Representing Metadata in Data Governance and Data Quality

Exploratory research on the Representation of Metadata and how this could affect Data Governance and Data Quality

Master Thesis

21-07-2023

Student name:	M. Coenraads
Student ANR/SNR:	651492 / 2077024
Master program:	Information Management
Educational institution / Faculty:	Tilburg University / Tilburg School of Economics and Management (TiSEM)
Supervisor:	Dr. A. Saghafi
Second reader:	Dr. H. Weigand

Preface

Dear reader,

First of all, I would like to thank you for finding this thesis. I hope it will bring you interesting insights.

This thesis is the final requirement to obtain my master's degree in Information Management. During the master's program I had a course called Advanced Data Management. I liked the content of the course. Therefore, I went on seeking for companies that had a strong data management/governance component. After searching, I had found my internship company, it was Clever Republic.

During the research, I have gained substantial knowledge about the three main concepts in this study, metadata, data governance, and data quality. Sometimes I even got lost in all the articles about one of the three concepts. However, I am still interested in the future of the combined concepts.

This research could not have been here without the help of some people;

I would like to thank my thesis supervisor Arash Saghafi for his time and feedback on my process. I would also like to thank Clever Republic for providing me with participants and a great place to work! Many thanks to Aura for providing feedback!

I would like to thank my girlfriend, Manon, for helping me out during the stressed times writing this thesis. I would also like to thank my brother for providing feedback.

Finally, I hope you will enjoy reading this thesis and learn from it as much as I have learned.

Mo Coenraads
Rotterdam, 21-07-2023

Management Summary

In today's digital environment, data plays a crucial role in decision-making, compliance, and gaining a competitive advantage for organizations. Effective management and utilization of data requires addressing the challenges associated with metadata and its representation, data governance, and data quality. Two problems were found in the metadata field (i) the lack of standardization and (ii) human interaction related. First, projects are hindered due to the lack of standardized metadata, information that is created is misinterpreted and models that were developed are unclear. Furthermore, the individual concepts metadata, data governance, and data quality have been studied comprehensively in the literature. However, the literature on the combined concepts of metadata, data governance, and data quality is limited. Therefore, with the aforementioned problems and the understudied literature, this research has explored and focused on the combinations of metadata & data governance and metadata & data quality. The following research question has been drawn up "How could the representation of Metadata affect Data Governance and Data Quality in organizations?"

Due to the understudied literature on the combinations of the concepts, this research used the Grounded Theory to answer this research question. Semi-structured interviews were conducted to collect data. In total nine participants have been interviewed at six different organizations.

After analyzing the interviews, a data structure was created. This structure consists of nine second order themes and three high over dimensions. These dimensions reveal how organizations interpret metadata representation in the contexts of data governance and data quality.

Metadata representation in the context of data governance includes presenting structured and organized information regarding data elements, definitions, quality standards, access rights, and compliance rules. This enables businesses to manage, govern, and control their data more efficiently. Metadata is critical for understanding and achieving compliance obligations, as well as assuring data quality and defining and enforcing data governance rules and standards.

In terms of data quality, metadata representation provides organized details about data characteristics and attributes, assisting in the comprehension and evaluation of data quality. It provides useful insights into data sources, verification procedures, and business regulations, while also encouraging a uniform language, better communication, and a shared understanding of data. Metadata connects technical elements to business goals, allowing for more effective data governance and quality control.

In conclusion, well-structured metadata representation improves data governance, compliance, comprehension, and quality management. It encourages stakeholder contact and collaboration. Organizations that have strong metadata representation succeed in data management and governance, ensuring high data quality standards.

Table of Contents

Preface 2
Management Summary 3
Table of Contents 4
List of figures and tables 6
1. Introduction 7
1.1 Problem Statement 8
1.2 Research Question 10
1.3 Research Method 10
1.4 Structure of Thesis 11
2. Theoretical Background 12
2.1 Data 12
2.2 Metadata 15
2.3 Metadata & Data Quality 26
2.4 Metadata & Data Governance 27
2.5 Summary and conclusion 29
3. Methodology 31
3.1 Research Method 31
3.2 Data Collection 32
3.3 Data Analysis 38
4. Results 41
4.1 Rating questions 41
4.2 Metadata 46
4.3 Metadata Representation 51
4.4 Metadata Representation in a Data Governance context 55
4.5 Metadata Representation in a Data Quality context 62
4.6 Key points of results 67
5. Discussion 69
5.1 Data Structure 69
5.2 Dimension I: Improving understanding 70
5.3 Dimension II: Data proficiency 74
5.4 Dimension III: Data Governance Lifecycle 78
5.5 Definition of participants 84
5.6 Implications 87
5.7 Relevance 88
5.8 Limitations 89
6. Conclusion 91

6.1 Future research....92
7. References....95
8. Appendices....105
8.1 Appendix I: Design of the literature review....105
8.2 Appendix II: Interview structure....110
8.3 Appendix III: The four-point approach to qualitative sampling....113
8.4 Appendix IV: Sample universe....114
8.5 Appendix V: Interviews....114
8.6 Appendix VI: Deleted category codes....115
8.7 Appendix VII: Citations in result chapter....116
8.8 Appendix VIII: Categories and themes to data structure....143

List of figures and tables

List of figures

Figure 1 - Graphical view of the pyramid from Russel Ackhoff's "Data to Wisdom" 13
Figure 2 - Data Structure 69
Figure 3 - Schematic overview of search for literature. 106
Figure 4 - Sample universe of this research 114

List of tables

Table 1 - Interview information 36
Table 2 - Category codes with Groundedness 39
Table 3 - Interview sections with codes 41
Table 4 - Results of the rating questions 41
Table 5 - Themes for 'Metadata general' 46
Table 6 - Themes for 'Metadata Representation general' 51
Table 7 - Themes for 'Metadata Representation in Data Governance' 56
Table 8 - Themes for 'Metadata Representation in Data Quality' 62
Table 9 - Dimensions and themes 92
Table 10 - Literature after selection 109
Table 11 – Rational for interview questions 112
Table 12 - The four-point approach to qualitative sampling from Robinson (2014, p. 26) 113
Table 13 - Deleted category codes 115
Table 14 - Category: Understanding 143
Table 15 - Category: (Meta)data management 143
Table 16 - Category: Organization 144
Table 17 - Category: Data management tools 144
Table 18 - Category: Privacy 145
Table 19 - From first order concepts to dimensions 145

1. Introduction

This chapter introduces the essential concepts used in this thesis: Data, Metadata, Data Governance, and Data Quality. Then the problem statement, research question, and method are stated. Finally, the structure of this thesis is given.

> "There were 5 exabytes of information created between the dawn of civilization through 2003, but that much information is now created every two days."
>
> - Eric Schmidt, former Executive Chairman at Google (2010)

The above quote of former Google Executive Chairman Eric Schmidt serves as an illustration to highlight the ever-increasing data stream in today's world. Data is critical in today's digital environment, as organizations rely on data for decision-making, compliance, and gaining a competitive advantage (Bansal et al., 2021; Brynjolfsson et al., 2011; Mcafee & Brynjolfsson, 2012; Rogers & Thompson, 2012; Yu et al., 2021). Organizations have issues in efficiently managing and utilizing this immense resource (Panian, 2010) as data grows exponentially (Edmond, 2021).

Data is fundamentally a collection of observations, information representations, or sets of facts (International Organization for Standardization (ISO), 2023). Emails, files, and other digital records are examples of digital information generated by human interactions with technology (Villars et al., 2011). Without adequate organization and structure, it is not easy to leverage the vast amount and diversity of data (Panian, 2010). To address these challenges, metadata has gained significant importance (IEEE, 2020; Loshin, 2015; Sundarraj & Rajkamal, 2019; van Helvoirt & Weigand, 2015). Metadata, commonly known as "data about data," contains critical information about the structure, content, and context of data (NISO, 2017; Pomerantz, 2015; Ulrich et al., 2022). It contains details such as the data source, format, lineage, and other relevant aspects (NISO, 2017; Pomerantz, 2015; Ulrich et al., 2022). Metadata, like a library catalog, facilitates efficient data discovery, understanding, and utilization (Pomerantz, 2015). It is critical in managing and organizing vast amounts of data, making it easier to find, access, and use (DAMA International, 2017; NISO, 2017; Pomerantz, 2015).

The Cambridge Analytica scandal is an example of the significance of metadata. Cambridge Analytica manipulated user behavior and influenced political outcomes using structured personal data from millions of Facebook users without agreement (Boldyreva, 2018; Kanakia et al., 2019). This incident underlined the need for data governance and controls to protect privacy and prevent unauthorized access to sensitive data (Viljoen, 2021).

In addition to metadata, strong data governance practices and comprehensive data management frameworks are required for managing data as a strategic asset (Panian, 2010). Data governance entails developing policies, procedures, and roles to govern data throughout its lifecycle (DAMA International, 2017; Panian, 2010). It provides a framework for data management, assuring compliance, improving overall data quality and reliability, and facilitating data accessibility (Abraham et al., 2019; DAMA International, 2017; Panian, 2010).

Furthermore, organizations must prioritize data quality to create accurate and reliable decision-making processes(Dyson & Foster, 1982; Harley & Cooper, 2021; Stvilia et al., 2007). Poor data quality can seriously affect analytical outcomes and restrict business intelligence efforts (Cichy & Rass, 2019; English, 2009; Grover et al., 2018). The European Commission has endorsed initiatives such as the Data Quality Act and the Data Quality Assessment Methods and Tools (DatQAM) to solve data quality issues and reduce the risks associated with poor data quality (Bergdahl et al., 2007).

In conclusion, effective data management, governance, and metadata utilization are critical as organizations navigate the digital landscape and harness the power of data. Organizations can use metadata to optimize data management procedures, improve data accessibility, maintain compliance, and limit risks.

1.1 Problem Statement

Problem

In the introduction, the importance of metadata, data governance, and data quality has been outlined. However, most studies about metadata, data governance, and data quality are about either. A recent literature review by Abraham et al. (2019) shows that much research on data governance has been conducted. However, these studies focused on subdomains, for example, cloud data governance, data governance principles, and agile data governance capabilities.

Another recent study by Timmerman and Bronselaer (2019) states that there has been much research on the meaning of and how to measure data quality. This shows the saturated field of data quality.

Finally, in a recent systematic review by Ulrich et al. (2022) that included 81 papers, it was discovered that there are 35 distinct metadata standards classified into three categories: (i) Structure standards, (ii) Technical standards, and (iii) Semantic standards. The review also identified five problem categories related to the processing and utilization of metadata. These categories are (i) Structural-related problems, (ii) Semantics-related problems, (iii) Human interaction–related problems, (iv) Metadata lifecycle–related problems, and (v) Metadata processing–related problems.

Ulrich et al. (2022) found problems related to structural- and semantic-related issues with using and processing metadata. While data has grown over the last decades, the scalability of projects is still hindered by metadata's lack of usage of standards (Bergmann et al., 2020; Elberskirch et al., 2022). For example, the data on buildings has grown over recent years, but the industry has not adopted standards to exchange, store, and use data (Bergmann et al., 2020). Another example is the use of standards and guidelines in the field of nanomaterials, with a focus on nanosafety to share data sets (Elberskirch et al., 2022). Exploring organizations' most used metadata presentation choices is essential to contribute to a standard metadata schema.

Al-Ruithe et al. (2019) mention in their literature review that data governance is mostly organized in siloes within organizations (Begg & Caira, 2012; Wende, 2007). This is also seen in practice: organizations separately organize their data governance and data quality. In contrast, data governance is an area that has arisen more recently. Due to the increasing volumes of data and data-driven decision-making, there is a necessity for reliable and accurate data.

Multiple studies (Al-Ruithe et al., 2019; Dhillon, 2019; Karkošková, 2023; Panian, 2010) focused on how to govern data (who makes what decisions) and the measurement of data quality (Batini et al., 2009; Bergdahl et al., 2007; Brodie, 1980; Cichy & Rass, 2019; Strong et al., 1997; Timmerman & Bronselaer, 2019; Zhang et al., 2019). When organizations treat data quality and data governance as separate entities, this may result in inconsistencies and inefficiency in the data. When these two are not aligned, it could be that the policies of both conflict with each other, which results in data that cannot be trusted, leading to potentially inaccurate or suboptimal decisions or more costs.

Ulricht et al. (2022) also found human interaction–related problems. They highlight that collaboration was a critical aspect of the articles. Not only did sharing and discussing the developed information present a chance to improve the designed data, but it was also an essential step in overcoming the challenges of misinterpretation (Trani et al., 2018). In addition, the problem extended to a conceptual level: the model would be unclear if stakeholders, users, and organizations differed slightly in their interpretation of the use cases (Eichenlaub et al., 2021).

Due to the lack of structured and standardized usage of metadata and the human interaction-related problems, the interest of this research lies in exploring how organizations present their metadata, what its challenges are, and how this affects data governance and data quality. Data governance and data quality will be treated as separate entities during this research.

Problem owner and result

This problem affects all organizations dealing with data governance and data quality. Nowadays, all organizations are dealing with data. The problem owner would be the management team responsible for the organization's data quality and Governance.

This thesis provides insight into the challenges that come with the representation of metadata, combined with data governance and data quality. The outcome is written in this research, providing a detailed understanding of metadata, its representation, and how this could affect data quality and data governance in organizations. The research will have an exploration focus on the 'how' aspect. The study will contribute to all topics addressed: Metadata, Data Governance, and Data Quality.

Scope and limitations

The scope is limited to organizations willing to discuss their metadata design and how it could affect data quality and data governance. Organizations that are not willing to talk are not useful for this study. Due to the nature and scope of this thesis, there is a limited time of four months, and there is no budget available to perform tasks.

1.2 Research Question

Given the lack of structured and standardized usage of metadata and human interaction-related problems, this study aims to provide insights for organizations into the challenges with metadata representation and what influence this has on data governance and quality. The following research question has been drawn up:

"How could the representation of Metadata affect Data Governance and Data Quality in organizations?"

Several sub-questions need to be answered to answer the research question:

1. How is metadata represented in organizations?
2. What is metadata representation in a data governance context in organizations?
3. What is metadata representation in a data quality context in organizations?

1.3 Research Method

This research explores the effects of metadata design on data governance and data quality within an organizational context. Specifically, the study has an exploration focus on the 'how' aspect of metadata design.

First, desk research is used to gain more knowledge and find literature about the most important concepts in this thesis: (i) Data; (ii) Metadata & Metadata Representation; (iii) Data Quality & Metadata; and (iv) Data Governance & Metadata.

Secondly, field research is conducted by interviewing organizations implementing data governance and data quality. These organizations are mainly large[1]. The interviewees are selected using purposive sampling. Nine participants are interviewed within six organizations.

The organizations are found through Clever Republic and my network. The semi-structured interviews are open-ended, meaning they are related to the specific themes (metadata, metadata representation, data governance, and data quality), have a general structure, but delve deeper when potential relevance arises during the interviews.

This research aims to provide insight into how organizations design their metadata and how it could affect data governance and data quality. This research is exploratory. The exploratory type of research has been chosen because the literature on metadata representation in combination with data governance and data quality is not mature. Multiple authors state that if insufficient literature is available, an exploratory approach can be used (Birkinshaw et al., 2011; Mills et al., 2009; Yin, 2018).

1.4 Structure of Thesis

This thesis is structured as follows:

- Chapter 1: introduced the most important concepts for this thesis (Data, Metadata, Data Governance, and Data Quality), the rational behind it, then the problem statement for this research is given, the research question, and the methodology used.
- Chapter 2: Chapter two provides a theoretical background on the most important concepts used in this thesis: (i) Data; (ii) Metadata & Metadata Representation; (iii) Data Quality & Metadata; and (iv) Data Governance & Metadata
- Chapter 3: outlines the research methodology, the selection of and the information about the organizations interviewed, and the data analysis process.
- Chapter 4: presents the results of the interviews conducted.
- Chapter 5: provides the discussion for this research. The results are analyzed in this section.
- Finally, in chapter 6, the conclusion of this research is provided and recommendations for future research are given.

[1]The organization fits the 'large organization' terms provided by The Ministerie van Economische Zaken en Klimaat (2021) define a 'large organization' as the following; a large firm has employed over 250 FTE; or
The net sales are over €50 million, and a more than €43 million balance sheet.

2. Theoretical Background

This chapter explains the most important concepts and theories to answer the research question. To answer these questions, knowledge is required about the key concepts in the research question. The key concepts are (i) Data; (ii) Metadata & Metadata Representation; (iii) Data Quality & Metadata; and (iv) Data Governance & Metadata. Below, each concept will be elaborated on. In Appendix I: Design of the literature review, the design for the semi-literature review and the selection of the articles can be found.

2.1 Data

In today's digitally enabled world, one cannot work without data. Data is connected with everything we touch in this world. Almost all actions we perform are digitally registered. For example, walking with a phone registers a step; this is just one event of a digitally registered action. Multiple sensors are built-in to a mobile phone. If GPS is on, it will track the steps you took at certain locations where you have been; an altimeter sensor can measure the height level of the current position. With all these events recorded and the information this provides. Some even say, "Data is the new oil" (DAMA International, 2017; Haupt, 2018). This implies that data will be as valuable as oil when found.

Definitions of Data and Information

Starting this chapter with the definitions of information and data provided by the International Organization for Standardization (2023):

- Information:
  "knowledge concerning objects, such as facts, events, things, processes, or ideas, including concepts, that within a certain context has a particular meaning" (p. 3).
- Data:
  "reinterpretable representation of information in a formalized manner suitable for communication, interpretation, or processing" (p. 3)

The definitions above show that context is needed to have knowledge. Without context, knowledge cannot be gained. The same can be read in the article of Ackhoff (1989), which is explained in the paragraphs below.

Data without context is nothing

Data comes in various types and formats (University of Cambridge, n.d.). Images, audio, spreadsheets, and transactional data are all different types. These different types have multiple formats that can be saved and stored. However, data is nothing without the proper context and knowledge (DAMA International, 2017; Gartner, 2016). Over time, models have been created that have presented thoughts about data to knowledge and even wisdom (Gartner, 2016). The

most influential one is the Ackhoff pyramid (Gartner, 2016). In Ackhoff's (1989) pyramid, he distinguishes five layers (i) Data, (ii) Information, (iii) Knowledge, (iv) Understanding, and (v) Knowledge. See Figure 1 for a graphical version of the pyramid.

Starting with the lowest layer of the pyramid: Data. This is the raw format of data. For example, a number in a spreadsheet is data. However, one number is not meaningful without its context.

Secondly, by moving up the pyramid, we go to the second layer: Information. Information is once we know the context in which the number is given. We can answer simple questions by providing context to the data (Ackoff, 1989; Gartner, 2016). For example, if the number is part of a dataset with sales data and the column name is 'Revenue,' we can assume the number represents the revenue. However, we do not know if the revenue is stated per month, daily, or yearly.

Third, moving one step up in the pyramid to Understanding, can be seen as linking more columns in a spreadsheet together, thus providing a broader context of the data (Ackoff, 1989; Gartner, 2016). With the example in mind, we can now see that the revenue is part of the monthly sales.

The fourth step is understanding the knowledge. This step consists of finding patterns when the knowledge of the data has been acquired (Ackoff, 1989; Gartner, 2016).

At last, step five, Wisdom is knowing what to do with the knowledge and applying it.

Gartner (2016) states that at all levels (except data), in the pyramid, different questions can be asked.

- Information: allows to provide answers to questions such as "Who?", "When?", "Where?" or "What?".
- Knowledge: allows one to provide answers to questions such as "How (things work)".
- Understanding: allows one to provide answers to questions such as "Why";
- Wisdom: should give answers to questions such as "What is best?" or "What is the right thing to do?". Data on itself cannot answer any questions without its context. This was in 1989 and is still applicable today (Ackoff, 1989; DAMA International, 2017; Gartner, 2016).

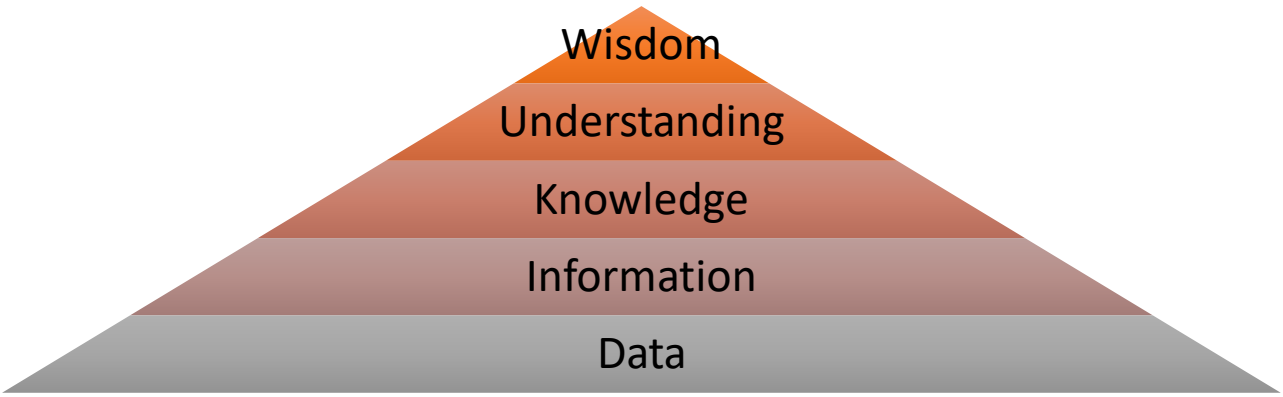

Figure 1 - Graphical view of the pyramid from Russel Ackhoff's "Data to Wisdom"

Big Data

In the Introduction is mentioned that data a vital asset is for organizations (Fisher, 2009; Panian, 2010). What is also mentioned is the amount of data produced today and what will be created in the future. According to Zikopoulos and Eaton (2011), enormous amounts of data are called 'Big Data'. Zikopoulos and Eaton (2011) define Big Data as: *"information that cannot be processed or analyzed using traditional processes or tools"* (p. 3). This means that data has outgrown the processes and tools used to derive information from data (Davenport et al., 2012). Big Data was first characterized by Laney in 2001 with the three Vs: Volume, Velocity, and Variety. Over the years, there have been added Vs due to the larger amounts of data: Veracity (Zikopoulos & Eaton, 2011), Value (Ishwarappa & Anuradha, 2015), and Variability (Fan & Bifet, 2013). Today, Big Data is commonly characterized by the four V's (Fan & Bifet, 2013; Ishwarappa & Anuradha, 2015; Shmueli et al., 2017; Zikopoulos & Eaton, 2011) :

(i) Volume: refers to the enormous amounts of data that in today's world are created, collected, and stored as a result of the rapid growth of digital technologies.
(ii) Velocity: describes the amount of data produced and how quickly it must be processed to get information out of it.
(iii) Variety: refers to the different types of data there. Examples are Audio; Text; Video; Image; and other forms. Due to all these different types of data, integrating and analyzing them is not easy.
(iv) Veracity: refers to the accuracy and trustworthiness of data. When data is collected, it could be incomplete, which brings risks. It is the level of trust in the data collected.

