company, these roles are often mixed or undefined. On the shared global process owner information, the company has defined business owners, which might be misunderstood or are not even known. More importantly, what is missing is the clear role matrix and assigned persons to each of the data sets.

The empirical finding shows that data corrections are reactive and the responsibility often falls by default to Master Data Managers or IT teams, rather than as a correction process driven by data owners who also own the business risk of those data sets. This gap between theoretical accountability and practical execution results in reactive and periodical maintenance culture, where one-time clean ups are favoured over the

continuous, ownership driven improvement cycles recommended by the literature and standards (Earley et al., 2017 p. 424). As highlighted in Table 3, the lack of formal escalation path for issue management means that when conflicts arise between systems or domains there is no "steering committee" to provide a final decision, leaving operational teams to make business critical choices by default.

> *"In my opinion, you must have the person who owns the core knowledge of the domain, and it includes the reference, transactional data and even the documentation. You must have the person who produces the data, controls the data such as Master Data Managers and in the end, you must also have the data consumers. All in all, these persons form the coalition i.e. this kind of a virtual matrix organisation and profiles who continuously process these data quality topics."* (Interviewee 7)

The case company data quality elements have been reviewed through the documentation and interviews. What was imminent is that all interviewees felt that the data quality measurement regarding master data is somewhat periodic and manual. In other words, the organisation lacks a demand from business stakeholders, resulting in absence of data ownership. Furthermore, there is no actionable and systematic measurement of data quality. Earley et al. (2017, pp. 464–465) highlight that organisations should recognise high quality as an asset, which allows you to reduce costs, improve efficiency and productivity and furthermore protect the company's reputation, often also driven by regulatory requirements. Poor quality data is often associated with direct costs and inability to invoice correctly resulting in revenue losses. Manual and periodic nature of data quality corrections is partly a result of low data literacy – employees may not know why quality is crucial.

> *"The overall feeling is that it is manual and periodic. It should be more automated and continuous, and it should be built into business processes. Periodic comparisons can be done occasionally, but automation should be the goal… And gradually*

> *we are understanding the issues better. Previously, we have lacked these structures to prevent future problems or they have been obscure. Now as process ownership is taking deeper root, a mechanism is emerging where learning of the issues is embedded in ways of working rather than issues remaining as obsolete observations." (Interviewee 8)*

Data quality standards are somewhat monitored in terms of the creation and approval process flow of master data creation within the case company. The data quality originates from the creation phase and is controlled with rules of creation in one system. This system is considered while not fully recognised as the single source of truth and the data is also replicated to other business operations relevant target systems. However, there are not clear 360-degree visibility views of different domains. For the customer domain for instance there is a standalone system, which is treated as a CRM, but it has been implemented with different objects related to the sales process. While the company has focused on the creation of master data, disposal and archiving standards should also be addressed as they might lead to increasing storage costs.

As a summary the existing data governance principle within the case company can be further developed and extended by transitioning from operationally excellent to a more strategically excellent understanding while driving data excellence. This involves defining physical frameworks to higher-level logical reference models and implementing unified role matrixes that assign accountability to data owners rather than to operational execution teams. Furthermore, an exceptionally rigorous process of established operational level of master data management can be referenced for further adaptation of different master data domains.

4.3 Identification of design requirements for the master data playbook

Based on the abductive content analysis, the following design requirements were identified to guide the development of master data playbook:

- **Centralisation**: The playbook must serve as a centralised “single source of truth” guideline for all Master Data Management functions to address scattered documentation and organisational silos. The target is to eliminate scattered documentation.
- **Pragmatism**: The playbook must serve as a practical document meaning it should be integrated into the daily active development.
- **Accountability**: The playbook must include clearly defined roles and responsibilities across systems with Global RACI matrixes.
- **Standardisation**: The playbook must serve as a guideline for driving common terminology across systems and provide instructions for data modelling and harmonize terminology.
- **Compliance:** The playbook must serve as a guideline for linking data management to legal requirements, such as GDPR, ISO standards to ensure proactive approach to regulatory compliance.
- **Scalability**: Overall, the playbook must serve as a repeatable process and reference for MDM activities for existing and new master data domains.