The value of Big Data lies in the insights and information we can retrieve by conducting data analysis (Zikopoulos & Eaton, 2011). In recent years, data-driven intelligence has transformed how organizations conduct their business and make well-informed decisions based on data (Bansal et al., 2021). Making the right decisions leads to achieving goals (Yu et al., 2021). During the COVID-19 pandemic, vast decisions had to be made to save economies and communities (Schippers & Rus, 2021; Yu et al., 2021). Due to the already available data about pandemics and the continuing COVID-19 data stream, well-informed decisions could be made (Yu et al., 2021). Studies found that organizations that make decisions based on accurate and trustworthy data performed superior to their competitors (Brynjolfsson et al., 2011; Mcafee & Brynjolfsson, 2012; Rogers & Thompson, 2012). In another study conducted by Brynjolfsson & McElheran (2016), it was found that long-established firms with multiple sites are (on average) more likely to switch to data-driven decision-making than young firms with a single site. Davenport (2006) accentuates that organizations benefit by analyzing the in-house and external data from different sources. The analysis of Big Data is increasingly common in today's digital world, recreating businesses and transforming industries (LaValle et al., 2010). According to Evans et al. (2012), the ability of organizations to create wealth in the current knowledge-based economy is no longer

solely dependent on tangible assets. They state that every organization depends on intangible resources like data, documents, content, and knowledge to function.

Database layers

Designing databases requires understanding what data should be captured and how data relates to each other. Three layers in database design contribute to the organization, administration, and utilization of data. The three modeling layers: (i) Conceptual, (ii) Logical, and (iii) Physical, play distinct roles in determining the structure and comprehension of data (Garcia-Molina et al., 2008; Ramakrishnan et al., 2003). According to multiple authors (Garcia-Molina et al., 2008; Groves, 2022; Ramakrishnan et al., 2003), the layers consist of the following:

(i) The conceptual layer is the least detailed layer where the 'what' of data is modeled. For example, designing a database for an online store needs to understand what data entities should be captured, such as the customer, payment, and product information. The entities are filled with attributes. The conceptual layer often comes from the business and its processes.

(ii) The logical layer shows 'how' the data is modeled. It is more detailed than the conceptual layer. For example, in the customer dimension, the customer's first name should be of the type of VARCHAR with a maximum of 20 characters, and the birth date should be of the type of DATE with the European date format. Architects and business analysts often create the physical layer.

(iii) The physical layer is the most detailed, specifying which type of database is used along with the infrastructure of the database. This layer often is created by database architects or developers.

In conclusion, data is essential in our digitally enabled world, but it requires context and knowledge to be relevant (DAMA International, 2017; Gartner, 2016). The Ackhoff (1989) pyramid illustrates the path from raw data to wisdom, highlighting the significance of understanding and applying knowledge (Gartner, 2016). Big Data, characterized by its volume, velocity, variety, and veracity, confronts organizations with both challenges and opportunities. Analyzing Big Data for insights allows for informed and data-driven decision-making (LaValle et al., 2010). With its conceptual, logical, and physical layers, effective database design facilitates data organization, administration, and utilization (Garcia-Molina et al., 2008; Groves, 2022; Ramakrishnan et al., 2003),. Overall, the importance of data rests in its ability to inform and transform decision-making processes in our data-driven world.

2.2 Metadata

As previously stated, data without context has no meaning (DAMA International, 2017; Gartner, 2016). DAMA International (2017) mentions that the context of data can be seen as the

representational system for the data; this system contains a shared vocabulary and a hierarchy of relationships between its parts. DAMA International (2017) states that metadata is important because it describes the data. This is what makes metadata valuable for organizations. DAMA International (2017) provide an example; A library has thousands, if not ten thousand of books, but no catalog exists. Without the catalog, finding books in the library would be hard, almost impossible. The catalog stores information about a book. This information creates a starting point to search for a book in the library based on the author, title, or genre. Without metadata, an organization would be the same as a library without its catalog. This makes metadata valuable for organizations.

Due to the limitations of the human brain (Mayer, 2003), we create shortcuts such as mnemonics[2]; this is why metadata exists (Gartner, 2016). When we exceed the limitations of a single brain, the need for metadata becomes ever more crucial (Gartner, 2016).

Guerra et al. (2013, p. 376) stated that metadata is a term that has multiple definitions and can be interpreted differently according to the context. The literature review of Ulrich et al. (2022) confirmed that there are many different definitions for metadata used in the literature. They provide a general definition *"metadata is a formal representation of data that defines and describes information in a (preferable) standardized and stable way"* (Li et al., 2012; Löpprich et al., 2014). DAMA International (2017) has a more specific definition: *"Metadata includes information about technology and business processes, data rules and constraints, and logical and physical data structures. It describes the data itself (e.g., databases, data elements, data models), the concepts the data represents (e.g., business processes, application systems, software code, technology infrastructure), and the connections (relationships) between the data and concepts"* (p. 417).

We can find overlapping concepts between both definitions. Combining these overlaps: Metadata is data that says something about the data; this includes database elements, data models, relationships, application systems, technology infrastructure, and concepts. In this research, the definition provided by DAMA is used. First, the definition captures all important aspects of metadata. Secondly, DAMA is a highly recognized organization in the data industry.

Organizations consider data as a vital asset (Fisher, 2009; Panian, 2010). It is used for decision making processes, compliance control and other important business operations. With the volume of data increasing, it becomes apparent that metadata or phrased differently the "data about data" also becomes increasingly more crucial for organizations. Next to that, organizations typically structure their business in departments, each having its own tasks and responsibilities.

[2] Mnemonics are designed to extract and represent the most notable features of the information they aim to preserve in our memory, for example "Richard of York Gave Battle in Vain," which represents the color spectrum: Red-Orange-Yellow-Green-Blue-Indigo-Violet (Gartner, 2016).

With that, an organization's knowledge is typically divided among multiple individuals (DAMA International, 2017) and thus without proper documentation and structuring between the data of these departments, data may be lost or unutilized with all its accompanied risks and consequences (DAMA International, 2017). Metadata helps capture the available data knowledge (DAMA International, 2017). The risk lies not only in losing knowledge, but it is also helpful in order to divide data into private or sensitive information and manage the lifecycle to meet compliance and exposure risks. Without reliable metadata, an organization is unaware of its data and what it represents (DAMA International, 2017). Finally, metadata is also necessary to manage data as an asset (DAMA International, 2017).

Three types of metadata

Now that background information on metadata is given, we continue by distinguishing between three types of metadata that are given by DAMA International (2017), Gartner (2016), and NISO (2017). These types are (i) Descriptive Metadata; (ii) Administrative Metadata; (iii) Structural Metadata.

(i) Descriptive Metadata

Descriptive Metadata is the most common type (Gartner, 2016). This type is created to help discover and locate where the data is stored (DAMA International, 2017; Gartner, 2016; NISO, 2017). With the analogy of a library, this type of metadata is necessary to find the book you seek. Examples of what kind of information this is, are genre, title, and author. Descriptive Metadata is often referred to as 'finding metadata.'

(ii) Administrative Metadata

The background information that ensures data can be stored, preserved, and accessed when needed is called Administrative Metadata (DAMA International, 2017; Gartner, 2016; NISO, 2017). Administrative Metadata is essential to the operation of any information system, whether electronic or not. It frequently exceeds Descriptive Metadata in size and scope (Gartner, 2016). This kind of metadata has always been important because every library has had to maintain information for its administration to function. This metadata has a higher volume and is frequently far more complex in information systems.

a. Technical Metadata

A specific form of Administrative Metadata is Technical Metadata (Gartner, 2016; NISO, 2017). This Technical Metadata is needed to run systems and deliver digital data to us in a manner that can be understood (Gartner, 2016). Technical Metadata has all the information it needs from an object to visualize it. For a digital image, this can include information about its dimensions in pixels, the file format, the size of the color palette, and many other details. Technical Metadata varies depending on the type of digital asset; for example, what is needed for photographs

differs from what is needed for video, audio, or text. This metadata is usually created automatically while a system is running and usually is not visible to the user (Gartner, 2016).

b. Rights Metadata

The data required for a system to enforce intellectual property rights (IPR) is also known as a form of Administrative Metadata (Gartner, 2016; NISO, 2017). This metadata includes information about the IPR's owners, copyright laws that govern their ownership rights, and the rights those owners give to those who access their data. IPR still holds outside of the digital realm, and every printed book has Rights Metadata connected to it. It is used to impose payment methods by an online newspaper like The New York Times, which charges users to view more than ten stories monthly.

c. Preservation Metadata

Preservation Metadata is the final type of Administrative Metadata used to ensure that digital data is accessible and useable for a long time (Gartner, 2016; NISO, 2017). It is complex and broad and requires careful maintenance to remain viable (Gartner, 2016). Large amounts of metadata are needed to support the procedures that make this possible (Gartner, 2016). This documentation of what is done to the data, how it is done, and by whom assures that it can still be used long after the original producers have passed away. Technical Metadata is invisible to the user, but it is crucial.

(iii) Structural Metadata

The last type of metadata can be required but is not always evident to the end user. Structural Metadata establishes connections between distinct data points to combine them into a more oversized, more complex item (DAMA International, 2017; Gartner, 2016; NISO, 2017). This information is known as structural Metadata and transforms a collection of unrelated pages into something we can identify as a book (Gartner, 2016; NISO, 2017). Even a physical book has this information due to the page numbers and arrangement of the leaves when fastened to the spine. If an item of any complexity will make sense in the digital world, it needs to be captured more specifically.

Metadata sharing and storing

Metadata is shared and stored similarly to 'normal' data. This can be done through databases as records, through a language made to describe data, or visualized into models. Below, the most common ways of sharing (meta)data are explained.

Relational Databases

Relational Databases were introduced by Ted Codd (1970). Todd suggests that databases should represent data in tables called relations[3]. Information in tables is stored in records. The practical design relies on normalizing database tables to improve query performance and enhance storage efficiency (Garcia-Molina et al., 2008; NISO, 2017; Ramakrishnan et al., 2003). Today, Application Programming Interfaces (APIs) are a standard method used by software systems that use this information model and wish to share it with others (NISO, 2017; Pomerantz, 2015). APIs consist of specification documents external software developers use to create tools that query the system and obtain relevant (Meta)data.

XML

XML (eXtensible Markup Language) became a popular method for the encoding, transferring, and storing metadata in the 2000s (NISO, 2017). A single root element is the base of an expanding tree representing an XML document (Garcia-Molina et al., 2008; NISO, 2017; Ramakrishnan et al., 2003). There are languages based on XML written for specific tasks and XML processing toolkits for major programming languages (Garcia-Molina et al., 2008; NISO, 2017). Metadata stored as XML is often loaded directly from external sources, generated through software user interfaces, or mapped from other data sources (NISO, 2017).

Linked Data and RDF

The most significant step toward putting the 'Semantic Web' (a worldwide network for actionable data) into practice was introduced by Tim Berners-Lee with the concept of Linked Data (NISO, 2017). Linked Data heavily relies on Resource Description Framework (RDF) standards, which model data as a network with no single object or piece of information having fundamental significance (Forum Standaardisatie, 2013; NISO, 2017). Individual triples comprise RDF graphs, in which a subject is coupled with an object by a predicate. Predicates are represented as properties, and class names begin with a capital letter (Forum Standaardisatie, 2013; NISO, 2017).

RDF properties' domains and ranges serve two purposes: they provide direction for implementers and make it possible for processing tools to create new associations (NISO, 2017; Pomerantz, 2015). Uniform Resource Identifiers (URIs) or International Resource Identifiers (IRIs) represent subjects, predicates, and objects. Class and property relationships are established using this structure (Forum Standaardisatie, 2016b; NISO, 2017; Pomerantz, 2015). Linked data is a powerful approach for connecting data from several sources (NISO, 2017; Pomerantz, 2015). Classes, properties, object datatypes, property domains and ranges, and hierarchical relationships between classes and subclasses are all defined using RDF Schema (RDFS) (Forum standaardisatie, 2021; NISO, 2017; Pomerantz, 2015). URIs should be dereferenceable, and

[3] See the website from oracle for more information about relational databases https://www.oracle.com/database/what-is-a-relational-database/#link2.

content negotiation is used to provide raw data to humans and software applications. Linked Data is a powerful tool for connecting information from multiple sources (NISO, 2017; Pomerantz, 2015).

According to Riley (2017), RDF data can be shared in various ways, such as microformats, large RDF triplestores, Linked Data, RDF/XML, RDFa (Forum Standaardisatie, 2016a), and Turtle. RDF/XML is a verbose encoding of RDF, while RDFa embeds RDF within HTML (Forum Standaardisatie, 2016a), and Turtle is a textual representation of an RDF graph. N-Triples is a text-based RDF serialization used to present multiple predicates of the same subject compactly. Metadata can also be stored and shared by embedding it in a digital file (NISO, 2017).

Several metadata other standards and languages have been explained clearly by the National Information Standards Organization (NISO) in the primer by Riley (2017) titled 'Understanding Metadata.'

In conclusion, metadata plays a crucial role in our understanding and utilization of data. It provides data context, description, and structure, helping organizations more effectively manage and utilize their data assets (NISO, 2017; Pomerantz, 2015; Ulrich et al., 2022). Metadata, like a library catalog, helps us in discovering and identifying useful information. Each of the three types of metadata (descriptive, administrative, and structural) serves a different purpose in capturing essential information about data and ensuring its proper storage, preservation, and accessibility (DAMA International, 2017; Gartner, 2016; NISO, 2017). Methods such as relational databases, XML, and Linked Data make it easier to share and store metadata in structured and standardized formats (Gartner, 2016; NISO, 2017; Pomerantz, 2015). By leveraging metadata, organizations can improve data discovery, governance, and decision-making processes (DAMA International, 2017; NISO, 2017; Pomerantz, 2015). Finally, metadata is essential for maximizing the value and potential of data in today's data-driven world.

Metadata Representation

Because our brains have limitations in processing raw information, visualizing information is critical for humans to understand complicated concepts and data (Mayer, 2003). Similarly, displaying metadata contributes significantly to our understanding of it. Metadata representation has a significant impact on data governance and data quality, making it critical for organizations to understand its implications and maximize the value of their data assets. Standardization and structured representation are critical in metadata representation to enable consistency and interoperability across multiple systems and domains (Dai et al., 2021; Mandal et al., 2016; Melo et al., 2021).

The study by Dai et al. (2021) emphasizes the importance of standardized metadata through the proteomics sample metadata representation framework for multiomics integration and big data analysis. This highlights the need for a structured and standard approach to metadata

representation, enabling seamless integration and analysis of complex biological data. Similarly, a study by Melo et al. (2021) highlights the significance of a strategy for archives metadata representation in supporting knowledge discovery within archives. By adopting a structured approach, archives can effectively organize and manage metadata, enabling researchers and users to more efficiently discover and access relevant historical information. The article of Mandal et al. (2016) extends metadata representation by collecting certain contextual aspects relevant to data collection, management, and utilization. The context-driven metadata representation approach for SaaS emphasizes the relevance of context-specific metadata in the Software as a Service environment for effective decision-making and quality control. SaaS providers can understand their data's specific conditions and requirements by including context-driven information, enabling more informed decision-making processes, and assuring greater data quality and governance standards.

While each metadata representation system has a different purpose, there are notable shared characteristics between them. Both the proteomics sample metadata framework and the method for archival metadata representation emphasize the importance of standardized and structured metadata representation. Standardization ensures standard formats and protocols, allowing seamless data integration and analysis across systems and domains. Structured representation systematically organizes metadata, making discovering knowledge and retrieving information easier.

Furthermore, combining advanced technologies with metadata representation improves its capabilities and effectiveness. Semantic-based metadata representation improves multimedia security and management by exploiting semantic information to secure multimedia asset integrity and confidentiality (Rytsar et al., 2003). Moreover, combining 3D modeling approaches with metadata representation improves information management (Yen et al., 2013). By merging metadata with 3D models, essential contextual information can be maintained, enhancing asset understanding and preservation.

Lastly, a recent systematic review of 81 papers by Ulrich et al. (2022) found that there are 35 different metadata standards used and structured them in three categories - subtracted from Costin & Eastman (2019) - (i) Structure standards; (ii) Technical standards; and (iii) Semantic standards. They also found five problem categories concerning the processing and usage of metadata. These are:

(i) Structural-related problems: this problem is related to the lack of standard usage of metadata. The reviewed articles expressed problems about the limited or overly broad range of available standards, which were considered confusing (Huang et al., 2017; Park & Tosaka, 2010).
(ii) Semantics-related problems: this problem is related to the sharing of metadata. The metadata elements included descriptions and definitions to provide clarity on the purpose of the data. However, these descriptions often contained synonyms, spelling variations,

or naming conflicts, which could lead to discrepancies in the shared data (Eichenlaub et al., 2021).

(iii) Human interaction–related problems: this problem is related to the human factor of understanding metadata. Sharing and discussing the developed information was not only an opportunity to improve the quality of the created data, but it was also a necessary step in overcoming issues linked to misinterpretation (Trani et al., 2018).

(iv) Metadata lifecycle–related problems: this problem relates to the mismatch between data and its corresponding metadata (Li et al., 2012). The data was frequently not in alignment with the metadata, making it unacceptable for reuse. There were several explanations for this inconsistency, including a lack of transparency regarding the source of the (meta)data (Maumet et al., 2016) or an unclear distinction or shifting perspective between data and metadata (Papež & Mouček, 2017).

(v) Metadata processing–related problems: this problem relates to the processing of metadata. The literature found pre-processing barriers in metadata, such as varied interfaces that cause siloization (Jeong et al., 2014). Automatic matching from broad to a specific level was practically difficult (Daniel et al., 2014), and automated mapping without human input was complex or impossible (Ashish et al., 2016; Kock-Schoppenhauer et al., 2019). It was difficult to get more testing data for algorithm enhancement (Deppenwiese et al., 2019). Furthermore, merging data sets encountered challenges such as confusing mappings (Song et al., 2014), varying obligation levels (Specka et al., 2019), and faulty mappings that led to misinterpretation (Ngouongo et al., 2013).

Representation Theory

The following section provides information about Representation Theory in information systems.

Representation Theory in information systems was originally developed by Wand & Weber (1990). Wand & Weber saw a need to create insight into the artifacts of information systems. They started with the knowledge that humans and organizations naturally desire information (Weber, 1997). Representation Theory is developed for all information systems (Wand & Weber, 1990) and focuses on how they convey information.

The concept of meaning is central to four of Representation Theory (RT) assumptions about the communication of information (Burton-Jones et al., 2017):

(i) RT can communicate meaning through symbols, and humans can obtain meaning from these symbols (Wand & Weber, 1995).

(ii) Information systems intend to represent someone’s or a group's view of the real world (Wand & Weber, 1995). Wand & Weber (1993) and Burton-Jones & Grange (2013) reason

that computerized representations provide a more efficient way to learn about the world than observations. This leads to the design of information systems.

(iii) Stakeholders can express their meaning about events that are of interest to them (Burton-Jones et al., 2017).

(iv) Users expect that an information system has more value when it represents a true view of the real world (Weber, 1997).

RT explores incorporating user perceptions of meaning into an information system (Wand & Weber, 1995). Wand & Weber (Wand & Weber, 1995) state that an information system holds three structures:

(i) Deep structure: the aspects of the expressions of stakeholders represented in the information system (for example, data objects and business rules created into code).

(ii) Surface structure: the aspects that let the information system user engage with the deep structure (for example, query interface and report generator).

(iii) Physical structure: the aspects that combine both structures (for example, computer networks and laptops).

RT focuses on the deep structure of the information system. Consequently, the main focus of RT is the degree to which the deep structure of an information system offers and sustains an accurate representation of the central real-world phenomena (Wand & Weber, 1995).

RT conceptualizes the deep structure of an information system as scripts representing real events. These scripts transform until a machine-executable script is obtained. Grammar and properties of scripts affect their ability to represent real-world phenomena and communicate meaning truthfully (Burton-Jones et al., 2017).

Three representation models

Wand and Weber (1993, 1995) developed three models (i) The Representation Model, (ii) The State-tracking Model, and (iii) The Good Decomposition Model, to explore how the deep structure of an information system can faithfully represent real-world phenomena and communicate meaning (Burton-Jones et al., 2017):

(i) The representation model (RM) focuses on generating scripts using grammar in information systems. The RM emphasizes maintaining a faithful representation of real-world phenomena in these scripts (Wand & Weber, 1993). Any loss of faithfulness is predicted to reduce the usability of the implemented information system. Wand and Weber (1993) evaluate the ability of grammars to provide faithful representation by examining the constructs used to represent specific types of phenomena. They identify four types of defects in grammar that can undermine their ability to generate clear and complete scripts: construct deficit, construct excess, construct redundancy,

and construct overload (Wand & Weber, 1993). Mapping constructs in grammar assess these defects to a reference ontology, such as Bunge's ontological theory (1977, 1979).

(ii) The state-tracking model (STM) proposed by Wand and Weber (1995) suggests four conditions an information system must satisfy to maintain a faithful representation of real-world phenomena as they change.

a. Mapping condition: The information system must map a single state of the real phenomena so that users can interpret the state of the information system and assign meaning to it that unambiguously corresponds to the real state.
b. Tracking condition: The information system must accurately track internal state changes within real-world phenomena caused by interactions within its limit.
c. External event condition: The information system must faithfully reflect changes in real-world phenomena caused by external environmental events.
d. Sequence condition: The information system must record and update the sequence of external events in the same order as perceived by the real phenomena to maintain a faithful representation of the sequence of internal events.

The STM and the RM are related, with the STM assuming that the machine-readable script implemented by the information system can provide a faithful representation of real-world phenomena. The conditions of the STM cannot be met if the script does not clearly distinguish between things and events.

(iii) The Good-Decomposition Model (GDM), proposed by Wand and Weber (1995), focuses on creating representations in information systems with "good decompositions" of real systems, which communicate more meaning about the modeled phenomena. The GDM sets several requirements that a decomposition (a hierarchical structure of systems and subsystems) must satisfy better to communicate the meaning of real-world phenomena to stakeholders.

The GDM assumes that the script representing real-world phenomena can provide a faithful representation, similar to the Representation Model (RM). For example, if the script cannot separate things from events, the requirements of the GDM cannot be met. Initially, three necessary conditions were proposed:

a. Determinism Condition: The decomposition must ensure that events induced in each subsystem are either external or well-defined internal events, allowing clear prediction of subsequent states.

b. Minimization condition: The decomposition must eliminate redundant state variables and ensure only necessary variables represent focal real-world phenomena.

c. Losslessness condition: The decomposition must preserve all heritable and emergent properties of the real-world phenomena.

Further drawn on works of, for example, Simon (1996) and Yourdon & Constantine (1979), two additional necessary conditions were later proposed by Weber (1997):

d. Condition of maximum cohesion: Subsystems must be maximally cohesive, meaning that output-state variables cannot be partitioned based on input-state variables. This ensures strong connectivity within subsystems.

e. Minimum coupling condition: The decomposition should have minimum coupling, focusing on reducing the number of external events in subsystems and promoting well-defined internal events. This increases predictability and facilitates understanding of events in the subsystem.

The GDM balances the criteria of coupling and coherence in good decomposition and strives for a harmonious relationship.

In conclusion, standardized metadata representation is crucial for communicating meaning and understanding complex data across systems and domains (Dai et al., 2021; Mandal et al., 2016; Melo et al., 2021). It enables data integration, knowledge discovery, and informed decision-making (Pomerantz, 2015). Advanced technologies like semantic-based approaches and 3D modeling enhance information management and asset understanding (Rytsar et al., 2003; Yen et al., 2013). However, challenges related to structure (Huang et al., 2017; Park & Tosaka, 2010), semantics (Eichenlaub et al., 2021), human interaction (Eichenlaub et al., 2021), metadata lifecycle (Li et al., 2012; Maumet et al., 2016; Papež & Mouček, 2017), and processing (Ashish et al., 2016; Daniel et al., 2014; Deppenwiese et al., 2019; Jeong et al., 2014; Kock-Schoppenhauer et al., 2019; Ngouongo et al., 2013; Song et al., 2014; Specka et al., 2019) need to be addressed for effective system interoperability and improved metadata representation (Ulrich et al., 2022).

Representation Theory provides insight on how information systems communicate meaning (Wand & Weber, 1995). The Representation Model, State-tracking Model, and Good Decomposition Model provide frameworks for evaluating grammar, state-tracking capabilities, and decomposition quality in order to ensure accurate representation and meaningful communication of real-world phenomena (Wand & Weber, 1993, 1995).

To maximize the value of data assets, organizations should embrace standardized and structured metadata representation, which improves data governance, data quality, and communication. This enables knowledge discovery, informed decision-making, and competitiveness in the digital landscape.

2.3 Metadata & Data Quality

Metadata of high quality has been shown to influence decision outcomes and increase decision-making accuracy and reliability (Dyson & Foster, 1982; Harley & Cooper, 2021; Price & Shanks, 2011; Shankaranarayanan et al., 2006, 2008; Shankaranarayanan & Zhu, 2021; Stvilia et al., 2007). Furthermore, metadata-driven data quality frameworks have been used in various contexts, including e-government and statistical agencies, demonstrating their practical relevance and impact (Dion, 2007; Myrseth et al., 2011).

Data Quality is not a concept from the last years. It has already been introduced in papers in the 1980s (Brodie, 1980; Woodward & Masters, 1989). These papers introduce data quality for information systems and seismic networks. When data is flowing through systems, it can degrade over time if there are no control mechanisms over the data (Batini et al., 2009; Dasu, 2013).

For organizations and processes that depend on data, the quality of the data must be reliable and correct to drive decisions and actions (Harley & Cooper, 2021; Stvilia et al., 2007). Poor data quality can cause organizations severe loss because it affects the analytical results of business intelligence or data warehouses (Cichy & Rass, 2019; English, 2009; Grover et al., 2018). Multiple initiatives have been launched to prevent these losses from poor data quality. Examples are the Data Quality Act and the Data Quality Assessment Methods and Tools (DatQAM) endorsed by the European Commission (Bergdahl et al., 2007). Wand & Wang (1996) found that data quality is mostly measured through the following dimensions: (i) Accuracy; (ii) Reliability/consistency; (iii) Timeliness (currency); and (iv) Completeness. As previously stated, there is the Data Management Body of Knowledge (DMBOK) from the Data Management Association (DAMA). The book describes eleven core areas with best practices to align data management in an organization, where one area is focused specifically on data quality.

In the field of data, many studies have already been conducted. However, most studies are focused on one theme within the broad data field (Abraham et al., 2019; Batini et al., 2009; Brynjolfsson et al., 2011; Harley & Cooper, 2021; Timmerman & Bronselaer, 2019; Ulrich et al., 2022; Zhang et al., 2019). There have been conducted many studies in the data quality area. However, organizations still struggle with data quality (Cichy & Rass, 2019; Liu et al., 2021; Zhang et al., 2019). According to a multi-region survey conducted by Côrte-Real et al. (2020), some organizations have not matured their processes for adopting data quality. This implies that data quality improvement is conducted in departments on a decentralized or as-needed basis (Karkošková, 2023). Often, no centralized data quality management is aligned with the strategic and operational needs (Karkošková, 2023).

Metadata is essential for data governance and data quality control in organizations. As mentioned in section 2.1, it provides critical information about data quality, helps decision-making, and improves data reliability and usability. Several studies have stressed the significance of metadata

in ensuring accurate, consistent, and reliable data (Dion, 2007; Myrseth et al., 2011; Verbitskiy & Yeoh, 2011).

Interoperability and standardization efforts are critical parts of metadata and data quality management. Standardized metadata representations and exchange methods allow seamless data sharing, collaboration, and comparison across organizations and domains (Becker et al., 2009). This standardization guarantees consistent data quality assessments and promotes effective data governance processes.

To take control over data quality, DAMA International (2017) states that a data quality program is more effective when it is part of a data governance program. Good data governance requires the development of a data management framework, among other principles. The data quality program team collaborates with numerous stakeholders and supporters to achieve it. The volume and complexity of data that organizations must deal with can be complex. Additionally, there may be significant challenges if this enormous amount of data needs to be managed appropriately (Dhillon, 2019).

In conclusion, metadata is critical in data governance and data quality management. Standardized representations facilitate data exchange, collaboration, and comparison (Ulrich et al., 2022). However, organizations continue to face challenges with controlling and increasing data quality (Cichy & Rass, 2019; Liu et al., 2021; Zhang et al., 2019). It is essential to integrate data quality programs with data governance tasks (DAMA International, 2017). A comprehensive data management framework based on best practices such as the DMBOK can assist in navigating data complexities and ensuring consistent data quality management (DAMA International, 2017). Organizations can increase data quality, manage risks, and unlock the full potential of their data assets by understanding the value of metadata, adopting comprehensive data quality programs, and embracing data governance principles.

2.4 Metadata & Data Governance

Several studies emphasize the importance of metadata in efficient data governance (Aamot, 2022; Loshin, 2015; NICOLESCU, 2019; Sundarraj & Rajkamal, 2019). Metadata provides important context and descriptions for data assets, supporting their discovery, understanding, and proper use in organizations. It supports numerous data governance operations and is an essential data quality control facilitator.

Another frequent theme is the challenges associated with metadata management. Studies highlight the difficulties of capturing, organizing, and retaining metadata (IEEE, 2020; Loshin, 2015; Sundarraj & Rajkamal, 2019; van Helvoirt & Weigand, 2015). Among these difficulties are guaranteeing metadata accuracy, consistency, and completeness and dealing with metadata integration and interoperability issues. Effective metadata management strategies are critical for overcoming these issues and assuring data assets' dependability and usability.

Several articles discuss the importance of metadata standards and models (Aamot, 2022; IEEE, 2020; Yan et al., 2022). Metadata standards provide a standardized framework for organizing and managing metadata, thus enhancing interoperability and consistency across systems and domains (Aamot, 2022; IEEE, 2020; Yan et al., 2022). Metadata standards development and acceptance help accelerate metadata governance activities and improve data quality management. Similarly, metadata models, such as ontology-based models, provide semantic structures that improve metadata understanding, integration, and utilization (Yan et al., 2022).

To align the organization with data, it is necessary to implement data governance. The definition of data governance given by the DMBOK is *"the exercise of authority and control (planning, monitoring, and enforcement) over the management of data assets"* (DAMA International, 2017, p. 67). Another definition given by Abraham et al. (2019) is *"Data governance specifies a cross-functional framework for managing data as a strategic enterprise asset. In doing so, data governance specifies decision rights and accountabilities for an organization's decision-making about its data. Furthermore, data governance formalizes data policies, standards, and procedures and monitors compliance"* (p. 425-426). This definition is extracted from papers found for their literature review. Looking more closely at both definitions, we can find an overlap between them. Data governance is the control over their data assets. It specifically creates policies, standards, and procedures and monitors compliance.

Data governance is about a people's perspective. There is a need for a data management approach to manage all these data sources. According to the DMBOK, data management is *"the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles "*(DAMA International, 2017, p. 17). This definition does not only imply how the technical aspect of data should be managed but also from a business perspective and people perspective to manage data throughout the entire lifecycle. Thus, data governance is more about the people's perspective and who makes what decisions. In contrast, data management is more about what is needed to bring the decisions into business (Dyché & Levy, 2006; Hagmann, 2013; Khatri & Brown, 2010; Otto, 2013).

In conclusion, metadata is crucial to effective data governance by facilitating data discovery, understanding, and quality control in organizations (Pomerantz, 2015). However, metadata management presents challenges in terms of capturing, organizing, and preserving correct and consistent metadata (IEEE, 2020; Loshin, 2015; Sundarraj & Rajkamal, 2019; van Helvoirt & Weigand, 2015). The implementation of defined metadata standards and models assists in addressing these challenges while also promoting interoperability and consistency (Aamot, 2022; IEEE, 2020; Yan et al., 2022). Data governance is essential for organizations to keep control and responsibility over their data assets, implement policies and procedures, and ensure compliance (DAMA International, 2017). Data governance is concerned with decision-making and responsibility, whereas data management is concerned with the technical and business

elements of data handling throughout its lifecycle (DAMA International, 2017; Dyché & Levy, 2006; Hagmann, 2013; Khatri & Brown, 2010; Otto, 2013). Effective data governance approaches require both data governance and data management (DAMA International, 2017). Organizations may improve the reliability, usefulness, and overall data quality of their data assets by recognizing the value of metadata, adopting standardized methodologies, and building effective data governance frameworks.

2.5 Summary and conclusion

The theoretical background emphasizes the critical significance of metadata representation in organizational data governance and data quality. Metadata is an important component that provides context, description, and structure to data assets. It functions like a catalog, making it easier to find and identify relevant information. Organizations can collect essential elements that assure optimal storage, preservation, and accessibility of their data using descriptive, administrative, and structural metadata.

Standardized metadata representation is critical for effortless data exchange, collaboration, and comparison across domains and organizations. Standardized metadata supports successful data integration, knowledge discovery, and informed decision-making by fostering consistency and interoperability. Technologies such as relational databases, XML, and Linked Data help to structure and standardize metadata storage and sharing, increasing its usefulness and value.

However, challenges arise as a result of the lack of structured and standardized metadata usage, as well as human interaction-related issues. These challenges emerge in the capture, organization, and retention of correct and consistent metadata, resulting in fragmented efforts and unsatisfactory consequences.

Moreover, existing studies on metadata, data governance, and data quality predominantly focus on either one of these topics separately (as mentioned in the literature reviews by Abraham et al. (2019), Timmerman & Bronselaer (2019), and Ulrich et al. (2022)). Subsequently, the semi-literature review (Appendix I: Design of the literature review) has selected seven articles for metadata representation, seven articles for metadata & data governance, and twelve articles for metadata & data quality. This highlights the understudied field of the topics. Finally, due to the understudied topics, it is difficult to establish a conceptual framework.

As a result of the lack of structured and standardized usage and the human interaction-related problem, and the understudied fields, the research will explore how organizations represent metadata and how this affects data governance and data quality. This leads to the following research question:

"How could the representation of metadata affect data governance and data quality in organizations?"

In conclusion, the theoretical background highlights the significance of metadata representation in data governance and data quality. The adoption of structured and standardized metadata practices is crucial for organizations to improve their data governance practices, enhance data quality, and harness the full potential of their data assets. By addressing the challenges related to metadata usage, presentation, and human interaction, organizations can establish effective data management frameworks and ensure the successful integration of metadata within data governance and data quality initiatives.

3. Methodology

This chapter provides an overview of the steps taken to conduct this thesis. First, the reasoning for the research method is explained. Secondly, the process of data collection is presented. Finally, an overview is given of how the data is analyzed.

3.1 Research Method

This study uses a qualitative approach throughout the research. The qualitative approach has been chosen because of the nature of the research question, *"How could the representation of Metadata affect Data Governance and Data Quality in organizations?"* as mentioned in the Problem Statement and Theoretical Background; the literature on the topics of metadata representation combined with data governance and data quality is limited and still developing, which necessitates an exploratory approach.

Exploratory research is used when the topic is understudied or new to a field (Birkinshaw et al., 2011; Mills et al., 2009; Yin, 2018). As addressed in the Problem Statement and Theoretical Background, the topics Metadata Representation & Data Governance, Metadata Representation & Data Quality are understudied and do not have sufficient literature to define a conceptual framework. The majority of the studies in relation to this - such as data governance, data quality, and metadata and its representation - are (i) limited to one of the three fields or (ii) too broad (Abraham et al., 2019; Timmerman & Bronselaer, 2019; Ulrich et al., 2022). Thus, this study aims to contribute by integrating these three topics.

To answer the research question, this study requires analyzing multiple cases to define a conceptual framework. This is an exploratory multiple-case study (Mills et al., 2009; Yin, 2018). According to Yin (2018), there are three conditions to adopt a case study as the research method; (i) the form of the research question is 'How' or 'Why'; (ii) it does not require control over observable events; and (iii) the focus of the research lies on the early past and present event. The research question checks the first condition, starting with 'How'. Secondly, the research does not require controlling the events, but rather the opposite, because the information must come from interviewees willing to provide input for this study. Lastly, the focus of this research is based on recent and current events. The form of a multiple case study is chosen because the data gathered from multiple cases is often more liable, implicating that the multiple case study is more robust (Yin, 2018).

As mentioned, this research has an exploration focus. Mills et al. (2009), states that an exploratory case study aims to explore an unknown (or understudied) concept. An exploratory case study would benefit from cases that have dealt with the research topic (Mills et al., 2009). In this study, those topics are Metadata Representation & Data Governance and Metadata Representation & Data Quality. To find and create a conceptual model based on multiple cases, this research chooses the Grounded Theory approach. The utilization of the Grounded Theory

approach aims to generate new theories and enhance the existing knowledge within the field. Exploratory studies have multiple benefits. Among these are: (i) exploring understudied and new areas to generate new insights, hypotheses, and ideas; (ii) requiring fewer resources and smaller sample size; (iii) providing the groundwork for follow-up research (Cuthill, 2002; Eisenhardt, 1989; Salkind, 2010; Streb, 2010; Taylor et al., 2002).

The Grounded Theory approach is used to generate new theories and contributes to the existing knowledge in the field, initially developed by Barney Glaser & Anselm Strauss in 1967 (Mills et al., 2009). Later versions of Grounded theory were added (Strauss & Corbin, 1998) that Grounded Theory leads to a set of relationships that provide a plausible reason for the subject of their study, connecting multiple facts logically, practically, and pragmatically. In addition, it can reveal the hidden or the unrecognized (Mills et al., 2009). Many qualitative studies use Grounded Theory and have shown its usefulness in modernist and interpretative schools (Bryant & Charmaz, 2007; Locke, 2001; Morse, 2016; Yamazaki et al., 2009). To read more about the modernist and interpretivist, see the work of Mills et al. (2009). The Grounded Theory approach has been developed and changed over the years by multiple authors and has formed three main views (Charmaz, 2000, 2016): (i) constructivist; (ii) objectivist; (iii) postpositivist. This research uses a constructivist view because of its nature which emphasizes the subject matter being studied and views the collection of data and analysis as being a product of relationships and experiences that are shared with participants (Bryant, 2002, 2003; Bryant & Charmaz, 2007; Charmaz, 1990, 2000, 2006). Different organizations are chosen to gather data, assuming each organization has different viewpoints. The relationship with participants is expected to be of importance. In contrast to the constructivist view, the objectivist view does not view data collection and analysis as a product of relationships and experiences shared with participants. Instead, they see themselves as objective analysts (Charmaz & Belgrave, 2012). At last, the postpositivist view aims at contributing new theories to the field of study (Charmaz & Belgrave, 2012), which this study does not aim for. Instead, provide a deeper understanding of the topic and insights for future research. This study is highly exploratory and interpretative, which is suitable for using the Grounded Theory (Charmaz, 2000, 2016; Charmaz & Belgrave, 2012).

3.2 Data Collection

Unit of Analysis

The unit of analysis for this study is the effect of metadata representation in organizations on data governance and data quality. The organizations differ and operate in multiple industries, from financial services to production. As mentioned in the Problem Statement, there is no standard use of metadata. Therefore, it is expected that each organization represents its metadata differently. To generalize the findings, finding organizations operating in different industries is essential.

Data sources

Primary source

Interviews

Yin (2018) states that to increase the validity of the research, different perspectives of organizations should be gathered to have different perspectives on the topic. This research has conducted multiple interviews with six organizations and nine experts. The participants' job functions ranged from the strategic level to the operational level. Therefore, the validity of the research is strengthened by the different perspectives that have shed light on the focus of this research.

This study uses semi-structured interviews as the primary resource for data collection. Using interviewing offers an interpretive analysis of the gathered data and the ability to ask follow-up questions during the interview (Charmaz & Belgrave, 2012; Mills et al., 2009). Despite the interview having an in-depth semi-structured format of open-ended questions, there were several side steps from these questions during the interview. The interviews were conducted online and on-site at the organization's premises. The interviewees have been identified through Clever Republic and personal networks. The interviews were recorded on-site through the audio recording app and online through the inbuild recording function of the meeting program. The interviews were conducted in Dutch or English. The language depended on what the interviewees' language was. The Dutch interviews have been translated into English since the primary language of this thesis is English.

The interview was structured as follows. The interview started with some initial questions, followed by intermediate questions, and at the end finalizing questions (Charmaz & Belgrave, 2012).

(i) Initial questions:
These questions are used to gather information about the participant, related to their job position, and have an open starting point (Charmaz & Belgrave, 2012).

(ii) Intermediate questions:
These questions are aimed explicitly at the core concepts in the study: (a) Metadata; (b) Metadata Representation; (c) Data Governance; (d) Data Quality. The questions will provide the participants' views on the topics and an understanding of how the topics are implemented in the organization.

(iii) Finalizing questions:
The questions at the end are used to transition the conversation back to normal (Charmaz & Belgrave, 2012).

See Appendix II: Interview structure for all interview questions and Table 11 for the rationale.

Secondary source

This research used desk research as a secondary resource to gather relevant information for this study. Academic articles, books, and web pages from newspapers, government agencies, and

standards organizations were used to gather secondary information. Appendix I: Design of the literature review shows the selection process of articles used in the Theoretical Background.

Case Selection

Yin (2018) states that to conduct a multiple case study, the design of the cases should follow replication. This is comparable to conducting multiple experiments, where the conditions should be almost identical. He states that the cases should be selected carefully so that all cases comply with at least one of the following conditions:

(i) Predicting the same results (literal replication).
(ii) Predicting opposite results but for foreseen reasons (theoretical replication).

This research focuses on the first condition as the second condition will be less likely to be achieved, due to the exploratory nature of this research. Therefore, the organizations and cases are selected carefully to have as many similarities as possible to estimate this before conducting the research.

The approach of case selection is important when conducting qualitative research (Robinson, 2014). Robinson (2014) mentions in his article that the Grounded Theory, among other qualitative methods, uses interviews as their main data source. Therefore, the cases were selected using the four-point approach to qualitative sampling created by Robinson (2014). The four-point approach consists of:

(i) Defining a sample universe.
(ii) Deciding the size of the sample.
(iii) A purposive sampling strategy.
(iv) Sourcing sample.

See Appendix III for the definitions and key points of the approach.

Sample universe

The cases used in this thesis are all cases that were selected through inclusion and exclusion criteria. By having inclusion and exclusion criteria, the sample can be reduced (Luborsky & Rubinstein, 1995; Patton, 1990). Robinson (2014) states that the more inclusion and exclusion criteria there are, the more homogenous the sample universe is. On the contrary, for this research, a heterogeneous sample is required since this research focuses on Metadata Representation in different organizations and uses the Grounded Theory approach to answer the research question (Strauss & Corbin, 1998). Robinson (2014) states that the rationale for a heterogenous sample is that it would be more generalizable if there were overlapping findings. See Appendix IV for the graphical sample universe used in this study.