To address centralisation issues the analysis revealed that documentation is scattered across systems, local folders, cloud servers and SME dependent tacit knowledge. The playbook must serve as a global guideline gathering all the relevant documentation under one framework with the linkages to relevant systems and repositories. The playbook is not only about creating rules but serves also as a knowledge management tool.
A significant risk was identified within the case company where operational master data activities were focused on a small group of employees due to the company lacking a clear global driving force or steering committee and data owners. The playbook must include a uniform RACI matrix reference that clarifies roles across all the systems and

domains from a more technical execution to business leaders. Accountability as such would be better to be embedded in a coalition of data owners who eventually own the business risks and are the driving force for continuous improvement. Clear escalation paths should be described for resolving data issues and conflicts between systems, eventually also improving data quality.

The analysis found that while current business processes are well-documented, the underlying information architecture is largely missing. To prevent unnecessary confusion over terms such as customer, business partner, account or material and product, the playbook must provide guidance in building a centralised data glossary for different attributes ensuring company level common definitions and mapping across all IT systems, clarifying the relationships between terms. Data modelling must be covered and provide understanding of different layers of the architecture. Eventually, this would help the company to move away from system-level thinking towards data-centric thinking.

To ensure alignment with all the regulatory requirements and industry-specific standards, compliance activities should be embedded within the initial design of the master data process. The company would be moving from reactive compliance checks and recurring audit findings to a more proactive approach, building compliant data from the start. This also allows companies to react to regulatory changes more efficiently.

Currently, the organisation struggles to systematically detect new data that should be governed under the MDM framework, resulting in uncontrolled models in different domains. The playbook must define a specific set of criteria to identify when a new data element qualifies as a Master Data and requires governance. Criteria could be for instance a data set, which is shared across multiple business units and serves as a critical element driving the business with financial aspects. At the same time the playbook serves as a reference for existing master data domains to review whether the framework is fully implemented, eventually improving master data management activities and referencing best practices from other domains. The playbook must also provide a

repeatable reference process for modelling and integrating new domains into the existing enterprise architecture.

4.4 Master Data Playbook

The proposed artifact, the Master Data Playbook, is designed to be a practical guideline that addresses the identified gaps from the interviews and documentation analysis. The concept of a "Playbook" was chosen as it is already familiar to the organization's personnel working between the business and IT organisation, making it easier to be adopted effectively. The idea of a playbook has been proposed as a Guideline for Master Data Management within the case company. Several years ago, a playbook was developed to guide sales and supply operational flows. In addition, the organisation has introduced other playbooks for various operative purposes. Against this background and based on the gap analysis and opportunities identified in this study, there is a clear need to develop an MDM playbook. It serves as a guideline for MDM, covering all relevant topics identified in the literature and complemented by the case company perspective.

To address the identified requirements, the proposed structure and content of the playbook are presented in this section. Overall, the playbook context mirrors the DAMA MDM wheel (Figure 4), while being adapted to the specific operational maturity of the case company. First, the playbook should include a short introduction of the purpose and strategy. It should describe the process for establishing an overall MDM framework covering all data domains treated as master data. In addition, it is considered good practice to explain why the document has been created and how it supports the management of master data domains within the context of the case company. The introduction should also provide guidelines and references to the key items that need to be addressed, clarify the intended audience within the case company, and explain the content and how it is organised. Furthermore, the business value of clean and well-governed data should be addressed and how process standardisation and getting the fundamentals right are essential for process efficiency and implementing artificial intelligence in the future.

As a process-driven company, there should be a clearly defined process describing how to establish a management structure for new master data domains. This process should begin with a discovery phase, where it is determined what constitutes master data and what the business rationale for the case is. For example, the trigger may be a recurring issue with a dataset in a specific system that is not integrated with other systems and for which data ownership or maintenance responsibilities are unclear to the relevant stakeholders.

Next phases should include an initiation stage, during which it is decided whether a separate project is needed for establishment and finalizing the business rationale behind the case. This phase may also include a formal approval requirement from a data governance council or steering group. The following phases should address items such as specifying the need and taking all relevant functions into account, and cover implementation and planning activities and confirming the exact go-live date for the new management structure.