As mentioned above, organizations have to meet certain criteria to be selected. The inclusion and exclusion criteria the organizations have to meet are:

Inclusion criteria

- The organization fits the 'large organization' terms provided by the Ministerie van Economische Zaken en Klimaat (2021). They state that a large firm has employed over 250 FTE; or
  The net sales are over €50 million, and a more than €43 million balance sheet.
- The organization is working with metadata.
- The organization utilizes data governance.
- The organization has data quality measures.
- The organization operates in the Netherlands but is not limited to the Netherlands.

Exclusion criteria

- The organization is not an SME.

Size of the sample

The sample size represents the number of cases that should be used in the case study to gain saturation and reach generalizable conclusions. In Grounded Theory, saturation is reached when no more contributions to the analysis are found (Strauss & Corbin, 1998). This research has conducted nine interviews at six organizations.

Sampling strategy

There are two main sampling strategies: (i) random/convenience sampling; and (ii) purposive sampling. Random sampling is selecting random cases from the sample universe. This is mainly used in opinion polls and social research surveys (Robinson, 2014), which is not used in this research. Convenience sampling is choosing cases within the interviewer's reach and with willing participants (Robinson, 2014). Purposive sampling is the non-random selection of cases from the sample universe that are important to select, according to the researcher (Robinson, 2014). There are different types of purposive sampling; one is theoretical sampling, which is used in this study. Theoretical sampling is associated with the Grounded Theory method (Robinson, 2014).

Sourcing sample

Sourcing a sample is recruiting participants for the study. The participants were found and selected through Clever Republic B.V. and my network. The participants were contacted through email, LinkedIn, and by phone. Six organizations and nine people were willing to participate in the research.

Sample

All organizations were granted full anonymity. Therefore, the participants and the organizations' names are completely anonymized. The organization names are named 'Organization X,' where 'X' is a letter of the alphabet. See Table 1 for information on how the organizations are referred, the function of the interviewees, and the date when the interview was conducted. If there are more interviewees from an organization, the interviewees are referred to as 'X[number].'

Table 1 - Interview information

Organization	ID	Job Function	Date (2023)
	A1	Head of Data Management	15th of May
Organization A	A2	Data Governance Specialist	24th of May
	A3	Data Governance Lead	31st of May
Organization B	B1	Data Quality Manager	16th of May
	B2	Jr. Business Consultant	25th of May
Organization C	C	Data Analyst	23rd of May
Organization D	D	Data Engineer	24th of May
Organization E	E	Data Governance Lead	26th of May
Organization F	F	Product Owner Metadata Management	31st of May

The following section briefly describes the organizations, the interviewees, and why the interviewee is relevant to this research.

Organization A

Organization A grew from a small company to a globally operating company today. Their main activities include providing the leasing of cars. A few years ago, they focused more on their digital department area, which has led to a strategy for their digital parts. Three participants have been interviewed from organizations A. Interviewee A3's job function is the Data Governance Lead. Interviewee A3 is operating between A1 and A2. The different levels at which the interviewees are operating will contribute to providing a broader view of the concepts used in this thesis. The job function of interviewee A1 is the Head of Data Management. Interviewee A1 has experience as a developer and went to the business side working as a business analyst. Interviewee A1 is relevant to this research due to its experience in operations as well as the strategic level of the organization. Interviewee A2 is working as a Data Governance Specialist. Interviewee A2 has experience in different fields in an analyst role. Interviewee A2 is working daily with metadata in his role. This is useful for creating a view based on daily operating tasks.

Organization B

Organization B is a major pension administrator and asset manager. They provide pension administration and advice, and responsible asset management services. The organization has over 450 employees working in the Netherlands. Two participants have been interviewed from

organization B. Interviewee B1 is the Data Quality Manager and has worked for the organization for six and a half years. Interviewee B1 has experience in the risk management field. To fulfill this task, interviewee B1 relies on data to fulfill his tasks. Currently, the tasks are more at the tactical and strategic level. Thus, the answers will contribute to this level of the concepts. Interviewee B2 is a Junior Business Consultant. B2 works daily with aligning and practicing metadata. This is useful for creating a view based on daily operating tasks.

Organization C

Organization E is an online grocery delivery service that offers consumers affordable groceries delivered directly to their residences. Customers can place orders via the app or website, and their groceries are delivered to their doorstep. Popularity has increased due to the service's innovative approach and commitment to customer experience. Interviewee C is working with data daily in the role of Data Analyst. The answers provided a view of the concepts in an organization that has focused on its digital infrastructure since it was founded.

Organization D

Organization D is a reputable insurance and asset management organization with over 16.000 employees worldwide. They provide an extensive selection of insurance products, pension plans, investment options, and banking services to meet the requirements of individuals and businesses. Interviewee D is a Data Engineer, working daily with data and data warehouses. Interviewee D has experience in different organizations as a Data Engineer. The experience with data warehouses interviewee D contributes to this research.

Organization E

Organization E organizes the supply chain of a worldwide beer brand. Their supply chain is part of the overall holding. It starts with the procurement of high-quality ingredients and advances to the production of beer in sophisticated breweries. Logistics and distribution are also a priority, as warehouses and distribution centers are strategically located. This supply chain guarantees their beer production, distribution, and delivery without delays. This is all based on data, which must be governed and of high quality. Interviewee E is the Data Governance lead within the supply chain branch in the Netherlands. Interviewee E has experience with data analysis and then became interested in the data governance field.

Organization F

Organization F is a technology organization in the semiconductor industry. The cutting-edge organizations' equipment enables them to be a world leader. This organization plays a crucial role in advancing the semiconductor industry. Currently, they have over 35.000 employees in locations worldwide. Interviewee F is the product owner of metadata management.

3.3 Data Analysis

> "You need to get lost before you can get found."
> (Gioia, 2004, p. 103)

This study used an inductive approach to analyze the data. The above quote represents the feeling when the interview data was being analyzed. A quote from Yin (2018) mentions "playing with the data" (p. 169) to understand and find concepts and patterns in the data. This could lead to the first step of getting insights from the data. This aligns with the Grounded Theory approach (Glaser & Strauss, 1967; Strauss & Corbin, 1998; Yin, 2018).

Understanding the steps taken to analyze the data requires a structured approach. This structure is outlined as follows: (i) Information about how the interviews were recorded and transcribed is given; (ii) The coding process and the program used to code the interviews.

First, Microsoft Word automatically transcribed the audio. Then the transcription was read over with the audio playing next to it to adjust words that were not transcribed correctly. Coding the interview has been done using the program ATLAS.ti[4]. This software is widely known for its tool to help coding in qualitative data analysis. The software is also called Computer-Aided Qualitative Data Analysis Software (CAQDAS) (Friese, 2016).

Before analyzing the interviews, a coding strategy had to be set up. Gehman et al. (2017) compared three major approaches to qualitative theory building. The approaches were from (i) Gioia et al. (2012); (ii) Eisenhardt et al. (2016); and (iii) Langley (1999). This research used the first approach from Gioia et al. (2012) to some extent. The approach is helpful for the interpretative Grounded Theory approach (Gehman et al., 2017; Gioia et al., 2012). The approach emerges around the idea that the interaction with the world is based on social constructs (Berger & Luckmann, 1967; Schutz, 1972; Weick, 1979).

Since the recent introduction of AI models, the world has started to use and implement the models in different ways. ATLAS.ti has also used this opportunity. Their beta feature, AI coding[5], can help with the coding process of the interviews. This feature decreases the time spent on the coding of the interviews. This study used AI coding to start the coding process. This is viewed as line-by-line coding in the approach of Gioia et al. (2012). After the AI coding, 298 codes were grouped into nine category codes. Since 298 codes were too many to have a good overview and some were almost identical. Codes that were set in the paragraphs from the interviewer were deleted. They resulted in 275 codes.

[4] See their website for more information https://atlasti.com/
[5] See this webpage for information about ATLAS.ti AI coding

Subsequently, three category codes were found to be of no value from the nine category codes. The category codes were Referencing, Research Methodology, and Attention to detail. Referencing and Attention to Detail were deleted because the codes did not seem to add insight. Research Methodology was deleted because the codes were mostly coded in questions and answers of the interviewer. See Appendix VI for the three categories with their sub-codes. The codes were replaced if they seemed to fit with those existing in the remaining category codes. Next, codes were created in a group called 'Section' to indicate the text sections of the structure of the interview. These were all added to the quotations. In total, five sections were created: (i) Metadata general; (ii) Metadata Representation general; (iii) Metadata Representation in DG; (iv) Metadata Representation in DQ; and (v) Finalizing.

Secondly, the synonyms or almost identical codes were grouped and merged. This resulted in 147 different codes spread over five category codes (the section category codes are only used for structure purposes). Table 2 provides an overview of the categories and their groundedness. Groundedness is the sub-code distinct counter quote when the sub-codes have been coded (ATLAS.ti, n.d.). The unique codes can be seen in ATLAS.ti with their groundedness for each category code.

The provided answers from the interviews were analyzed through the four main concepts in this study: metadata, metadata representation, metadata & data governance, and metadata & data quality. The codes per section were compiled to create first-order concepts. Each theme that resulted from the analysis of the provided answers were assigned to one another. This resulted in five high-over categories (See Table 14, Table 15, Table 16, Table 17, and Table 18 in Appendix VIII: Categories and themes to data structure). The unique codes per category have been deduplicated. The codes per category were divided into groups of codes, and the first-order concepts were made from these groups of codes. The second-order themes were derived from these concepts. Finally, the second-order themes have been aggregated into three dimensions: (i) Improving understanding; (ii) Data proficiency; and (iii) Data Governance Lifecycle. Categories. (See Table 19 for the first-order concepts, themes, and dimensions in Appendix VIII: Categories and themes to data structure). In Figure 2 in the Discussion, the data structure can be seen.

Table 2 - Category codes with Groundedness

Category code	Groundedness
Data management	152
Technology	105
Personal and organizational development	71
Clarity/understanding	66
Business management	53
Data quality	28

Validity and reliability

This research has adopted multiple strategies to ensure the validity and reliability of the results. Respondent validation, together with Clever Republic the participants were selected for the research. As mentioned earlier in the unit of analysis and data sources section, multiple views of organizations and participants have been captured in the research to generalize the findings over the sample universe. Finally, the results have been presented in a session with multiple employees of Clever Republic, where the results have been found of practical relevance. According to Backman & Kyngäs (1999), in the Grounded Theory, the results can be validated by experts who are not part of the research. The employees of Clever Republic have not been interviewed.

Summary of the methodology

This chapter has outlined the reasoning for the chosen research methodology, the Grounded Theory approach. The method of data collection and what samples have been selected was presented. The samples have been selected through the four-point approach of Robinson (2014). Finally, the process of data analysis has been provided. To some extent, the approach of Gioia et al. (2012) has been used to analyze the interview data. In the next chapter, the results from the interviews will be given. This will be done in the structure of the interview. The results start with the rating questions.

4. Results

This chapter will present the results found in the analysis of the interviews. The chapter has the structure of the interviews that were conducted. The answers from participants will be "quoted" and put in *italic*. Next, all participants have a corresponding ID, as seen in Table 1. For each interview section, codes have been created. See Table 3 for the section and codes. The quotes will be referred to as 'X[number].[code].

Table 3 - Interview sections with codes

Section	Code
Personal information	P
Company information	C
Metadata general	MD
Metadata representation general	MDR
Metadata representation in data governance	MDRDG
Metadata representation in data quality	MDRDQ

The outline of the results will be presented as follows. First, the questions with ratings will be presented. Then, metadata in the view of the participants is given. Subsequently, a view on Metadata Representation is provided. Finally, the representation of metadata in a data governance and data quality context with the definitions provided by the interviewees of data governance and data quality is given.

4.1 Rating questions

This section provides the results from the six rating questions asked during the interviews. The results are outlined in Table 4. A high number means more use or a better rating for that part, whereas a lower number means less or a lower rating. The questions were:

1. To what extent do you have to deal with metadata in your current role? (range 1-6)
2. To what extent do you see metadata as a beneficial asset to your organization? (range 1-6)
3. To what extent would you rate your organization to be data-driven? (range 1-6)
4. How does this rate compare to your competitors? (range 1-6)
5. Do you feel your organization makes sufficient use of potential information? (range 1-6)
6. How important is metadata in your organization? Could you rate this? (range 1-6)

Table 4 - Results of the rating questions

Question/ Participant ID	Q1	Q2	Q3	Q4	Q5	Q6
Participant A1	3	5	4	4	3	4
Participant A2	6	4	5	4	5	3
Participant A3	-	-	4	-	-	3

Participant B1	4	5	3	5	3	5
Participant B2	4	5	3	5	4	5
Participant C	6	6	6	-	5	6
Participant D	5	6	3	4	3	3
Participant E	4	6	3	4	3	3
Participant F	4	-	4	-	-	-
Average	4,5	5,3	3,9	4,3	3,7	4,0

A few questions were not answered or asked during the interviews. This was due to time constraints or interruptions and the semi-structured nature of the interview to have the interview be more of a conversation. Therefore, some participants have a dash in their row. Specifically, participants A3 and F have multiple dashes in their rows. This was due to the time constraint of 45 minutes and the choice to focus on different questions. The following paragraphs provide deeper insight per question.

Q1. To what extent do you have to deal with metadata in your current role? (range 1-6)

This question was asked to see how much the participants' position had to deal with metadata and to find if there were any relations between the answers given and the position.

First of all, what is seen is the lowest number given by participant A1. A1 mentions that they do not call out managing their metadata. They see it more as a byproduct of gaining information out of data.

"We don't do that. No, we do it as a kind of fallout from a byproduct of the fact that we're focusing on data relevance" – A1.P

Participants A2 and C both give the use of metadata the highest number. Participant A2 mentioned the use of metadata in his daily tasks.

"Yes then it is a 5 or a 6. I mean, I'm obviously just dealing with a daily one. Yes a daily with metadata working to connect it, but also to make sure it. That is verified, or it just gets put down in the right. What is it called? Yes in the right place, yes." – A2.P

Participant C mentions the use of a data warehouse every day.

"…Just indicated at those fact and dimension tables that is metadata for you." – Interviewer with C

"So, did I understand a little? Yes, and but in Maybe the general, the data warehouse we really use Everyone every day super much." – C.P

Other participants mentioned that they use metadata ranging from some extent to daily. It was found that participants have different views of metadata and responded with this in mind. In the last questions is asked for their definition. The average for question one is 4.0.

Q2. To what extent do you see metadata as a beneficial asset to your organization? (range 1-6)

This question was asked to capture the participants' view of the importance of metadata in their organization. This is particularly useful for the view of participants in the same organization.

Organizations A and B provided multiple participants. Organization A has one response missing. However, participant A3 states in P, *"…I live and breathe metadata, and when I say metadata, more on the business metadata than more on the technical metadata…"* This implies A3 greatly uses metadata in their daily tasks and sees it as a valuable asset. In organization B, participants rate the same number (five) for metadata as a beneficial asset. All other participants also rate metadata as a beneficial asset for their organization. Participant F did not mention a number. However, F speaks about assets in general and then the value of data and metadata as an asset.

"Yes, I don't think there's a company where metadata has no value [….]. The word asset suggests that something has a value, doesn't it? Asset Management is therefore of course also possible. To invest In the financial world, so assets that are you have assets and liabilities, so things of value [….]. Statements like data are an asset, aren't they? So it has value, because it has yes, it has a lot. Intrinsic meaning of how you run your company and what goes on in your company, eh? So all your products and customers and suppliers are represented in that data. In other words, all processes depend on them. Data and your information provision to be able to control depends on data and that. Is is is data, so to speak. The raw material with which you generate your end product from your business processes, in other words. Absolute data is an asset, metadata are assets [...]" – F.C

Question two has the highest average score of 5.3.

Q3. To what extent would you rate your organization to be data-driven? (range 1-6)

This question is asked to create insight into the data-drivenness of the organization. When looking at the numbers given, it is seen that there are more average numbers (3 and 4) than there are higher numbers.

Participants mention that their organizations are average when it comes to data-drivenness. Participant E mentions data quality as an essential factor in being data-driven.

"Yes, we use a lot of data. Without them realizing it and data driven means that you are also in control of the quality of your data and we are really not there yet. Only that a lot of People do not see that because they just take it for granted. They believe it and to really work data

driven means that you do indeed have full data dictionary and full data lineage mapped out, we are not there yet." – E.C

E.C mentions that people do not see that to go up on the data-driven scale, they have to be in control of the data and the quality of the data.

Participant B1.C states that data-drivenness consists of three parts: (i) metadata, Data Governance, and Data Quality are in place and control; (ii) Data tooling is needed to align the first three; and (iii) Data culture/literacy is in place. Specifically, the latter costs energy and takes a while before it is in place. B1.C also mentions that it is different in the organization per department. This is something also mentioned by participant F:

"[….] I don't think that's it the top priority is, but one of the many important capabilities within data management that absolutely need attention. And that realization is there. Not everywhere, because People who really who work in the supply chain and who only negotiate prices with suppliers and import duties. Et cetera that one. Are also constantly working on data, but they might claim that they are. That give me good vendor Data and product data and financial data and metadata management. [….]" – F.C

Participant C.C defines their organization as highly data-driven because their focus was from day one at their data warehouse. This has increased their use of data.

Because of the many average numbers of three and four, the average is 3.9.

Q4. How does this rate compare to your competitors? (range 1-6)

This question was asked to find how organizations that operated in the same industry thought about their data-drivenness compared to their competitors. No organizations are operating in the same industry. Therefore, this question is not

The participants stated that it was a guess and not based on facts. When looking at the numbers, it can be seen that all numbers are four and above. Therefore, the average is 4.3.

Q5. Do you feel your organization makes sufficient use of potential information? (range 1-6)

Question five has the lowest average number, 3.7. Participants mention multiple statements for these lower numbers.

"[….] there we do not make full use of data with the full potential that is in it and that has mainly with the data quality and also just conservative In the nature of the employee." – E.C

Participant E mentions that the potential information is not reached if the data quality is incorrect. E also mentions the nature of employees that are still too conservative.

"I think Maybe a 2.5 or so or or 3. Of the 6. Yes, I think it's quite a lot being done with it already, but the possibilities are just endless, I guess. I think much more is possible. And yes, now with that one. Yes with the data we have now. Yes, we do have a number of sources in it now, but there is still years of work to be done, say to get everything in and then you can really link even more data and get even more value from it. Yes, we are not there yet, so to speak." – D.C

Participant D states that the years of work there still need to be done to fill the data warehouse. This decreases the use of data for potential information.

"[….] And I think the more you make your data transparent, the how, the more you find out that there's actually a lot more of us, so. My point of view, is that? Yes that is. That we're not making the most of it yet, so I always think we're using that data, but not yet in an efficient way." – B2.C

Participant B2 outlines the use of data in an efficient way. B2 also mentions the transparency of data to create insight. Participant B1.C also mentions the availability of the data.

Q6. How important is metadata in your organization? Could you rate this? (range 1-6)

This question was asked to understand the importance of metadata in the organization given in the participants' view.

"[…] so from my perspective that's very important and well a 5 or so of that per scale of 6. But it's not going to be like that, I think by Everyone seen, so It's just like? I think a lot of People will ask, okay, what is that metadata? […..]" – B1.MD

From the perspective of participant B1, metadata is essential. However, when the whole organization is in scope, the rate decreases to three.

"Just strive towards. Four you're never going to get you're you're kind of the North Star or. But yeah, we. We should strive towards being at least 4." – A3.MD

"And and and. How do you think it's right? It's a three or less." – Interviewer with A3

"Given the size of [COMPANY NAME]. I would say. We are. Yeah, three. If we're looking from the [COMPANY NAME] whole." - A3.MD

Participant A3 sees that metadata does not yet have the awareness it needs. The same holds for participants B1 and A2.

"Yes and the awareness, so how important they think it is to say yes how important is metadata for your organization. if You ask me is Of course. 6 nothing is more important than almost more important than that. That is especially when something breaks or when you have

to do impact analysis. If you really need to go back and look at your data to see what's actually happening […] you're super happy that you have that metadata, […] it's just keeping an eye on what's going on huh? Data should not be a Black box, you just have to know exactly what is happening with all the key data you have. Not all the data you have, that's nonsense, but all the key data you have just know what happens to that?" – A2.MD

Participant A2 also points out the need for metadata when something breaks or there is a need for tracking. A2 also mentions the insight into critical data and what is happening with it.

"[…] I'm mainly about more the definitions of Some data. Where we often use a lot of different names for the same thing, so It's more about the vocabulary around a not so much data, but also just even machine names. We have a lot of different names for that and that's already metadata. [….] actually they all mean the same thing and That's all metadata. Things are already going wrong at that level, which makes it very difficult. Each time you have to switch between different samples. As it were. And, you see that across the board. That there is no unambiguous definition there. English, Dutch mixed together. So yes, yes, it's a bit of chaos." – E.MD

Participant E mentions the description and vocabulary of data that is important. Everyone knows what an object is called in different systems.

4.2 Metadata

This section provides the results for the metadata-related questions. One of the questions was the definition of metadata in the view of participants. Not all experts have the exact definition of metadata. The themes found for the section 'Metadata general' are given in the following paragraphs. Finally, the view of the participants is given.

First, the codes related to the code 'Section: Metadata general' were filtered. This resulted in a list of 66 different citations. The codes that have been assigned to these citations were grouped into themes. The codes that occurred three times or more have been used to create themes. See Table 5 for the themes with the most important codes.

Table 5 - Themes for 'Metadata general'

Theme	Participant ID	Codes
Lack of clarity	A3; B1; B2; C; D; E	• Clarity/Understanding: Confusion • Clarity/Understanding: Uncertainty • Clarity/Understanding: Difficulty understanding • Data Management: Data literacy • Clarity/Understanding: Ambiguity • Personal and Organizational Development: Inefficiency

Recognition/Awareness	A1; A2; A3; B1; B2; C; E	• Data Management: Importance of metadata • Personal and Organizational Development: Awareness • Data Management: Importance of data • Clarity/Understanding: Clarifying • Data Management: Recognition of metadata
Data Management	A1; A2; A3; B1; B2; C; D; E; F	• Data Management: Data management • Data Management: Data analysis • Data Management: Data governance • Technology: Technical expertise • Business Management: Business processes
Metadata type	A1; A2; A3; B1; C; E ;F	• Technology: Metadata • Business Management: Business metadata • Technology: Technical metadata
Data type	A2; A3; C; E; F	• Data Management: Data privacy • Data Management: Data Quality

The citations can be found in Appendix VII. This is done to create a clear overview and not have many citations in the main text. Behind the paragraph's headings, the count of participants that mention the codes is given.

Lack of clarity (Participants = 6)

A3.MD highlights the need for guidelines or a defined strategy for managing metadata. A3.MD underlines the importance of understanding what metadata is for an organization, what the value of it is, and why it is important. A3.MD also highlights the risks of assuming of knowing what they do without the right guidelines and standards. This indicates the need for an organized approach to metadata management.

E.MD underlines the importance of describing data accurately and the establishment of good ownership and relationships. E.MD states that from the beginning, data quality should be monitored, making the right adjustments and ensuring the right system integration. E.MD acknowledges that the scope of metadata has an impact on the quality of data. E.MD suggests that by making effective use of metadata and providing clear definitions, metadata can have a positive influence on data quality.

All participants highlighted the lack of clarity and a defined strategy for metadata management. A3.MD emphasizes ensuring effective metadata management, guidelines, compliance with

standards, and clear direction is needed. E.MD highlights the importance of understanding metadata has a role in data quality and emphasizes the need for advanced planning and clear data descriptions. Both emphasize the importance of clarity, standards, and proactive approaches to achieve effective metadata management and improve data quality.

Recognition/Awareness (Participants = 7)

E.MD points out that there is a lack of standard metadata language and unclear data definitions. This leads to confusion and difficulties in metadata management. The ongoing struggle to act when there are metadata-related challenges is also mentioned.

In contrast, A1.MD recognizes that the awareness of metadata has been growing. A1.MD describes this as a concept that needs to be captured. A2.MD elaborates on the importance of awareness in understanding the role of metadata for managing key data and avoiding the black box phenomenon. This indicates a change to recognize the need for transparency and knowledge of data processes.

A3.MD highlights the different layers of metadata management. A3.MD highlights the importance of closing the gap between the technical and business metadata. A3.MD also mentions that technical metadata is often seen to be easier to deal with than business metadata. A3.MD emphasizes the need for efforts to help people to understand the value of metadata and what the implications are for the business.

The participants highlight the recognition and awareness of metadata. Despite the recognition of the importance of metadata, there still remain challenges in terms of vocabulary, definitions, and closing the gap between technical and business metadata. The findings highlight the need to increase awareness, education, and efforts to communicate the value and importance of metadata at different layers of the organization.

Data Management (Participants = 9)

A2.MD mentions the need for transparency in data management because they have to be aware of what is happening with their key data. B2.MD further highlights the importance of metadata in dealing with the exponential growth of data since metadata supports structuring and providing insight into the enormous amount of data. A2 and B2 both emphasize the increasing role of metadata in today’s data-driven world.

A3.MD recognizes the role of metadata management in the field of privacy issues such as GDPR and personally identifiable information (PII). Metadata can help with identifying and managing PII data. A3.MD also highlights the importance of metadata management in the storage and deletion of data.

F.MD and A2.MD emphasize that metadata management should focus on addressing specific pain points and solving problems rather than trying to capture all data. This highlights the importance of a focused metadata management strategy and stresses the importance of identifying domains where metadata management has the most impact.

All the participants highlight that managing data can be done effectively through metadata. This includes the understanding of key data, structuring metadata to deal with the growth of data, managing compliance and privacy issues, and handling specific pain points for effective metadata management. The insights from these observations will increase the understanding of data management within the metadata framework and support the formulation of resilient data management strategies.

Metadata type (Participants = 7)

Different participants mention the different types of metadata. They categorize metadata into two categories: technical and business metadata.

A3.MD mentions the need to split metadata to increase understanding and to find out where it adds value. A3.MD also highlights the need to implement data governance tools and metadata management practices to assure control and create an audit trail with data. This expresses the need for operationalizing metadata management to ensure effective data governance.

B1.MD also highlights the distinction between technical and business metadata. B1.MD mentions that business metadata is about the meaning, quality, and data governance aspects of data. In contrast, technical metadata is more technical information, such as storage and information systems. The aforementioned points highlight the understanding of the different forms and application areas of metadata.

F.MD explores the topic of metadata, covering its descriptive aspect as well as its manifestations in data models, error logs, and operational metadata. F.MD emphasizes the significance of linking business and technical metadata, as well as the challenges. In addition, F.MD recommends the use of trusted datasets as a solution to the need for data governance and trust in an era of complex data landscapes.

The participants demonstrate the variety of metadata. The perspectives range from categorizing metadata into business and technical categories to understanding the distinctions between various information kinds. The findings emphasize the growing significance of metadata in data governance, data quality, and building confidence in complex data settings.

Data type (Participants = 5)

E.MD evaluates their previous concept of metadata, which was seen primarily as labels and descriptions for fields and files. In recent years, E.MD has come to appreciate the greater

broadness of metadata and sees it as a discipline that requires the involvement of a Chief Data Officer. E. MD's attention has switched to the function of metadata in data quality, particularly in guaranteeing accurate and reliable data.

C.MD examines the difficulties and changes that arise with handling data types. When definitions change, C.MD mentions the need to rebuild data models. For example, when modifying 'orders' to 'delivery.' C.MD also emphasizes the significance of data lineage or tracking data changes over time. C.MD also underlines the rising necessity of data security and privacy, especially in light of GDPR regulations. As the organization grew, it became more cautious and vigilant about restricting access to personal information and sensitive data, emphasizing the ethical and privacy concerns involved.

The participants highlight the evolving understanding of data types, especially with regard to the comprehensive role of metadata. The perspectives range from viewing metadata as labels and descriptions to recognizing its impact on data quality, GDPR compliance, data retention, and privacy. The citations highlight the challenges in aligning data models with changing definitions and the increasing importance placed on data security and privacy in an organization.

Metadata definition

The definition of metadata was asked during the interview. Below are the definitions provided by the participants.

- A1.MD states that metadata is capturing information about the data models, data elements, and domains. It also involves gathering a lot of information about those data elements, which can be considered metadata.
- B1.MD first provides the broader definition: data about data. Then mentions that it can be categorized into different types, such as business metadata, which focuses on the meaning, quality, and data governance aspects of data. Technical metadata concerns the technical storage of data and information systems, while operational metadata relates to processing details and data access in systems.
- C.MD defines metadata as the information around an object. C.MD mentions as an example, imagine a picture, the picture itself is the data and the information about the picture is the metadata, such as when and where the picture was taken.
- A2.MD states that metadata is the combining of data and its description. A2.MD explains that it's information about the data, like the structure, format, and relationships between the data. As an example, A2.MD mentions that metadata can provide details about table names, column names, and more that will help in understanding the data.
- D.MD defines metadata from a data engineering perspective. D.MD states that metadata represents the description of data, which includes the columns, data types, nullability, and other characteristics that define the data.
- B2.MD defines metadata as information about data that is relevant and provides insights into the data. It includes various types of information related to data, both physical and conceptual, and plays a crucial role in facilitating data discovery and understanding.

- E.MD mentions that metadata can also include data quality scores, which provide an indication of the reliability of the data. However, assigning scores can be complex as it depends on the relationship between the data and its usage.
- A3.MD states that metadata can be divided into business metadata and technical metadata. Business metadata helps people understand the meaning and value of data, while technical metadata focuses on the technical details and supports data governance activities.
- F.MD defines metadata as descriptive data that provides a deeper understanding of the underlying data. It can take different forms, including data models, error logs, and the relationship between business metadata and technical metadata. The focus is on linking and aligning the business perspective with the technical aspects of data.

4.3 Metadata Representation

This section provides the results for the metadata representation related questions. One of the questions was the view of participants against the representation of metadata. Not all experts had the same view. In the next paragraphs the themes found for the section 'Metadata Representation general' are given. Finally, the view of the participants is given.

First, the codes that were related to the code 'Section: Metadata Representation general' were filtered. This resulted in a list of 45 different citations. The codes that have been assigned to these citations were grouped in themes. The codes that occurred two times or more have been used to create themes. See Table 6 for the themes with the most important codes.

Table 6 - Themes for 'Metadata Representation general'

Theme	Participant ID	Codes
Understanding the data	A1; A2; A3; B1; B2; E	• Clarity/Understanding: Confusion • Clarity/Understanding: Uncertainty • Clarity/Understanding: Ambiguity • Technology: Technical Jargon
(Meta)Data Management	A1; A2; A3; B2; C; D; E ; F	• Data Management: Data management • Business Management: Business processes • Data Management: Metadata Management • Personal and Organizational Development: Teamwork
Organization: Old VS New system	A1; A2; A3; B1; C; D; E;	• Personal and Organizational Development: Awareness • Data Management: Data capture • Technology: Legacy systems • Personal and Organizational Development: Efficiency
Data Management: Tools	A1; A2; B1; B2; C; D; E; F	• Data Management: Data modeling • Technology: Programming • Technology: Technical expertise

		• Technology: Technology • Data Management: Data cleaning • Technology: AI • Technology: Tooling

Understanding the data (Participants = 5)

A2.MDR emphasizes the need of making metadata descriptive and accessible to a wide audience, with an emphasis on keeping them understandable and avoiding technical jargon. This implies an emphasis on ensuring that metadata is user-friendly and easily understood by business users.

E.MDR explores the challenges of making additional actions or data details visible. E.MDR suggests that such information be made available only upon particular request, as it may need people using unfamiliar tools. This suggests a cautious approach to providing access to additional metadata information. E.MDR also references Collibra, a data management application, and its purpose of making data easily searchable. However, E.MDR acknowledges that users must still know what they are looking for and be able to efficiently utilize the tool, drawing a similarity with how Google search requires precise terms and the user must choose which results to click on.

Participants highlight the significance of making metadata descriptive, understandable, and accessible to a broad audience. It also recognizes the difficulties in providing access to extra metadata details, as well as the need for user familiarity with tools and search methods. The references indicate an emphasis on usability and the purpose of enabling people to understand and explore metadata in order to gain an improved understanding of the underlying data.

(Meta)Data Management (Participants = 8)

B2.MDR emphasizes the importance of extracting value from metadata and advocates for extensive information and data acquisition that provides concrete benefits. However, simplicity is still important because the goal is to make metadata straightforward and easy to understand for business users. The emphasis is on clearly explaining the data and assisting users in comprehending its purpose and limits. Focusing on user-friendly metadata presentation is critical for promoting its successful use.

E.MDR discusses the challenges and potential improvements in metadata management, particularly in the context of employing Collibra. E.MDR emphasizes the importance of properly integrating metadata with data consumption and input processes. In an ideal world, relevant information appears directly next to the data, maybe via augmented reality overlays or context-specific tooltips. This strategy seeks to reduce the number of activities required by users to get metadata, making it more accessible and seamlessly integrated into their processes. The

overarching goal is to create a data ecosystem in which users can easily retrieve and use metadata in order to improve data quality and decision-making processes.

A3.MDR highlights the significance of creating awareness and promoting understanding of metadata inside organizations. A3.MDR proposes illustrating the usefulness and relevance of metadata management with specific use cases. Organizations can promote a data literacy culture by including metadata management into these use cases and showing its impact. This method assists people in recognizing the importance of metadata in supporting data quality, analytics, and decision-making processes, hence driving metadata management practice adoption and acceptability.

F.MDR highlights the tooling element of metadata management, identifying Collibra as a regularly utilized platform within their organization. F.MDR also acknowledges the presence of various metadata management approaches, such as Azure Purview and SAP HANA Cloud. Collibra, on the other hand, stands out due to its enterprise-wide scale and sophisticated data modeling capabilities. The ultimate goal is to deliver a unified front-end experience for consumers while properly handling the underlying complexity of metadata storage and modeling.

Participants emphasize the significance of effective metadata presentation in data management. The primary themes identified include the requirement for simplicity, integration with data consumption processes, increasing awareness through use cases, and selecting appropriate metadata management technologies. By addressing these issues, organizations can improve the usability and utility of metadata, resulting in better data interpretation, quality, and decision-making across the enterprise.

Organization: Old VS New (Participants = 7)

A2.MDR emphasizes the necessity of using logical naming standards and descriptions to make information easier to understand, particularly for business users. This strategy tries to involve businesses in the data environment by providing explicit metadata that allows users to comprehend data flow and operations. E.MDR addresses the desire to integrate metadata close to to the consumption or input of data. In a comparable way, suggesting the use of augmented reality or overlays to provide immediate details about the data being viewed or inserted. These perspectives emphasize the need for user-friendly and contextualized metadata representation.

A3.MDR highlights the use of specific use cases and roadshows to raise metadata awareness and its value inside the company. The goal is to foster a change in mindset and encourage metadata adoption throughout the company by demonstrating the value of metadata management in use cases and increasing data literacy. This method recognizes the importance of metadata in enhancing data literacy and overall data management practices.

B1.MDR, on the other hand, specifies Collibra as a tool for unambiguously capturing metadata and connecting it to architecture models. The emphasis here is on metadata consolidation and

selecting appropriate technologies to pull together metadata from various sources. This stresses the importance of certain metadata management technologies in enabling appropriate metadata organization and display.

Participants highlight the importance of logical and clear metadata presentation, close integration with data consumption or input, awareness, and the use of proper metadata management systems. Organizations may increase data understanding, decision-making, and general data literacy by taking these factors into account.

Data Management: Tools (Participants = 8)

E.MDR highlights Collibra's collaborative character, viewing it as a library where multiple users can collaborate on data. However, issues have been raised about usability and user experience, implying that the tool should be made more accessible. E.MDR also advocates for the incorporation of metadata into data consumption, as well as the use of augmented reality and overlays to provide real-time information on data viewed or entered.

B1.MDR, on the other hand, concentrates on Collibra's role in gathering metadata and linking it to architectural models. To successfully gather metadata, the emphasis is on selecting the right tools. B1.MDR emphasizes the need of metadata aggregation for successful data management by recognizing the significance of selecting the correct tools.

A1.MDR highlights Collibra's status as the leading metadata management technology that is widely utilized in regulated industries. It acts as a hub within businesses, allowing users to search for specific data and learn about data ownership and systems. The necessity for a centralized metadata management system is becoming apparent, which will contribute to improved data governance and organization-wide accessibility.

F.MDR supports several types of metadata representation, such as simple list views and visualizations. While Collibra is considered as a reliable and consistent representation tool, F.MDR prefers more appealing and current capabilities for diagrams. The significance of striking a balance between consistency and complexity in metadata models is highlighted as well so that the average end user can effectively understand and use metadata.

The participants explain how Collibra acts as a metadata management solution. In order to achieve efficient metadata management, they emphasize the need of teamwork, usability, integration with data consumption, and accurate representation. Furthermore, the complexity challenges and the need to balance consistency and accessibility in metadata models are underlined.

View of Metadata Representation

The view of the representation of metadata was asked during the interviews. These are the views provided by the participants.

- A1.MDR states that the representation of metadata involves establishing a centralized hub within the organization, such as Collibra, where anyone can access information about data ownership, systems, and other metadata attributes.
- B1.MDR mentions that the representation of metadata includes capturing and defining business terms, data elements, data quality requirements, and critical data through a dictionary, glossary, and control framework.
- C.MDR states that representing metadata varies depending on the purpose and audience. It can range from SQL queries and dashboards to analytical tools, and customized presentations.
- According to A2.MDR metadata representation aims to provide a logical and understandable structure for data, focusing on logical names and descriptions rather than technical details.
- D.MDR states that the representation of metadata often is done in digital formats like JSON, CSV, or Excel files, which describe tables, data types, and configurations.
- B2.MDR mentions that metadata representation should be simple, intuitive, and valuable, enabling clear interpretation of data and its capabilities.
- E.MDR states that the representation of metadata can be enhanced through augmented reality or overlay features, providing relevant information about data directly where it is consumed or entered.
- A3.MDR mentions that the representation of metadata requires creating awareness, conducting roadshows, and showcasing use cases to demonstrate the importance and value of metadata management.
- F.MDR states that metadata representation exists in various forms and should be preserved to ensure data literacy and support data-driven decision-making within an organization.

4.4 Metadata Representation in a Data Governance context

This section provides the results for the data governance metadata representation related questions. One of the questions was the view of participants against the representation of metadata in a data governance context. Not all experts had the same view. In the next paragraphs the themes found for the section ‘Metadata Representation in DG’ are given. Finally, the view of the participants is given.

First, the codes that were related to the code ‘Section: Metadata Representation in DG’ were filtered. This resulted in a list of 83 different citations. The codes that have been assigned to these citations were grouped in themes. The codes that occurred two times or more have been used to create themes. See Table 7 for the themes with the most important codes.

Table 7 - Themes for 'Metadata Representation in Data Governance'

Theme	Participant ID	Codes
Understanding	A1; B1; B2; C; D; E	• Clarity/Understanding: Confusion • Clarity/Understanding: Uncertainty • Clarity/Understanding: Unclear • Technology: Difficulty • Technology: Technical constraints
Organizational management	A1; A2; B1; B2; C; D; E; F	• Personal and Organizational Development: Accountability • Business Management: Organization governance • Business Management: Risk management • Personal and Organizational Development: Compliance • Business Management: Organizational structure • Business Management: Business processes • Business Management: Workflow processes • Technology: Organization
Data Management and Tooling	A1; A2; A3; B1; C; D; E	• Data Management: Data Management Tools • Data Quality: Data Quality Control • Technology: Tooling • Data Management: Data analysis • Data Management: Data capture • Data Management: Data classification • Data Management: Data visualization • Technology: APIs
Privacy	A3; B2; C; D	• Data Management: Data privacy • Technology: Privacy
Data Management and Quality	A1; A2; A3; B1; C; D; F	• Data Management: Data modeling • Data Management: Data Quality • Technology: Technical requirements • Data Quality: Data quality assurance
Metadata Management	A1; A2; A3; B1; C; D; E; F	• Technology: Metadata • Personal and Organizational Development: Awareness • Business Management: Business adoption • Data Management: Importance of correct data definitions • Data Management: Metadata cataloging • Data Management: Metadata Management

Understanding (Participants = 6)

B1.MDRDG emphasizes the many data management positions inside the firm, such as data stewards and data owners. Despite the fact that B1.MDRDG acknowledges that they are only responsible for data management in their role, B1.MDRDG emphasizes the existence of several functions that contribute to good data management. The question is how these roles interact to carry out their different responsibilities. B1.MDRDG also highlights the shift in data management, acknowledging that it has occurred in stages. To establish what data management was, there was manual labor and collaborative brainstorming at first. Consultants were even considered for providing a data dictionary. However, it is acknowledged that purchasing a solution does not entirely meet the need for company-wide ownership and understanding. The value of time and the ability of individuals to go through the process of taking ownership of data management are both emphasized. Instead of relying entirely on external solutions, the organization must embrace data management.

B2.MDRDG explains the distinction between data governance and data management. B2.MDRDG defines data governance as policies and guidelines that describe how data should be stored and the specific needs required for compliance. On the contrary, data management is defined as the application and enforcement of these policies. This quotation emphasizes the viewpoint that data governance establishes the rules, while data management ensures that the rules are obeyed.

The participants provide a sense of the organization's understanding of data management. The roles and collaboration of data stewards and data owners are considered crucial to efficient data management. The progress of data management, as well as the importance of internal ownership and comprehension, are also emphasized. Finally, the sources distinguish between data governance and data management, emphasizing the importance of policies and their application in efficiently managing data.

Organizational management (Participants = 8)

C.MDRDG considers the difficulties of organizing cross-functional efforts across teams, particularly with regard to data warehouses and production systems. The challenge is keeping track of data management practices across the company. This emphasizes the importance of efficient coordination and collaboration in order to achieve uniform data management methods across teams and systems.

A1.MDRDG defines the data governance team's tasks and stresses the distinction between governance and operational management. The team organizes workshops to bring together data owners for talks on topics such as data definitions. A1.MDRDG is also responsible for defining rules, standards, and frameworks, as well as assuring data ownership and accountability. In addition, A1.MDRDG facilitates training to increase data management principles understanding.

This highlights the functional component of data management and the data governance team's role in its implementation.

B1.MDRDG emphasizes the significance of organizing data governance inside the organization. This includes identifying responsibilities and resolving any potential problems. The aim is to provide a framework that ensures correct data management while also integrating with the organization's processes and protocols. The focus is on establishing and administering data governance in a way that fosters efficient and effective data management methods.

C.MDRDG underlines the need of showing data access and sensitivity. A clear and visible depiction of who has access to sensitive data can help with monitoring and control. This emphasizes the significance of metadata in providing insight into data use, access, and compliance, and so helping to successful data management.

A2.MDRDG emphasizes the importance of an organization having the correct data, processes, and accountable people in place. It is essential for good data management to ensure that the data matches with the needs, operations, and protocols of the organization. A2.MDRDG emphasizes the significance of responsibility and accountability in data management.

D.MDRDG highlights the importance of metadata in data compliance and management. Metadata allows you to specify desired data parameters, allowing for targeted data retrieval and regulatory compliance. D.MDRDG emphasizes the importance of metadata in establishing data compliance and efficient data management techniques.

B2.MDRDG connects data governance to compliance and the obligation to preserve specific metadata components in order to meet the requirements of regulatory organizations such as the Dutch Data Protection Authority. This remark emphasizes the significance of metadata in ensuring regulatory compliance and legal requirements associated with data governance.