Key terminology should be clearly defined, including what is considered master data and how it differs from reference data and transactional data. In the context of the case company, the importance of building a unified data glossary cannot be overstated. The glossary should be transparent and accessible across the entire organisation, serving as a foundational guide that aligns both IT and business stakeholders and ensures a shared understanding of the same language, data concepts and attributes. To ensure that these definitions remain consistent as data flows through integration layers, metadata and naming standards should be systematically reviewed and maintained. These standards should provide company-wide naming conventions and require the documentation of both technical and functional metadata, eventually transforming the technical attributes to clear business assets.

The playbook should include a description of roles and responsibilities, supported by a unified RACI matrix. While these roles may vary across departments or data domains,

they should operate under the established data governance steering group. It is critical to ensure a common understanding and the allocation of sufficient resources for the roles, responsibilities, and accountabilities within the context of operational data stewardship and data ownership within the case company. Insufficient resourcing poses also significant risks leading to reactive issue management and overreliance on tacit knowledge. Therefore, clear escalation paths and accountability structures must also be defined.

Unified data modelling is a key topic in the context of the case company and should be carefully addressed and reviewed. Building the information architecture framework, including descriptions of conceptual, logical and physical data models should be explained. It is recommended that all data models are formally defined, reviewed and approved to ensure consistency across system boundaries. Furthermore, this transparency should be paired with a data quality framework that extends beyond periodic monitoring towards more continuous management.

A data quality framework should be established, where topics such KPIs and monitoring requirements are defined. While many of these are domain-specific, the general data quality framework metrics can be described, such as process efficiency from data creation to approval, lifecycle standards, system-level validation rules, and enforcing building automatic monitoring reports. Under the data quality framework, regulatory compliance and applicable industry standards should also be explicitly addressed, as they are the key drivers behind data quality rules. The playbook should highlight that domain-specific compliance requirements are listed and defined. Although these requirements are typically known by domain-specific stakeholders, formal documentation is necessary to support business continuity and to ensure that compliance obligations are secured and actively reviewed.

A strong connection to the business context should be established, as the playbook is intended to provide linkages to other relevant systems and internal documentation. This

ensures that master data domains are not managed in silos. It is strongly recommended that master data domains are connected to business processes and models, enabling users to understand the real-world context and rules of the dataset, moving away from system-centric view. Ultimately, this approach would enable the playbook to be up-to-date and create a continuous, self-balancing cycle of governance that would adapt organically as the business grows and new data domains are identified and incorporated into the MDM framework.

5 Conclusions

The purpose of this thesis was to identify the most effective practical tools and practices for the case company to govern master data within a global framework. This was done by combining literature review, documentation analysis and semi-structured interviews to build a practical Master Data Management (MDM) playbook for the company to follow. The literature review was conducted to understand the theoretical foundation around MDM, information architecture and data governance, and to support the empirical research carried out within the case company. By integrating insights from the literature with empirical findings from interviews and document reviews, the study was able to address all the three research questions.

The study addressed three research questions, the first being:

1. *What are the key challenges and critical information architecture considerations in achieving uniform master data modelling and governance across system boundaries?*

The study identified that the primary obstacles to uniform modelling stem from siloed documentation and the absence of a standardised integration layer across systems. Achieving successful data governance across system boundaries requires a shift from a system-centric approach to a domain-centric perspective, completed with a strong information architectural focus. Finally, this transition is secured by establishing accountability and defining role matrices across all the data domains.

The second research question the study addressed was:

2. *How can the existing Master Data governance model and current master data domains (e.g., Customer and Product) be extended and applied to new master data domains and across different IT systems?*

The findings indicate that the existing governance models of the core ERP data domains can be scaled by adopting rigorous and consistent creation processes throughout the master data domains. This includes defining a unified set of core data attributes that remain constant across systems while allowing local system-specific extensions where necessary. Furthermore, data governance steering group ensures consistent adaptation of established data governance principles across the organisation.

The third research question the study addressed was:

3. *What kind of functional structure and practical content are required for an effective Master Data Playbook designed to guide Master Data Management (MDM) practices within the organization?*

An effective Master Data playbook must function both as a strategic roadmap and as a practical guidance document. It requires a functional structure comprising overall role definitions and responsibility definitions, quality standards and KPIs, and centralised architectural models to ensure companywide consistency in master data management practices. The following recommendations are proposed for the case company:

1. Shift to domain-centric approach: Transfer the accountability and ownership from IT-system owners to business-driven data owners who will manage data domains across the system landscape and throughout its lifecycle.
2. Implementation of the Playbook: The developed artifact should be used as a mandatory tool when onboarding new master data domains, IT-systems and business units to ensure alignment with the MDM framework from the start. Furthermore, existing domains should be reviewed against the playbook periodically to ensure their alignment.
3. Standardization of definitions: Define a company wide data glossary and architectural data models to prevent unstructured domains and organisational silos.

4. Proactive monitoring: Establish monitoring KPIs and metrics for data governance and data quality, transition from reactive data fixing to proactive management and ensure satisfaction and understanding of data governance activities among the users.

Finally, existing literature highlights that MDM programs can support the adoption of new master data domains under the MDM framework and deliver clear benefits on efficiency and financial aspects. The practical contribution of this study was to develop a Master Data playbook tailored for the case company and the subject matter experts in applying MDM principles. The playbook will provide high-level guidance for governing new master data domains and serves as a tool for reviewing existing domains to identify potential gaps or missing elements. This way, the playbook helps direct organizational focus and effort toward the most critical aspects of master data management.

While this study provides a structured MDM framework for the case company, potential future research could focus on the investigation of automating the playbook's standards. This could be done through various MDM software solutions by exploring implementation of artificial intelligence and how automation could further improve efficiency and data quality. Furthermore, analysis on the long-term effects of adopting the playbook standards within the case company culture and people should be investigated. Primarily, this case study represents a valuable contribution for the case company by introducing a structured governance approach to master data management. The study enhances organizational understanding of the Master Data Management framework and allows the company to further extend its master data activities as business requirements and data needs evolve.

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

Adapa, C. S. R. (2025). Enterprise Master Data Management: Trends and Solutions. *European Journal of Computer Science and Information Technology*, *13*(12), 96–105. https://doi.org/10.37745/ejcsit.2013/vol13n1296105

Baghi, E., Otto, B., & Oesterle, H. (2013). *Controlling Customer Master Data Quality: Findings from a Case Study*.

Bourne, K. C. (2014). *Application administrators handbook: Installing, updating and troubleshooting software*. Morgan Kaufmann.

Brous, P., Janssen, M., & Vilminko-Heikkinen, R. (2016). Coordinating Decision-Making in Data Management Activities: A Systematic Review of Data Governance Principles. In H. J. Scholl, O. Glassey, M. Janssen, B. Klievink, I. Lindgren, P. Parycek, E. Tambouris, M. A. Wimmer, T. Janowski, & D. Sá Soares (Eds), *Electronic Government* (Vol. 9820, pp. 115–125). Springer International Publishing. https://doi.org/10.1007/978-3-319-44421-5_9

Cleven, A., & Wortmann, F. (2010). Uncovering Four Strategies to Approach Master Data Management. *2010 43rd Hawaii International Conference on System Sciences*, 1–10. https://doi.org/10.1109/HICSS.2010.488

Dahlberg, T., & Nokkala, T. (2015). A FRAMEWORK FOR THE CORPORATE GOVERNANCE OF DATA – THEORETICAL BACKGROUND AND EMPIRICAL EVIDENCE. *Business,*

Management and Education, *13*(1), 25–45. https://doi.org/10.3846/bme.2015.254

Das, T. K., & Mishra, M. R. (2011). A Study on Challenges and Opportunities in Master Data Management. *International Journal of Database Management Systems*, *3*(2), 129–139. https://doi.org/10.5121/ijdms.2011.3209

Ding, W., Lin, X., & Zarro, M. (2025). *Information Architecture and UX Design: The Integration of Information Spaces*. Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-72138-0

Earley, S., Henderson, D., & Data Management Association (Eds). (2017). *DAMA-DMBOK: Data management body of knowledge* (Second edition). Technics Publications.