F.MDRDG discusses the concept of 'governance' in data governance. The term was originally connected with bureaucracy and compliance but has now developed to include involvement and accountability. It is considered a beneficial feature of efficient data management since it requires individuals to take ownership and responsibility for data. The importance of participation and awareness in governance is underlined, emphasizing the need to involve stakeholders and develop a data management culture inside the company.

The participants shed light on several elements of organizational management within the context of data governance. They discuss cross-functional cooperation problems, data governance team roles and responsibilities, the necessity of metadata for compliance and data access control, and the shifting perception of governance as an active and accountable practice.

Data Management and Tooling (Participants = 7)

A1.MDRDG believes that focusing too much on tooling, where data governance is solely about setting up and linking tools like Collibra, is a potential issue. The essential essence, however, is

in data governance itself, in employing tools as instruments rather than as the center. To effectively assist governance, E.MDRDG highlights the significance of capturing roles, responsibilities, and other critical factors in a tooling system.

A1.MDRDG brings up another problem about monitoring, which can be difficult in data governance. Although data governance analysts can develop norms and standards, data ownership remains with the appropriate business units. It becomes clear that coordination and clear communication between data governance teams and business units are required. C.MDRDG highlights the significance of structured metadata representation, particularly for distinguishing and controlling sensitive information.

B2.MDRDG acknowledges Collibra's significance for showing lineages of data and departmental connections but observes that text-based information predominates. To successfully categorize and represent metadata, A3.MDRDG recommends employing relevant tools and combining them with cross-application and business intelligence solutions.

Participants emphasize the importance of data management and metadata representation tools in data governance. It is critical to achieve a balance between employing tools as resources and focusing on data governance. The effective use of tools and organized techniques improves governance standards, simplifies monitoring, and provides visualizations that aid in the understanding and management of data assets.

Privacy (Participants = 4)

C.MDRDG highlights the importance of customer data sensitivity and the requirement for a structured approach to distinguish and preserve unique information, such as a customer's address or household composition. Clear structures and metadata provide for effective governance to ensure privacy compliance and data management.

C.MDRDG also underlines the importance of showing access rights to sensitive data. A clear visual representation of who has access to what data enables improved management and control of data use, promoting accountability and privacy compliance. B2.MDRDG further highlights the need for specific metadata storage to meet Data Protection Board privacy regulations.

D.MDRDG adds that while end users and administrators may be interested in technical metadata, it plays a significant role in data management. Technical metadata, which may not be sensitive in and of itself, provides valuable information about the data and enables efficient data management. Access to metadata becomes critical when it comes to answering data-related questions, such as finding missing tables or columns and requesting data changes or additions.

In general, the participants emphasize the relevance of privacy issues in metadata representation for data governance. They stress the importance of structured approaches, clear visualizations of access rights, and the inclusion of relevant metadata to ensure compliance,

protect sensitive information, and make data available to the right people at the right time for the right reasons, all while maintaining data security.

Data Management and Quality (Participants = 7)

D.MDRDG emphasizes the importance of managing data flow and maintaining data quality while conforming to relevant regulations. This emphasizes the significance of creating processes and policies to properly manage and ensure data quality.

B1.MDRDG defines the methodical process to capture metadata, beginning with business term definitions in a dictionary or glossary and progressing to data elements and other information such as quality requirements. The availability of key data frameworks, techniques, and regulations emphasizes the necessity of metadata in supporting data governance activities.

A2.MDRDG outlines ongoing work on logical data models and logical quality criteria in the procurement area. The goal is to create a link between logical quality standards and data properties within the model. This highlights the desire to link quality rules at several levels, influencing physical data and encouraging data governance across the data lifecycle.

F.MDRDG offers an example of a smaller-scale implementation of data governance. Even with limited resources, such as a glossary recording critical information about data objects, appointing responsible individuals, and encouraging cooperation, significant progress toward data governance objectives can be made. This demonstrates the significance of metadata in promoting knowledge sharing, control, and order inside a company.

Participants highlight the importance of metadata in data management and quality within the context of data governance. Metadata supports the creation of data definitions, captures critical information about data elements and their related rules, and facilitates the connection of logical and physical data. Organizations may improve data governance procedures, maintain data quality, and efficiently manage their data assets by leveraging metadata.

Metadata Management (Participants = 8)

F.MDRDG mentions a smaller financial institution that established a data board (or data governance board) to stimulate conversations about data-related difficulties and foster information exchange. They began by compiling a dictionary or catalog of data definitions, places, and standards. This modest solution enabled greater data control and organization, demonstrating that metadata is important even in the early phases of data governance.

B1.MDRDG highlights the value of metadata in guiding data quality and delivering precise definitions to intended recipients. Metadata allows for the monitoring, measurement, and assurance of data quality, which is especially important for crucial data. It becomes difficult to efficiently manage and govern data without sufficient metadata.

D.MDRDG states that metadata is important not only for compliance but also for supplying precise datasets tailored to individual requirements. Metadata enables data selection based on predefined criteria, ensuring that the desired data is retrieved and that service level agreements (SLAs) or data license agreements (DLAs) are followed. Metadata is becoming increasingly important for data handling and compliance.

A2.MDRDG emphasizes the need for metadata in efficient data governance implementation. Without metadata, it is difficult to properly apply data governance rules, and knowing the amount or context of the data becomes difficult. Metadata offers the context and information required to assist data governance efforts.

All of the participants highlight the importance of metadata in data governance. Metadata offers effective control, compliance, evaluation of data quality, and data selection based on predefined criteria. Even in the early stages of deployment, it serves as a guiding force for data management and governance.

Definition of Data Governance

The definition of data governance was asked during the interview. These are the definitions provided by the participants.

- A1.MDRDG defines data governance as the operational management of data, including sessions to discuss data topics, definitions, and working on policies, standards, frameworks, ownership, responsibility, and training.
- B1.MDRDG states that data governance involves structuring and organizing how data is dealt with in a company.
- C.MDRDG states that data governance focuses on ensuring data quality, accuracy, reliability, and controlling access to data.
- A2.MDRDG emphasizes that data governance is about how procedures and roles within an organization ultimately affect data, ensuring that the right data is available to keep processes running smoothly.
- D.MDRDG states that data governance is related to managing the flow of data, maintaining data quality, and complying with rules and regulations.
- B2.MDRDG defines data governance as the policy that outlines how data should be stored and the descriptions and requirements needed to comply with certain standards. B2.MDRDG also mentioned that data management is more about adhering to the policy
- E.MDRDG states that data governance is about setting up roles and responsibilities within an organization to ensure data quality, covering all necessary tasks and processes.
- A3.MDRDG defines data governance as making data fit for purpose with the right level of controls, security, availability, and ensuring it is accessible to the right people at the right time.
- F.MDRDG first mentions that data governance had a bureaucratic sense. However, today it is seen as a positive engagement. It involves taking accountability, setting standards, managing data quality, and appointing responsible individuals within the organization.

4.5 Metadata Representation in a Data Quality context

This section provides the results for the data quality metadata representation related questions. One of the questions was the view of participants against the representation of metadata in a data quality context. Not all experts had the same view. In the next paragraphs the themes found for the section 'Metadata Representation in DQ' are given. Finally, the view of the participants is given.

First, the codes that were related to the code 'Section: Metadata Representation in DG' were filtered. This resulted in a list of 88 different citations. The codes that have been assigned to these citations were grouped in themes. The codes that occurred two times or more have been used to create themes. See Table 8 for the themes with the most important codes.

Table 8 - Themes for 'Metadata Representation in Data Quality'

Theme	Participant ID	Codes
Understanding	A1; A3; B1; B2; C; D; E; F	• Clarity/Understanding: Clarifying • Clarity/Understanding: Confusion • Clarity/Understanding: Uncertainty • Clarity/Understanding: Unclear • Business Management: Business glossary • Data Management: Data definition • Personal and Organizational Development: Challenges
Managing quality of data	A1; A2; A3; B1; B2; C; D; E; F	• Data Management: Data management • Data Management: Importance of accurate data • Data Quality: Data quality assurance • Data Quality: Data Quality Control • Business Management: Business processes • Business Management: Organization governance • Data Management: Data governance • Technology: Documentation • Technology: Organization • Data Management: Metadata Management • Technology: Hierarchy • Data Management: Data privacy
Insights from data	A1; A2; A3; B1; B2; C; D; E; F;	• Data Management: Data analysis • Technology: Tooling • Data Management: Data capture • Data Management: Data processing • Data Management: Data visualization • Data Management: Importance of data quality

		• Data Quality: Data quality assessment • Technology: Analysis • Technology: Automation • Technology: Integration • Technology: Representation • Technology: Technology
Data Quality measure	A1; A2 ;A3; B1; C; D; E; F	• Technology: Metadata • Technology: Technical expertise • Data Quality: Validity • Data Quality: Accuracy • Data Quality: Consistency • Data Quality: Reliability

Understanding (Participants = 8)

B1.MDRDQ highlights the importance of metadata in assessing the quality of a dataset, including information about its refresh rate, verification processes, and other important aspects. This understanding serves as the foundation for the next steps in data governance. Without metadata, understanding the nature of the data elements becomes challenging, and B1.MDRDQ advises that a dictionary and glossary are required to establish common agreements and definitions.

C.MDRDQ agrees that metadata is critical to the formulation of rules for data quality assessments. Users can improve their knowledge of the data and evaluate its quality more effectively by utilizing metadata. As A3.MDRDQ emphasizes, understanding data quality involves a business perspective in addition to technical considerations. Organizations frequently place an undue emphasis on technological complexities while ignoring wider business objectives. It is critical to understand the purpose and goals of data governance.

A3.MDRDQ goes on to say that metadata can help uncover potential data quality issues. While metadata is not data in and of itself, it does provide information on the industries, structure, and business rules that govern the data. This knowledge is useful for assessing and improving data quality.

Furthermore, A3.MDRDQ states that metadata, particularly business metadata documented in a glossary of words, aids in the establishment of a single language throughout the firm. It promotes effective communication and a unified picture of data throughout the company. The proper capture and representation of metadata, whether done top-down or bottom-up, helps to achieve this enterprise-wide understanding.

Participants emphasize the importance of metadata in understanding data quality. Metadata helps users to assess and express data quality attributes, aids in the development of quality check criteria, and adds to an organization-wide understanding of data. It facilitates efficient data governance by bridging the gap between technical elements and business objectives.

Managing quality of data (Participants = 9)

To enhance understanding, trust, and correct decision-making, C.MDRDQ highlights the importance of explicit data structures and descriptions, as well as data reliability. B2.MDRDQ highlights the need for reliable data and excellent data quality, as low-quality data can lead to uncertainty and incorrect analysis.

A1.MDRDQ emphasizes the relevance of metadata in capturing and portraying data quality; A1.MDRDQ describes the usage of Collibra for capturing metadata at the business and logical levels, as well as the application of data quality rules at the physical level. A2.MDRDQ emphasizes the need of linking data quality to metadata and understanding data reliability. A1.MDRDQ also suggests breaking down data quality into concepts and entities for greater comprehension and trust.

E.MDRDQ highlights the link between data quality and business processes, emphasizing that data quality should correspond with the needs of diverse processes and not be limited to specific silos. F.MDRDQ also emphasizes the significance of examining data quality aspects such as accuracy, currency, completeness, and integrity, as well as choosing key data pieces for quality management. F.MDRDQ also emphasizes the difficulties of data quality management, such as the difficulty of rectifying and preventing errors caused by hidden logic and authorization issues.

Participants also emphasize the importance of continuous improvement in data quality processes. D.MDRDQ emphasizes the necessity of creating processes for maintaining data quality and addressing concerns such as completeness. A3.MDRDQ highlights the importance of enterprise-wide standards and business-rule-based monitoring to provide a uniform language and consistency in data quality. D.MDRDQ also cites the transition from early data pipeline construction to an emphasis on quality checks and automation.

Finally, the participants emphasize the importance of data quality management, the relevance of metadata, the relationship between data quality and business processes, and the need for continuous improvement. These insights provide helpful perspectives on the problems and issues involved in maintaining and improving data quality.

Insights from data (Participants = 9)

A1.MDRDQ highlights the difficulty of displaying data quality as a single number and proposes a more nuanced method by segmenting it into various data concepts and entities. Examining data quality on a per-entity level, such as leasing contracts or consumers, can provide a better understanding of the data's trustworthiness and accuracy. This method enables a more detailed evaluation, such as concluding that Brazil's contract data is 75% correct.

B1.MDRDQ acknowledges the availability of tools and frameworks inside the organization but emphasizes the lack of a centralized mechanism for data quality evaluation. The goal is to combine these tools into a centralized platform. The current situation includes the use of many instruments, but the goal is to finally bring everything together. This unification would provide a consistent perspective of data quality and streamline the monitoring process.

A3.MDRDQ highlights the broad nature of data quality, noting that it must be "fit for purpose" and include enterprise-wide standards. Data quality is critical for developing a shared language and understanding across all business divisions. It underlines the significance of matching data quality with business standards and putting in place efficient monitoring methods. It is possible to create a consistent approach to data quality by converting business rules into data monitoring methods.

A3.MDRDQ emphasizes the significance of metadata in building a shared language. A thorough vocabulary of words promotes shared understanding and harmonization within the company. The manner of gathering metadata, whether top-down or bottom-up, has an impact on the capacity to achieve a consistent, enterprise-wide view of data.

The citations also discuss the methods used to analyze data quality and visualize it. A1.MDRDQ suggests using Collibra to display data quality test results. B1.MDRDQ, on the other hand, claims that Power BI is currently utilized to create dashboards and acquire insights from data monitoring. These technologies help enterprises to effectively monitor, assess, and analyze data quality.

D.MDRDQ highlights the subjective sense of data usefulness and the link between good data quality and gaining more value from it. Data of poor quality is frequently made useless and serves no useful function. It recognizes the difficulty of dealing with low-quality data; as the phrase goes, "garbage in, garbage out." Nonetheless, organizations may maximize the value of their data and unlock potential insights by ensuring it is in the proper format and adhering to high data quality standards.

A3.MDRDQ mentions data quality visualization through lineage and visualization methods. These visual representations show users exactly where data quality is being tested and provide a clear representation of the accompanying scores. Based on data quality assessments, such visualizations improve understanding and promote informed decision-making.

Finally, A3.MDRDQ highlights the importance of alert systems that alert data stewards or owners when data quality goes below a predefined threshold. This ensures that actions may be taken to remedy any errors and maintain report integrity. This information should ideally also flow up to the reporting line, giving report owners confidence in the certification of their reports based on the observed data quality score.

Participants emphasize the significance of metadata representation in data quality, as well as its function in building a common understanding, trust, and trustworthiness in organizational data.

They emphasize the importance of centralized tools, established processes, and visualizations in monitoring and assessing data quality. Organizations can maintain the integrity and usefulness of their data by prioritizing data quality and establishing alarm systems, resulting in improved decision-making capabilities and commercial outcomes.

Data Quality measure (Participants = 8)

F.MDRDQ emphasizes the significance of capturing metadata representation in a way that corresponds to end-user familiarity and preferred system interfaces. The objective is to easily integrate metadata tools into current programs. The objective is to deliver a uniform user experience and reduce the need for individuals to navigate numerous systems, resulting in increased productivity and user acceptance.

A2.MDRDQ highlights the interdependence between metadata and data quality measurement. The intended data quality criteria and norms are defined using logical reasoning and then translated onto the actual data using metadata representation. The dependability and accuracy of the data can be checked and validated by associating metadata models with data quality tests. This link emphasizes the relevance of metadata in identifying the origins and dependability of data.

According to A2.MDRDQ, the concept of data quality includes both completeness and correctness. It is insufficient for data to be correct if it is incomplete. Although a table with only 10% of its data filled in is technically correct, it lacks relevant information. Both aspects influence the amount of reliability, and different characteristics contribute to data quality, such as adherence to stated norms and suitable formatting. To ensure accuracy, data quality measurements should ideally be performed as close to the data source as possible.

The participants emphasize the importance of metadata representation in measuring data quality. The alignment of metadata tools with user interfaces and existing systems is critical for user adoption and efficiency. Furthermore, combining metadata models with data quality measurement allows for the assessment and verification of data reliability and accuracy. Organizations may improve their understanding of data quality and make educated decisions based on trustworthy and full data by considering both completeness and accuracy.

Definition of Data Quality

The definition of Data Quality was asked during the interview. These are the definitions provided by the participants:

- A1.MDRDQ mentions that data quality is ensuring data is fit for its purpose. They use seven quality dimensions and have a separate team dedicated to data quality.

- B1.MDRDQ states that data quality is based on agreements between users and data suppliers, taking into account various data quality dimensions.
- C.MDRDQ emphasizes the importance of clear data structure and modeling, along with clear field descriptions. C.MDRDQ also mentions the reliability of the data itself, ensuring that it is accurate and trustworthy.
- A2.MDRDQ states that data quality involves both completeness and accuracy. It is essential that data meets expectations and is correct, with proper reliability and adherence to data standards.
- D.MDRDQ highlights the usefulness of data as a measure of its quality. High data quality enables extracting more value from the data, while low-quality data is unusable. The source of data also plays a role in data quality.
- B2.MDRDQ considers data quality in terms of compliance with requirements and dimensions such as availability, integrity, accuracy, and completeness. B2.MDRDQ emphasizes the usability and reliability of data.
- E.MDRDQ states that data quality should be linked to the business processes that requires the data. It extends beyond individual silos and involves ensuring the accuracy and correctness of data across related processes and consumer-producer relationships.
- A3.MDRDQ defines data quality as establishing common standards across the enterprise to ensure a shared understanding. Fit for purpose is a crucial aspect, and data quality should be driven by business rules and monitored accordingly.
- F.MDRDQ defines data quality as ensuring that data meets self-imposed standards. It involves focusing on critical data elements and applying control based on various dimensions. Consideration is given to existing process and technology layers.

4.6 Key points of results

The key points of the results are outlined in this section.

Metadata:

- Participants recognized the value of metadata in data management and data quality.
- There were calls for structured approaches and guidelines for effective metadata management.
- Awareness and understanding of metadata's importance among participants were highlighted.
- Clear definitions and bridging the gap between technical and business perspectives were identified as challenges.
- Effective metadata management was crucial for data governance, transparency, compliance, and addressing pain points.
- Business and technical metadata were distinguished, emphasizing their respective roles.

Metadata Representation

- Metadata representation should be user-friendly, intuitive, and accessible to a wide audience.
- Metadata management tools like Collibra were highlighted for their usability and enterprise-wide scope.

Data Governance:

- Data governance involves the operational management and organization of data within a company.
- Roles, ownership, policies, standards, and responsibility are emphasized.
- Data governance ensures data quality, accuracy, and reliability and controls access to data.
- Data governance supports management, compliance, decision-making, and trust in complex data environments.
- Metadata is critical in data governance, enabling control, data quality measurement, and compliance.

Data Quality:

- Data quality is defined as ensuring data is fit for its intended purpose.
- Dimensions such as completeness, accuracy, reliability, and compliance contribute to data quality.
- A clear data structure, modeling, and reliable data are essential for quality.
- Data quality involves establishing common standards, adhering to data standards, and aligning with business rules.
- Data usefulness and value are closely tied to high data quality.
- Tools for data quality assessment and visualization, such as Collibra and Power BI, were mentioned.
- Metadata is integral to understanding, managing, measuring, and improving data quality.

5. Discussion

This chapter will discuss the results that have been provided in chapter four. Then, the definitions of the participants are interpreted and compared to the definitions used in this research. Finally, the implications, relevance, and limitations of this study are outlined.

5.1 Data Structure

First, the data structure (Figure 2) has been created from the results. This data structure consists of first-order concepts, second order themes, and dimensions. The results sections' themes have been aggregated to create first-order concepts. Each theme has been assigned to one another. This resulted in five high-over categories. These are (i) Understanding, (ii) (Meta)Data management, (iii) Organization, (iv) Data management tools, and (v) Privacy. See Appendix VIII: Categories and themes to data structure, Table 14, 15, 16, 17, and 18 for the aggregated themes and codes.