Gregory, A. (2011). Data governance — Protecting and unleashing the value of your customer data assets: Stage 1: Understanding data governance and your current data management capability. *Journal of Direct, Data and Digital Marketing Practice*, *12*(3), 230–248. https://doi.org/10.1057/dddmp.2010.41

Guerreiro, L., Martins, J., Do Rosário Bernardo, M., Mamede, H., & Branco, F. (2025). Impact of a Master Data Management Framework to Trigger Data Governance Maturity: A Systematic Literature Review. *IEEE Access*, *13*, 190644–190667. https://doi.org/10.1109/ACCESS.2025.3627855

Haneem, F., Kama, N., Taskin, N., Pauleen, D., & Abu Bakar, N. A. (2019). Determinants of master data management adoption by local government organizations: An empirical study. *International Journal of Information Management*, *45*, 25–43. https://doi.org/10.1016/j.ijinfomgt.2018.10.007

Hannila, H., Silvola, R., Harkonen, J., & Haapasalo, H. (2022). Data-driven Begins with DATA; Potential of Data Assets. *Journal of Computer Information Systems*, *62*(1), 29–38. https://doi.org/10.1080/08874417.2019.1683782

Haug, A., Staskiewicz, A. M., & Hvam, L. (2023). Strategies for Master Data Management: A Case Study of an International Hearing Healthcare Company. *Information Systems Frontiers*, *25*(5), 1903–1923. https://doi.org/10.1007/s10796-022-10323-z

Haug, A., & Stentoft Arlbjørn, J. (2011). Barriers to master data quality. *Journal of Enterprise Information Management*, *24*(3), 288–303. https://doi.org/10.1108/17410391111122862

Haug, A., Stentoft Arlbjørn, J., Zachariassen, F., & Schlichter, J. (2013). Master data quality barriers: An empirical investigation. *Industrial Management & Data Systems*, *113*(2), 234–249. https://doi.org/10.1108/02635571311303550

Helo, P., Tuomi, V., Kantola, J., & Sivula, A. (2019). *Quick guide for Industrial Management thesis works*.

Hikmawati, S., Santosa, P. I., & Hidayah, I. (2021). Improving Data Quality and Data Governance Using Master Data Management: A Review. *IJITEE (International Journal of Information Technology and Electrical Engineering)*, *5*(3), 90. https://doi.org/10.22146/ijitee.66307

Hillard, R. (2010). *Information-driven business: How to manage data and information for maximum advantage*. John Wiley & Sons. https://doi.org/10.1002/9781119200338

Hubert Ofner, M., Straub, K., Otto, B., & Oesterle, H. (2013). Management of the master data lifecycle: A framework for analysis. *Journal of Enterprise Information Management*, *26*(4), 472–491. https://doi.org/10.1108/JEIM-05-2013-0026

Ibrahim, A., Mohamed, I., & Safie, N. (2021). Factors Influencing Master Data Quality: A Systematic Review. *International Journal of Advanced Computer Science and Applications*, *12*(2). https://doi.org/10.14569/IJACSA.2021.0120224

Inmon, W. H., & Lindstedt, D. (2015). *Data architecture: A primer for the data scientist: big data, data warehouse and data vault*. Morgan Kaufmann.

Kotusev, S., Kurnia, S., & Dilnutt, R. (2022). The concept of information architecture in the context of enterprise architecture. *Aslib Journal of Information Management*, *74*(3), 432–457. https://doi.org/10.1108/AJIM-05-2021-0130

Loshin, D. (2008). *Master Data Management*. Elsevier. https://doi.org/10.1016/B978-0-12-374225-4.X0001-X

Merwe, E., Pisa, N., Gideon, E., & Chakamera, C. (2024). Effects of material master data management on supply chain performance at FLSmidth: The moderating role of PiLog external service provider. *Acta Logistica*, *11*(4), 569–578. https://doi.org/10.22306/al.v11i4.547

Osakwe, J. (2026). A UNIFIED THEORY OF DATA GOVERNANCE: INTEGRATING FRAMEWORKS FOR ENHANCED DATA PROTECTION AND COMPLIANCE. *INTERNATIONAL JOURNAL OF INFORMATION SECURITY*, *5*(1), 1–25. https://doi.org/10.34218/IJIS_05_01_001

Otto, B. (2012). How to design the master data architecture: Findings from a case study at Bosch. *International Journal of Information Management*, *32*(4), 337–346. https://doi.org/10.1016/j.ijinfomgt.2011.11.018

Otto, B., Hüner, K. M., & Österle, H. (2012). Toward a functional reference model for master data quality management. *Information Systems and E-Business Management*, *10*(3), 395–425. https://doi.org/10.1007/s10257-011-0178-0