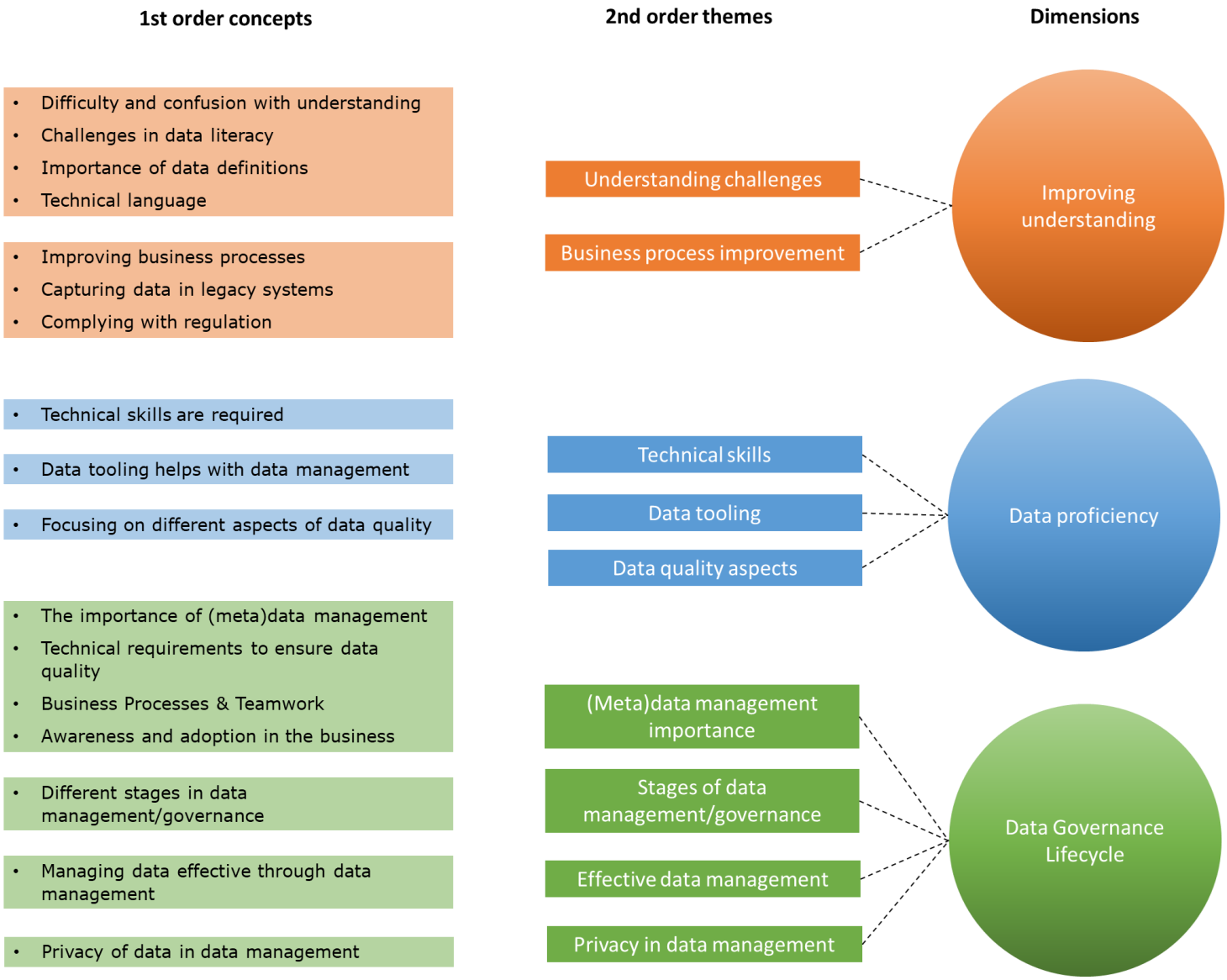

Figure 2 - Data Structure

Secondly, the unique codes per category have been deduplicated. The distinct codes per category were grouped, and the first-order concepts were created from these groups. See Figure 2Table 19 for the first-order concepts.

Third, the first-order concepts were combined and from this combination the second-order themes were created. See Appendix VIII: Categories and themes to data structure, Table 19 for the combined first-order concepts and second-order themes.

Finally, the second-order themes have been aggregated into three dimensions: (i) Improving understanding; (ii) Data proficiency; and (iii) Data Governance Lifecycle. See Figure 2Table 19 for the dimensions.

The paragraphs below will elaborate on the themes in the dimensions. The participants will be mentioned according to their ID and section code. At the end of each theme, informal propositions are given based on the overall theme metadata representation. The propositions will be used for future research recommendations in chapter 6.

5.2 Dimension I: Improving understanding

This dimension comes from the two themes: (i) Understanding challenges and (ii) Business process improvement. The themes will be analyzed and discussed in-depth.

Understanding challenges

Understanding challenges is the first theme, which has led to the first dimension Improving understanding. In the sections below, the results will be analyzed that were obtained during the interviews in relation to understanding the challenges. The findings will be compared with the literature which has been discussed in chapter 2 Theoretical Background. On the basis of this comparison and the theme, a conclusion will be drawn, and a proposition will be proposed.

One of the most significant challenges revealed by the participants' insights is the difficulty and confusion encountered when grasping metadata and its management (A3.MD, E.MD, B1.MDRDQ). This highlights the need for organizations to address this issue by providing clarity, guidelines, and standards that can assist individuals in efficiently navigating the complicated world of metadata. This challenge of understanding metadata has also been found in the literature review of Ulricht et al. (2022). They mention the human-interaction problem of understanding the metadata. Another problem they mention is related to structure, which means the lack of standard usage. Multiple studies highlight the importance of standardized metadata and its representation (Dai et al., 2021; Mandal et al., 2016; Melo et al., 2021).

Another challenge is the difficulties involved with data literacy. Participants emphasized the need to make metadata clear, accessible, and understandable to a broad audience (E.MD, A2.MDR, E3.MDR, F.MDR). In metadata management and data governance, technical jargon presents a barrier (A2.MDR). Complex vocabulary makes it difficult for business users and non-technical

stakeholders to understand metadata and its implications (A3.MD, E.MD, A3.MDRDQ). Organizations need to foster data literacy efforts to overcome these issues that enable individuals to understand and use metadata properly. The above concludes that organizations must educate employees on various data concepts. In other words, they need to increase data literacy within the organization. Only when everyone, or everyone involved, understands the intended meaning of specific data and terms, data can be used to its fullest potential. Like Ackhoff's (1989) pyramid, employees must understand the intended meaning of specific objects and terms. A data glossary and a data dictionary are good examples to help understanding the concepts. Understanding the context gets us to the information we seek. Ultimately, by connecting the information, the right decisions can be made (Ackoff, 1989; Gartner, 2016).

Participants highlight the significance of data definitions (E.MD, C.MD, A1.MDRDG, B1.MDRDG, F.MDRDG, B1.MDRDQ). Users can understand the nature and characteristics of the underlying data when data definitions are clear and well-defined. Using metadata dictionaries and glossaries to establish common agreements and definitions can help overcome the difficulty of correctly interpreting metadata and increase data understanding. Metadata dictionaries and glossaries do not only help with interpreting and the understanding of metadata, but DAMA International (2017) mentions that they also help with not losing the knowledge that is in the organization.

The literature mentions another human challenge. Because of the limitations of the human brain, the brain needs help remembering things (Gartner, 2016; Mayer, 2003). This is where metadata comes into the picture. As the results and literature indicate, metadata should be presented in a clear, accessible, and understandable way so that non-technical individuals and business users can understand them. Representation Theory (RT) can be used to help present the metadata and make it understandable. RT has four assumptions on how information is communicated:

(i) RT can communicate meaning through symbols, and humans can obtain meaning from these symbols (Wand & Weber, 1995).
(ii) Information systems intend to represent someone’s or a group's view of the real world (Wand & Weber, 1995). Wand & Weber (1993) and Burton-Jones & Grange (2013) reason that computerized representations provide a more efficient way to learn about the world than observations. This leads to the design of information systems.
(iii) Stakeholders can express their meaning about events that are of interest to them (Burton-Jones et al., 2017).
(iv) Users expect that an information system has more value when it represents a true view of the real world (Weber, 1997).

When examining these assumptions in the light of the literature and results, the following can be concluded:

- The first assumption is that communication of meaning through symbols is a simple and understandable way to extract meaning. This can be used to present (meta)data to non-technical and business users.
- Secondly, it can be seen in the second assumption that information systems intend to represent someone's or a group's view of the real world. The participants mentioned that this should be done clearly, easily, and friendly.

In conclusion, the participants and literature emphasize the problems and ambiguity of metadata management and data governance. They highlight the difficulties of data literacy, the need for data definitions, and the need to avoid technical jargon. By addressing these issues, businesses can improve understanding, increase data literacy, create clear data definitions, and effectively represent metadata to allow efficient metadata management and data governance procedures. Thus, the suggested proposition is:

Proposition 1: Understanding the challenges has a positive effect on the representation of metadata.

Business process improvement

Business process improvement is the second theme, which has led to the first dimension Improving understanding. In the sections below, the results will be analyzed that were obtained during the interviews in relation to business process improvement. The findings will be compared with the literature which has been discussed in chapter 2 Theoretical Background. On the basis of this comparison and the theme, a conclusion will be drawn, and a proposition will be proposed.

From the results, C.MDRG highlights the emphasis on coordination and collaboration across teams and systems, which underscores the importance of streamlining and optimizing business processes. Effective coordination ensures consistency in data management processes throughout the organization, eliminating redundancies, reducing errors, and enhancing overall efficiency. Organizations can establish clear responsibilities and standards by establishing effective data governance frameworks, resulting in easier data flows and better decision-making. It has been indicated by Aamot (2022), IEEE (2020), and Yan et al. (2022) that standards for metadata contribute to a standardized framework for organizing and managing metadata, thereby improving operability and consistency across systems and domains.

The data governance team, as mentioned by A1.MDRDG, plays an important role in business process improvement. The data governance team helps the adoption of data management concepts and practices by arranging workshops, developing rules and standards, and assuring data ownership and accountability. These actions foster a data-driven culture within the organization, with business processes aligned with data requirements and regulations (B1.MDRDG). Furthermore, training initiatives can help employees better comprehend data

management concepts, allowing them to make more informed decisions and contribute to process development.

Some participants have noted the incorporation of metadata into corporate operations provides useful insights about data access, sensitivity, compliance, and legal needs (A2.MDRDG, B2.MDRDG, C.MDRDG,, D.MDRDG). Organizations may effectively monitor and control data usage by exploiting metadata, guaranteeing compliance with relevant rules and securing sensitive information. This level of visibility and control allows firms to make better informed decisions, enhance operations, and eliminate data management risks. Looking into the definition by DAMA International (2017) *"Metadata includes information about technology and business processes, data rules and constraints, and logical and physical data structures. It describes the data itself (e.g., databases, data elements, data models), the concepts the data represents (e.g., business processes, application systems, software code, technology infrastructure), and the connections (relationships) between the data and concepts"* (p. 417). It can be read that the results from the participants align with the definition of DAMA. Thus, metadata includes information about the processes and rules. DAMA also indicates that metadata helps distinguish between sensitive and private data to comply with laws and regulations. It is also indicated that high-quality metadata contributes to decision-making (Dyson & Foster, 1982; Harley & Cooper, 2021; Shankaranarayanan et al., 2006, 2008; Shankaranarayanan & Zhu, 2021; Stvilia et al., 2007).

Participants also underline the necessity of having the right data, processes, and accountable people in place (A2.MDRDG, C.MDRDG). Aligning business processes with the organization's specific needs, operations, and standards is critical for successful data management. Organizations may streamline processes, decrease inefficiencies, and improve overall performance by ensuring that data is fit for purpose and aligned with business requirements. The literature states that the business and its processes create the conceptual layer within the database domain. There are few technical details stated in this, it is rather the high-level view about assets that need to be connected (Garcia-Molina et al., 2008; Groves, 2022; Ramakrishnan et al., 2003).

Finally, the concept of governance as an active and accountable practice is critical in promoting business process improvement (F.MDRDG). Encouraging stakeholder participation and understanding develops a culture of data accountability and ownership. Individuals within an organization are more likely to actively contribute to process improvement activities that drive innovation and efficiency when they understand the importance and influence of data management on business processes.

In conclusion, the interviews indicated the significance of business process improvement, with a focus on coordination, cooperation, and data governance. Streamlining procedures, defining roles, and using metadata all lead to greater knowledge and consistency. Aligning business processes with organizational needs and cultivating an accountability culture promotes creativity

and efficiency. Businesses can streamline operations, improve decision-making, and promote overall progress by addressing these difficulties. Thus, the suggested proposition is:

Proposition 2: Incorporating metadata positively affects business processes.

5.3 Dimension II: Data proficiency

This dimension comes from the three themes: (i) Technical skills; (ii) Data tooling; and (iii) Data quality aspects. The themes will be analyzed and discussed in-depth.

Technical skills

Technical skills are the third theme, which has led to the second-dimension Data proficiency. In the sections below, the results will be analyzed that were obtained during the interviews in relation to technical skills. The findings will be compared with the literature which has been discussed in chapter 2 Theoretical Background. On the basis of this comparison and the theme, a conclusion will be drawn, and a proposition will be proposed.

Technical skills are essential for utilizing tooling such as Collibra, which is highlighted by participants as an effective metadata management system and supporting data governance tasks (E.MDR, B1.MDR, A1.MDR, A1.MDRDQ). Participants frequently highlight Collibra's collaborative aspect, which allows numerous users to interact on data and promotes teamwork in metadata management (E.MDR, B2.MDRDG, A3.MDRDG). However, fully utilizing Collibra's (and other tools) functionality and encouraging communication among data governance teams involves the use of technical expertise. Participants also underline the significance of technical skills in properly applying metadata representation (E.MDR, F.MDR). To ensure that end users can understand and use metadata successfully, metadata models must strike a balance between consistency and complexity (F.MDR). This includes combining visually appealing elements while retaining consistency and readability for the average end user. The results found that technical skills are needed to integrate metadata efficiently into the business processes and to represent metadata effectively. As mentioned in the definition of Big Data by Zikopoulos and Eaton (2011): *"information that cannot be processed or analyzed using traditional processes or tools"* (p.3); Big Data cannot be processed or analyzed using traditional tools. Technical skills are required to analyze Big Data.

Furthermore, technical skills are required in the context of measuring data quality. Participants emphasized the importance of representing metadata in a form that corresponds to end-user familiarity and preferred system interfaces (B1.MDRDQ, A3.MDRDQ). This requires technical expertise to easily incorporate metadata tools into existing applications, minimizing the need for individuals to navigate different systems, and increasing productivity. Technical skills are not only needed when incorporating metadata but also needed when storing data, specifically designing databases. As mentioned in chapter 2 Theoretical Background there are three layers

in database design. The most technical layer is the physical layer. Besides defining the data attributes, this layer describes the type of database used and the database's underlying infrastructure(Garcia-Molina et al., 2008; Groves, 2022; Ramakrishnan et al., 2003). When designing a database, there should be multiple aspects considered. A technical skilled person can help with the technical and preferably business side of designing.

The ability to develop data quality criteria and measures using logical reasoning and transfer them into actual information using metadata representation demands technical expertise (A3.MDRDQ). Thus, technical skills are also required when combining metadata models with data quality measurement. This enables evaluating and confirming data reliability, accuracy, and key data quality components. What is also found is the aim to strike a balance between consistency and complexity in metadata models, incorporating visually appealing and current visualization features while ensuring that the average end user can understand and utilize metadata effectively. The Representation Theory by Wand & Weber (1990) focuses on incorporating user perceptions of meaning into an information system (Wand & Weber, 1995). RT could be useful when designing metadata models for information systems.

In conclusion, technical skills are critical in many aspects of data management, including the skilled use of data management tools, the effective implementation of metadata representation, and the accurate data quality assessment. Organizations should emphasize growing technical capabilities within their data governance teams to ensure optimal use of solutions like Collibra, seamless integration of metadata models, and accurate data quality assessment. Organizations may strengthen their data governance procedures, decision-making processes, and the value gained from their data assets by fostering technical knowledge. Thus, the suggested proposition is:

Proposition 3: Technical skilled employees have a positive effect on the representation of metadata.

Data tooling

Analysis of the results

Data tooling is the second theme, which has led to the second-dimension Data proficiency. In the sections below, the results will be analyzed that were obtained during the interviews in relation to data tooling. The findings will be compared with the literature which has been discussed in chapter 2 Theoretical Background. On the basis of this comparison and the theme, a conclusion will be drawn, and a proposition will be proposed.

Data tools are essential in data management, (meta)data representation, and data quality management. Participants emphasize the importance of choosing the correct tools for information aggregation and highlight Collibra's role in gathering metadata and relating it to

architectural models (B1.MDR, B2.MDRDQ, A1.MDRDQ). The tooling system should record roles, responsibilities, and essential aspects to support governance properly (A1.MDRDG, E.MDRDG). Participants also emphasized the importance of collaboration and teamwork and the requirement for a centralized metadata management system that promotes data governance and accessibility throughout the company (A1.MDRDG, E.MDRDG, B2.MDRDG, A3.MDRDG). As mentioned above, data tools are important to implement in the organization. This is also what has been happening in recent years. Bansal et al. (2021) highlight that organizations have transformed their way of conducting business through data-driven intelligence. This can only be achieved through the implementation and utilization of tooling.

Collibra is a well-known solution for metadata management, particularly in regulated industries. It is a centralized hub for searching specific data, determining ownership, and understanding system designs (A1.MDR). However, some complaints about Collibra's usability and user experience have been made, implying that it needs to be more accessible (E.MDR, F.MDR). To enable successful understanding and utilization by end users, the participants emphasize the need to balance consistency and complexity in metadata models. Integrating metadata into data consumption processes using augmented reality and overlays is also addressed. When looking at the design of databases where data can be stored, the three modeling layers are important. Modeling database design can be done with pen and paper but is more efficient digitally. In some applications[6], modeling can be done and implemented instantly. These digital tools are needed to create an understanding of data and its relationships. The Representation Theory by Wand & Weber (1990) can help represent the users' view of the objects and relations needed in the database.

Explicit data structures, descriptions, and reliable data are critical in data quality management (C.MDRDQ, B2.MDRDQ). Collibra is suggested for recording data quality metadata, allowing organizations to assess accuracy, currency, completeness, and integrity (A1.MDRDQ). Integrating data quality with business processes is highlighted, emphasizing data quality aligning with different needs rather than being constrained to silos (E.MDRDQ, A3.MDRDQ). Continuous improvement, enterprise-wide standards, and automated monitoring mechanisms are emphasized to ensure data quality and usefulness (A1MDRDG, A3.MDRDQ, D.MDRDQ). Tooling is needed when (meta)data needs to be stored and shared. As mentioned earlier, data can be stored in databases. APIs can then access this data (Pomerantz, 2015; Riley, 2017). Data can be shared in various ways, from visualization to implementation in a language specifically tasked to transfer metadata. Visualization can be done through modeling using the Resource Description Framework, which models data as a network (Forum Standaardisatie, 2013; Riley, 2017). Metadata can be shared through languages like XML or JSON (Riley, 2017).

[6] Applications such as MySQL Workbench, HeidiSQL, and DataGrip among many others.

In conclusion, data tools like Collibra are critical in data management, metadata representation, and data quality management. While Collibra is considered a good metadata management tool, usability and user experience might be improved. Integrating metadata into data consumption processes and using the appropriate metadata aggregation technologies are critical for efficient data management. Furthermore, specific data structures, descriptions, and accurate data help to preserve data quality. Continuous improvement, enterprise-wide standards, and automated monitoring systems are critical for guaranteeing the utility and value of data. Organizations may improve their data management processes, decision-making, and business outcomes by adopting data tooling. Thus, the suggested proposition is:

Proposition 4: Implementing specific data tools has a positive effect on the representation of metadata.

Data quality aspects

Data quality aspects is the third theme, which has led to the second-dimension data proficiency. In the sections below, the results will be analyzed that were obtained during the interviews in relation to data quality aspects. The findings will be compared with the literature which has been discussed in chapter 2 Theoretical Background. On the basis of this comparison and the theme, a conclusion will be drawn, and a proposition will be proposed.

From the results, it was obtained that there is a strong connection between metadata and data quality. The participants highlight the use of metadata in creating data quality rules and norms (A1.MDRDG, A2.MDRDQ). Organizations can examine and certify the reliability and accuracy of their data by associating metadata models with data quality tests. This link emphasizes the significance of metadata in identifying the origins of data and assuring its reliability. The results mention that metadata improves data quality. This, in turn, contributes to decision outcomes and decision-making accuracy and reliability (Dyson & Foster, 1982; Harley & Cooper, 2021; Price & Shanks, 2011; Shankaranarayanan et al., 2006, 2008; Shankaranarayanan & Zhu, 2021; Stvilia et al., 2007). Several studies have stressed the significance of metadata in ensuring accurate, consistent, and reliable data (Dion, 2007; Myrseth et al., 2011; Verbitskiy & Yeoh, 2011). These studies all highlight the importance of metadata in high data quality. As with all improvements, there are challenges in guaranteeing (meta)data accuracy, consistency, and completeness and dealing with (meta)data integration and interoperability issues. Implementing effective (meta)data management strategies will help to solve these issues (IEEE, 2020; Loshin, 2015; Sundarraj & Rajkamal, 2019; van Helvoirt & Weigand, 2015).

Furthermore, the results underline the importance of data quality in completeness and correctness (A2.MDRDQ). It is highlighted that data can be technically correct but incomplete, resulting in a lack of meaningful information. Data quality measurements should ideally be performed as close to the data source as possible to ensure accuracy (A2.MDRDQ). This

realization emphasizes the need to consider data quality dimensions - completeness and correctness - when assessing data reliability and trustworthiness. Wand & Wang (1996) provided in their literature review that the most common dimensions for data quality were (i) accuracy, (ii) reliability/consistency, (iii) timeliness (currency), and (iv) completeness. The dimensions obtained from the results have overlapped with the dimensions found in the literature, namely completeness and correctness (accuracy). This implicates the importance of measuring data quality through its dimensions.

The results also include challenges and changes in dealing with various (meta)data types. The importance of rebuilding data models when definitions change is highlighted, for example, when changing terms like 'orders' to 'delivery' (C.MD). This requirement emphasizes the dynamic nature of metadata and the need to appropriately modify data models to represent an organization's growing concepts. As mentioned in chapter 2.2 Metadata, there are three types of metadata (DAMA International, 2017; Gartner, 2016; NISO, 2017). Descriptive metadata is the type of metadata that describes an object.

Furthermore, the growing importance of data security and privacy is stressed, particularly in light of GDPR rules (C.MD). C.MD mentions data lineage and track changes over time to support compliance. Organizations are becoming more careful and vigilant in controlling access to personal information and sensitive data, putting ethical and privacy concerns first. In chapter 2.2 Metadata, DAMA International (2017) states that metadata is useful for capturing knowledge but also to help divide data into private and sensitive information. Because metadata is known as “data about data” it can help with identifying different types and changes in data.

In conclusion, there is a strong relationship between metadata and data quality, with metadata playing a critical role in creating data quality criteria, identifying data origins, and assuring reliability for better decision-making. Despite accuracy and integration concerns, effective metadata management solutions can address these issues. Data quality measurements close to the data source improve accuracy, while modifying data models to changing (meta)data types reflects metadata's dynamic character. Furthermore, the growing relevance of data security and privacy, particularly in relation to GDPR compliance, underlines the role of metadata in tracking data lineage and changes for data protection and compliance. All of the above contribute to measuring data quality, which contributes to high data quality. Thus, the suggested proposition is:

Proposition 5: High metadata quality has a positive effect on the representation of metadata.