Pansara, R. (2023). DESIGN OF MASTER DATA ARCHITECTURE. *International Journal of Engineering Applied Sciences and Technology*, *8*(4), 58–61. https://doi.org/10.33564/IJEAST.2023.v08i04.007

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Reif, J. A. M., Kugler, K. G., Stockkamp, M. T., Richter, S. S., Benning, V. M., Muschaweck, L. A., & Brodbeck, F. C. (2022). An employee-centered perspective on business processes: Measuring “healthy business processes” and their relationships with people and performance outcomes. *Business Process Management Journal*, *28*(2), 398–418. https://doi.org/10.1108/BPMJ-06-2021-0375

Saunders, M., Lewis, P., & Thornhill, A. (2007). *Research methods for business students* (4. ed). Financial Times Prentice Hall.

Schmuck, M., & Georgescu, M. R. (2025). Master Data Management and its optimisation: An assessment of the maturity level on the example of a medium-sized company.

Procedia Computer Science, *263*, 575–582. https://doi.org/10.1016/j.procs.2025.07.069

Schmuck, M., & Rieder, E. (2025). MASTER DATA MANAGEMENT: A CONCEPTUAL FRAMEWORK FOR SUCCESSFULLY SUPPORTING DATA-DRIVEN TRANSFORMATION AND CORPORATE SUCCESS. *Review of Economic and Business Studies*, *18*(1), 177. https://doi.org/10.47743/rebs-2025-1-0011

Silvola, R., Jaaskelainen, O., Kropsu-Vehkapera, H., & Haapasalo, H. (2011a). Managing one master data – challenges and preconditions. *Industrial Management & Data Systems*, *111*(1), 146–162. https://doi.org/10.1108/02635571111099776

Simsion, G., Milton, S. K., & Shanks, G. (2012). Data modeling: Description or design? *Information & Management*, *49*(3–4), 151–163. https://doi.org/10.1016/j.im.2012.01.003

Smith, H. A., & McKeen, J. D. (2008). Developments in Practice XXX: Master Data Management: Salvation Or Snake Oil? *Communications of the Association for Information Systems*, *23*. https://doi.org/10.17705/1CAIS.02304

Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior*, *51*, 1068–1076. https://doi.org/10.1016/j.chb.2014.09.030

Srikant, S. (2006). Logical Data Modeling: A Key to Successful Enterprise Data Warehouse Implementations. *DM Review*, *16*(9), 13–16. (22546837).

Väre, T. (with Alma Talent). (2019). *Master data* (1. painos). Alma Talent.

Vilkka, H. (with Santalahti-kustannus). (2025). *Tutki ja kehitä* (6., uudistettu painos). Santalahti-kustannus.

Vilminko-Heikkinen, R., & Pekkola, S. (2017). Master data management and its organizational implementation: An ethnographical study within the public sector. *Journal of Enterprise Information Management*, *30*(3), 454–475. https://doi.org/10.1108/JEIM-07-2015-0070

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

Vom Brocke, J. (with Hevner, A. R., & Mädche, A.). (2020). *Design Science Research. Cases*. Springer International Publishing AG. https://ebookcentral-proquest-com.proxy.uwasa.fi/lib/tritonia-ebooks/detail.action?docID=6356690.

Wang, R. Y., & Strong, D. M. (1996). Beyond Accuracy: What Data Quality Means to Data Consumers. *Journal of Management Information Systems*, *12*(4), 5–33. https://doi.org/10.1080/07421222.1996.11518099

Weber, K., Otto, B., & Österle, H. (2009). One Size Does Not Fit All—A Contingency Approach to Data Governance. *Journal of Data and Information Quality*, *1*(1), 1–27. https://doi.org/10.1145/1515693.1515696

Zhao, C., Ren, L., Zhang, Z., & Meng, Z. (2020). Master data management for manufacturing big data: A method of evaluation for data network. *World Wide Web*, *23*(2), 1407–1421. https://doi.org/10.1007/s11280-019-00707-8

Appendices

Appendix 1. Interview at Case Company

Main concepts: Data Governance, Master Data Management, Information Architecture

Introduction

- What is your name and could you describe your current position and primary responsibilities, including how long have you been working in your current role?
- What is your relationship or involvement with Master Data Management and Data Governance?
- To what extent do you interact with different master data domains?
- Describe how the responsibilities are divided in master data management and data governance in the company?

Data Governance Framework: Scaling Proven Success & Standards

Focus on leveraging existing maturity and governance to build a scaling foundation

- How would you describe the success of the current master data governance model in the company (e.g. domains already managed customer, product, finance objects)
- Are there any metrics to measure data governance activities and compliance?
- Does the company have a Data Governance Council or a steering group established?
- What are the most critical roles and responsibilities that currently ensure data quality?
- In your view, have the master data standards been established and integrated into the company's data governance lifecycle?
- In your view, what are the best practices currently used that should be documented in a Global Playbook?
- What is missing from our governance documentation that causes confusion in day-to-day operative work?

- When different systems or departments disagree on a data definition, how is that conflict currently resolved and should the Playbook formalize this process?

Data Quality Excellence: KPI Standards & Automated Monitoring

Focus on measuring quality and defining success for the Playbook

- To what extent are data quality standards, criteria or KPIs defined for master data elements today?
- How are these data quality standards maintained and are they reviewed periodically to ensure relevancy to business?
- Is master data quality monitoring automated and continuous or manual and periodic?
- Are the data quality issues assessed and what is the process for communicating findings and taking corrective actions? Are the actions one-time clean-ups or continuously improved?
- How well are the data quality issues understood across the organisation and is the knowledge used to prevent future issues?
- How can the Playbook help automate or standardize the “Data Quality Dashboard” requirements for new domains?

Data Model, Information Architecture and Integrations

Focus on defining the unified logical reference model across IT landscape

- To what extent is there a unified definition for master data elements? (What is material/product, customer, supplier, vendor)
- Is there a Master Data data model defined for the company?
    - If yes, does it map the relationships between systems?
- Do you know whether any tools are used and available in the company for data modelling and to ensure data consistency?
- In your view, is data model compatibility ensured and reviewed between systems?
- Beyond master data management, what roles and responsibilities should exist for data modelling?

- Is there a centralized Data Glossary or dictionary maintained that allows users to view Master Data attributes, definitions, and formats?
- Could you describe metadata maintenance process structure, are the responsible persons for these updates defined?
- What are the most common points of failure or inconsistency when data is replicated between systems?
- In your view, what are the biggest obstacles to adopting a shared data model today (e.g., legacy system limitations, lack of centralized ownership)
- What could be the primary benefit of having a “unified logical reference model” for our new master data domains?
- How should the Playbook address constraints imposed by legacy systems that cannot comply with the global reference model?

System Architecture & End-to-End Data Lifecycle

Focus on mapping the “Single source of truth” and master data lifecycle

- How are the roles of each IT system defined regarding data sovereignty e.g. which systems act as the primary source for creation and which are the consumers of data?
- Has a Single source of truth been identified and defined for every master data domain?
- How frequently are the core master data lifecycle processes, such as storage, system integrations, archiving and deletion reviewed for efficiency and alignment with business needs?
- To what extent can current lifecycle processes and subject matter expertise be reused for new data domains?
- What kind of organizational knowledge and technical competencies, or practical experience from managing our current master data lifecycles could be most effectively reused when onboarding new data domains?

Playbook Utility & Scalability & Adoption

Focus on ensuring the Playbook is a practical tool

- What other data sets do you consider “business-critical” for a company's operations and what criteria/-s should trigger their inclusion in the MDM framework?
- If you were tasked to onboard a new data domain today, what would be the first instruction or template you would look for?
    - What format or tool would be most useful for your daily work?
- What are your top three requirements for data quality when we start managing a new data domain?
- What format should the Playbook take (e.g., process maps, checklists, modelling templates) and where should the Playbook be hosted (shared drive, management system?) to ensure it is part of the active development/business workflow?
- In your view, what are the important factors for the Playbook scalability so that it remains relevant as IT systems and business needs to evolve?
- What kind of training or onboarding kit should accompany the Playbook to ensure stakeholders actually adopt the standardized practices?
- What would prevent you or stakeholders from following the guidelines in the Playbook, and how can we design it to avoid those obstacles?