5.4 Dimension III: Data Governance Lifecycle

This dimension comes from the four themes: (i) Data management importance; (ii) Stages of data management/governance; (iii) Data quality aspects; and (iv) Privacy in data management. The themes will be analyzed and discussed in-depth.

Data management importance

Data management importance is the first theme, which has led to the third dimension Data Governance Lifecycle. In the sections below, the results will be analyzed that were obtained during the interviews in relation to data management importance. The findings will be compared with the literature which has been discussed in chapter 2 Theoretical Background. On the basis of this comparison and the theme, a conclusion will be drawn, and a proposition will be proposed.

Data management, especially efficient metadata management, is critical for businesses to utilize the full potential of their data assets. The findings emphasize the importance of implementing data management methods that maintain data quality, facilitate data understanding, support compliance requirements, and improve decision-making processes (A2.MD, A3.MD, A3.MDR).

Ensuring data quality is based not only on technological needs but also on well-defined business processes and successful teamwork. The participants mentioned the significance of having explicit business processes and allocating roles and responsibilities for data governance (D.MDRDG, A2.MDRDG). Collaborative efforts, such as data governance boards, may start discussions about data-related issues and encourage information exchange, ultimately leading to more effective data management procedures (F.MDRDG).

It is critical for successful data management to raise awareness and drive acceptance of metadata management techniques across businesses. Participants emphasized enhancing metadata awareness and its role in supporting data quality, analytics, and decision-making processes (A2.MD, A3.MDR). Organizations may build a culture of data literacy and encourage the adoption of metadata management techniques across the business by leveraging specific use cases to demonstrate the value of metadata management (A3.MDR).

Understanding the importance of metadata is critical for good data governance and management (A2.MD, A3.MD, A3.MDR). Metadata offers crucial context, definitions, and rules for data assets, allowing organizations to define, maintain quality, and control data across its lifecycle (E.MDR, A2.MDRDG). Organizations may secure their data assets' correctness, dependability, and integrity by including metadata in data governance activities, facilitating data-driven decision-making and regulatory compliance.

What can be derived from the results is that the fields of data management and data governance are deeply connected. Abraham et al. (2019) specifies data governance as *"a cross-functional framework for managing data as a strategic enterprise asset"* (p.425). They elaborate that data governance has multiple components. It specifies decision rights and accountabilities based on data, the formalization of policies, standards, and procedures, and monitors compliance. These components have to be managed and executed. This is done through a data management program. DAMA International (2017) states that data management is *"the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles."* (p.17)

Data governance is about who makes and what decisions have to be made; management is more about what is needed to bring the decisions into business (Dyché & Levy, 2006; Hagmann, 2013; Khatri & Brown, 2010; Otto, 2013). Both definitions overlap with the findings. The roles and accountability, policies, and procedures in data governance. As well as the supervision and execution of these policies, procedures to enhance the value of data.

In conclusion, effective metadata management is critical for businesses to maximize the potential of their data assets. The findings underline the importance of well-defined business processes and collaborative data governance activities in achieving data quality, knowledge, compliance support, and enhanced decision-making. Organizations can foster a culture of data literacy and increase acceptance of metadata management approaches by raising knowledge of metadata and its role in data quality, analytics, and decision-making. Understanding the significance of metadata is critical to preserving data integrity, control, and reliability, allowing for data-driven decision-making and ensuring regulatory compliance. Thus, the suggested proposition is:

Proposition 6: Managing metadata has a positive effect on the representation of metadata.

Stages of data management/governance

Stages of data management/governance is the second theme, which has led to the third dimension Data Governance Lifecycle. In the sections below, the results will be analyzed that were obtained during the interviews in relation to stages of data management/governance. The findings will be compared with the literature which has been discussed in chapter 2 Theoretical Background. On the basis of this comparison and the theme, a conclusion will be drawn, and a proposition will be proposed.

The participants highlight multiple aspects of data management/governance. The stages are not explicitly mentioned, however, after analysis of the results the stages below have been derived.

The first stage is becoming aware of and understanding metadata and data types. Initially seen as simple labels and descriptions, metadata has evolved into a discipline requiring the engagement of a Chief Data Officer (E.MD). Participants also mention the distinction between technical and business metadata, emphasizing the differences in their objectives and applications (B1.MD, F.MD). The participants define technical metadata as storage and information systems, whereas business metadata concerns data meaning, quality, and governance (B1.MD, F.MD, E.MDRDQ). Similarly, understanding data types has developed to include their roles in data quality, GDPR compliance, data retention, and privacy. This stage involves knowledge about metadata and data types of various formats and application areas. The participants highlight the distinction between business metadata and technical metadata. The literature provided a distinction between three types of metadata: descriptive, administrative, and structural (DAMA International, 2017; Gartner, 2016; Riley, 2017). Descriptive metadata can be placed in the

category of business metadata because this type of metadata is not technical but more expressive. At the same time, administrative and structural metadata are both technical. They provide details about the storage, access, and retrieval of metadata and the relations between data points (DAMA International, 2017; Gartner, 2016; Riley, 2017).

The following stage involves metadata maintenance and the application of tools. Participants address the importance of metadata management solutions like Collibra, which acts as a centralized hub for metadata (E.MD, B1.MDR, A1.MDR, F.MDR, A1.MDRDG, B2.MDRDG, A1.MDRDQ). These tools improve user cooperation, allow data searching, provide insights into data ownership and systems, and deliver real-time information. However, difficulties with usability, user experience, and effective metadata representation are also highlighted. This stage focuses on implementing metadata management strategies and technologies to provide effective metadata governance, control, and accessibility. Data tooling has been elaborated on in the theme data tooling (p. 75).

Data quality management is considered to be an important stage in data governance. Participants underline the significance of data quality in building trust, improving understanding, and making sound decisions (B2.MDRDQ, A2.MDRDQ,A3.MDRDQ, D.MDRDQ). They emphasize the importance of detailed structures, descriptions, and accurate data. Metadata is essential for capturing and displaying data quality since it allows for assessing and validating data reliability and accuracy. Capturing and expressing data quality using metadata, integrating data quality into business operations, defining data quality standards, continual improvement, and visualizing data quality using lineage and visualization approaches are all aspects of data quality management. The participants highlight that data quality is needed in building trust, improving understanding, and making sound decisions. The literature supports these statements. Several studies have highlighted the importance of metadata in data quality, ensuring that data is accurate, consistent, and reliable (Dion, 2007; Myrseth et al., 2011; Verbitskiy & Yeoh, 2011). Other studies have found that representing metadata quality (meta)data support understanding and communicate meaning (Dai et al., 2021; Mandal et al., 2016; Melo et al., 2021). Finally, the influence on making decisions and its outcomes has been mentioned by multiple studies (Dyson & Foster, 1982; Harley & Cooper, 2021; Price & Shanks, 2011; Shankaranarayanan et al., 2006, 2008; Shankaranarayanan & Zhu, 2021; Stvilia et al., 2007).

The final stage focuses on gaining insights and making sound decisions. Participants explain how data quality and metadata help them get insights and make sound decisions (B2.MDRDQ, A2.MDRDQ,A3.MDRDQ, D.MDRDQ). They discuss the difficulties associated with assessing data quality and propose segmenting data quality into data concepts and entities for greater understanding and trust. The significance of a centralized process for data quality evaluation and using tools such as Collibra and Power BI (B1.MDRDQ) for data quality assessment and visualization is stressed. At this point, businesses can extract valuable insights and make data-driven decisions by employing metadata and data quality information. The paragraph before has

elaborated on decision-making. The literature also supports the difficulties with assessing data quality (Cichy & Rass, 2019; Liu et al., 2021; Zhang et al., 2019).

In conclusion, the participants' views provide useful insights into the data management/governance stages. These stages include metadata and data type awareness and knowledge, metadata categorization, metadata management and appropriate technologies, data quality management, and the use of metadata and data quality information for insights and decision-making. The findings emphasize the importance of effective metadata management, data quality evaluation, and data quality information utilization to ensure trustworthy and valuable data assets. Thus, the suggested proposition is:

Proposition 7: Having insight in all data governance/management stages has a positive effect on the representation of metadata.

Effective data management

Effective data management is the third theme, which has led to the third dimension Data Governance Lifecycle. In the sections below, the results will be analyzed that were obtained during the interviews in relation to effective data management. The findings will be compared with the literature which has been discussed in chapter 2 Theoretical Background. On the basis of this comparison and the theme, a conclusion will be drawn, and a proposition will be proposed.

The operationalization of metadata was emphasized as important by participants. Using data governance tools and procedures to operationalize metadata management helps establish control, create an audit trail, and ensure successful data governance (A3.MD, F.MD). DAMA International (2017) mentions that proper data governance requires, among other things, the development of a data management framework. In the literature is mentioned that the standardization of metadata increase metadata governance practices and improves data quality (Aamot, 2022; IEEE, 2020; Yan et al., 2022).

Participants highlighted the importance of data quality for effective data management (B2.MDRDQ). Data quality management is critical for improving knowledge and trust and making sound decisions. Data quality management should also be aligned with various business processes and require continual improvement to address concerns such as correctness, currency, completeness, and integrity. The above statements have been elaborated on throughout this discussion. It is highlighted in chapter 2.3 that the data must be of reliable and correct for organizations and processes to drive decisions and actions (Harley & Cooper, 2021; Stvilia et al., 2007).

Continuous improvement and standardization were also important aspects of effective data management. Participants emphasized the significance of developing processes to ensure data quality, conduct quality checks, and automate operations (D.MDRDQ, A3.MDRDQ).

Standardization through enterprise-wide standards and business-rule-based monitoring helps establish a uniform language and data quality (A3.MDRDQ). These techniques enable organizations to reach higher levels of data quality, reduce errors caused by hidden logic and permission concerns, and constantly enhance their data management operations. Throughout this research the benefits of standardization of metadata has been pointed out.

In the literature multiple challenges were found with capturing, organizing, and retaining metadata (IEEE, 2020; Loshin, 2015; Sundarraj & Rajkamal, 2019; van Helvoirt & Weigand, 2015). The challenges include ensuring metadata accuracy, consistency, and completeness, and addressing metadata integration and interoperability issues. Employing effective metadata management strategies is crucial to overcome these challenges and ensure the dependability and usability of data assets.

In conclusion, effective data management needs a comprehensive approach that includes metadata management, a thorough understanding of data types, and a focus on improving data quality. Organizations must establish data governance tools, metadata management techniques, and continuous improvement procedures to enable effective data management. The views shared by the participants provide unique perspectives on the issues and tactics involved in preserving and increasing data quality, data governance, and trust in complex data contexts. Thus, the suggested proposition is:

Proposition 8: Managing data effectively has a positive effect on the representation of metadata.

Privacy in data management

Privacy in data management is the fourth theme, which has led to the third dimension Data Governance Lifecycle. In the sections below, the results will be analyzed that were obtained during the interviews in relation to privacy in data management. The findings will be compared with the literature which has been discussed in chapter 2 Theoretical Background. On the basis of this comparison and the theme, a conclusion will be drawn, and a proposition will be proposed.

The participants in the study highlight several key points regarding privacy in data management, which provide valuable insights into addressing privacy concerns effectively.

One of the key results is the significance of an organized strategy in distinguishing and preserving unique customer data (C.MDRDG). This organized approach ensures that sensitive data, such as client addresses or household compositions, are properly identified and protected. Organizations can build effective governance processes to ensure privacy compliance and secure data handling by implementing explicit structures and metadata. As mentioned before in the theme data quality aspects, metadata is helpful in dividing data into sensitive and private information (DAMA International, 2017). Organizations have to comply with regulations such as

the GDPR. Therefore, they have to implement data governance and management practices to know what sensitive information it has.

Participants mention that visual representation of access rights is important in establishing accountability and privacy compliance (C.MDRDG, B2.MDRDG). A clear visual overview of who can access what data allows for better data management and control. This transparency enables enterprises to monitor and control data access, lowering the risk of unauthorized or inappropriate data usage and improving privacy protection.

The need for particular metadata storage in following privacy standards specified by Data Protection Boards is also emphasized (B2.MDRDG). Metadata is vital for privacy compliance since it provides critical information about the data and its associated privacy needs. By adding essential information, organizations can ensure compliance with privacy requirements and the proper management and protection of sensitive data. Throughout this discussion the importance of metadata has been highlighted.

The use of technical metadata in data management and privacy is another critical consideration. While technical information may not be directly relevant to end users or administrators, it is essential for efficient data management. Technical metadata gives useful information about the data, making operations like discovering missing tables or columns and requesting data updates or changes easier. When dealing with data-related requests while ensuring privacy and security safeguards, access to metadata becomes critical. As mentioned in the chapter 2.2, technical metadata is needed to run the systems and delivering digital data that can be understood (Gartner, 2016).

In conclusion, the participants' perspectives highlight the importance of privacy considerations in data management. Organizations should use organized techniques, establish clear visualizations of access rights, and integrate essential metadata in their data governance policies to maintain privacy compliance and secure sensitive information. Organizations can achieve a balance by doing so between preserving data security, complying with privacy rules, and making data accessible to authorized individuals for legal purposes. Privacy in data management is an ongoing undertaking that necessitates constant attention and modification to satisfy increasing privacy concerns and regulatory obligations properly. Thus, the suggested proposition is:

Proposition 9: Insight into privacy with metadata has a positive effect on the representation of metadata.

5.5 Definition of participants

In this section the definitions given by the participants are compared to the definitions used for this research. The definitions can be found at the end of each sub-chapter in the results section.

Metadata

First, all definitions of the participants state that metadata is the information about data to help to understand it. Thus, the broad definition "data about data" applies. Next to this definition, three participants mention the differentiation of metadata into business and technical metadata. Two participants mention data models and structures in their definitions.

The definition used in this research was adopted from DAMA International (2017) *"Metadata includes information about technology and business processes, data rules and constraints, and logical and physical data structures. It describes the data itself (e.g., databases, data elements, data models), the concepts the data represents (e.g., business processes, application systems, software code, technology infrastructure), and the connections (relationships) between the data and concepts"* (DAMA International, 2017, p. 417).

Overlap can be seen when comparing the definitions. Both definitions emphasize business and technical aspects. Next, two participants describe metadata as information about data, meaning the structure, format, and relationships between the data. This aligns with the definition of DAMA.

Metadata Representation

The different views of the participants show some overlap. Two participants take into consideration the context of when metadata is represented. Others highlight the importance of clear, easy, and friendly metadata representation.

The first view where the context has to be considered can be linked to Ackhoff's (1989) pyramid, where the data layer has no meaning without context. The lowest layer will step up from data to information by providing context.

This research also mentioned the Representation Theory of Wand & Weber (1990). They state that the communication of information in Representation Theory (RT) is based on four assumptions that revolve around the central concept of meaning (Burton-Jones et al., 2017):

(v) RT can communicate meaning through symbols, and humans can obtain meaning from these symbols (Wand & Weber, 1995).
(vi) Information systems intend to represent someone's or a group's view of the real world (Wand & Weber, 1995). Wand & Weber (1993) and Burton-Jones & Grange (2013) reason that computerized representations provide a more efficient way to learn about the world than observations. This leads to the design of information systems.
(vii) Stakeholders can express their meaning about events that are of interest to them (Burton-Jones et al., 2017).
(viii) Users expect that an information system has more value when it represents a true view of the real world (Weber, 1997).

Holding the view of the participants against the four assumptions in the Representation Theory, the first assumption is that communication of meaning through symbols is a simple and understandable way to extract meaning. This can be used to present (meta)data to non-technical and business users. Secondly, it can be seen in the second assumption that information systems intend to represent someone's or a group's view of the real world. The participants mentioned that this should be done clearly, easily, and friendly.

Data Governance

The definitions given have overlapping results. Four participants mentioned the importance of ownership and roles in data governance. They also state that it is important to set up policies and create standards in the organization. Two participants recognize that data governance is needed to ensure the right level of controls, security, and availability to make data fit for purpose. Others emphasize the need for data governance to comply with policies and standards. It is also highlighted that data governance is needed to manage data properly.

The definition used in this research given by Abraham et al. (2019) was *"Data governance specifies a cross-functional framework for managing data as a strategic enterprise asset. In doing so, data governance specifies decision rights and accountabilities for an organization's decision-making about its data. Furthermore, data governance formalizes data policies, standards, and procedures and monitors compliance"* (p. 425-426).

Comparing both definitions, it can be seen that the participants also mention decision rights and accountability. They also mention that data governance helps with complying with policies and standards.

Data Quality

Three participants mentioned the importance of data quality dimensions. They name dimensions, including completeness, accuracy, and reliability, highlighting the need to measure data quality according to dimensions. Two participants highlight the role of data structure, documentation, and descriptions. They also highlight the need for accurate and trustworthy data to ensure efficient processes. Others highlight the business side of data quality. They mention that it is important to align business objectives with data quality.

This research used the results from an article of Wand & Wang (1996). They stated that data is usually measured in terms of data quality dimensions. In their literature review, the following dimensions were found to be most used: (i) Accuracy; (ii) Reliability/consistency; (iii) Timeliness (currency); and (iv) Completeness.

Comparing the views of the participants against the article of Wand & Wang (1996), the dimensions accuracy, reliability, and completeness are mentioned by both.

5.6 Implications

This chapter highlights the implications of metadata representation in the context of the research questions.

The research question is "How could the representation of Metadata affect Data Governance and Data Quality in organizations?" Several sub-questions were created, to answer the research question:

1. How is metadata represented in organizations?
2. What is metadata representation in a data governance context in organizations?
3. What is metadata representation in a data quality context in organizations?

The representation of metadata has an important role in data governance because it provides structured information on data elements, definitions, access rights, and compliance rules. It enables organizations to successfully develop and implement data governance rules and standards, resulting in better data management, control, and governance throughout their lifecycle.

Organizations obtain an accurate understanding of their data assets, their characteristics, and their relationships by capturing metadata in a systematic and organized manner. This adds context and helps with data comprehension, assisting data users in understanding the meaning and purpose of the data. Organizations can make better-informed decisions and lower the risk of data misinterpretation or misuse by improving data comprehension.

In addition to improving data comprehension, metadata representation supports compliance efforts by documenting information on compliance requirements and standards. Organizations can track and manage data compliance requirements such as data privacy rules (e.g., GDPR), industry-specific regulations (e.g., HIPAA), or internal policies. Organizations can reduce compliance risks and possible sanctions by properly representing metadata.

Moreover, metadata representation is strongly related to data quality. Organizations can set guidelines and criteria for analyzing and improving data quality by documenting metadata associated with data quality standards. Metadata contains information about data sources, transformation methods, lineage, and validation criteria, all of which are necessary for understanding data quality issues and establishing suitable data quality controls. This proactive approach enables organizations to discover and address data quality issues, resulting in more accurate and reliable data.

Furthermore, the results found that there are multiple aspects important to consider representing metadata effectively:

(i) Technical skills: organizations should have technical skilled employees to effectively represent metadata. There should be a balance between complexity and consistency, enabling end users to engage with metadata. It is highlighted that the representation of metadata should be integrated into applications by visualizing them. The ability to generate data quality criteria and measures using logical reasoning and translate them into information using metadata representation also requires technical skills.

(ii) Data tooling: organizations should implement tools to represent metadata. The objective would be to integrate the tools into already existing processes. Collibra is mentioned by the participants as a centralized center for metadata and is to be well-known in the industry.

(iii) Privacy: visualizing the access and control rights allows for improved data management and control. Which lowers risks and helps with compliance.

(iv) Context: participants emphasized that context should be provided when representing metadata. This will increase the understanding of metadata.

5.7 Relevance

Academical relevance

The literature about the combinations of metadata & data governance and metadata & data quality have not been researched extensively, as mentioned throughout this research. The focus of the current literature is about one of the three concepts or focused on a specific domain within the concept.

In Table 10 can be seen that after a careful selection of the literature, there are seven articles on metadata representation, ranging from years 2003 to 2021. Twelve articles are selected on metadata & data quality with a range of 2006 to 2021. Furthermore, seven articles on metadata & data governance were selected ranging from 2015 to 2022. As can be seen from the article years, data governance is a relatively new concept. This research contributes to the literature by exploring the combinations of metadata & data governance and metadata & data quality, more specifically the representation of metadata in combination with data governance and data quality in (large) organizations.

The founded themes in the proposed data structure contribute to the literature about metadata representation, data governance, and data quality by introducing nine themes aggregated into three dimensions. The results are in line with previous studies and theories. For example, the lowest layer ‘raw data’ in the pyramid of Ackhoff (1989) requires context to step up a layer. This

is supported by multiple participants. The results also mentioned that metadata should be represented in a clear way, for non-technical and end users. The Representation Theory of Wand & Weber (1990) can help with representing metadata in information systems and make it understandable. See the aforementioned sections after the analysis of the themes for more links to the theory.

Practical relevance

As mentioned throughout this research, the combinations of metadata & data governance and metadata & data quality have not been researched extensively. This research explored these topics and from the results a data structure has been created. Metadata can help with capturing knowledge about data, dividing data into private or sensitive and managing the lifecycle of data to meet with compliance and regulations. Representing metadata can help with several challenges. For example, structured representation of metadata helps with the discovery of patterns and retrieval of information. Furthermore, it helps with data literacy among employees. Representing metadata in a clear, accessible, and understandable way helps with showing how important metadata is. The data structure that has been created can be used by organizations working with metadata, data governance, and data quality. The themes in the structure can be used to help organizations with representing their metadata. Focusing on multiple themes of the structure will improve (meta)data management/governance and data quality. The practical relevance also seemed high by Clever Republic.

5.8 Limitations

This chapter discusses the limitations of this research and possible improvements to reduce the issues.

As with all studies conducted, this research also has its limitations. First of all, this research had a time constraint of four months in which the study had to be conducted and there was no funding for resources.

Secondly, this research is generalizable only to the given sample universe mentioned in the methodology in chapter 3. The organizations selected had to be of a 'large' scale according to the definition of the Ministerie van Economische Zaken en Klimaat (2021), which stated that an organization had over 250FTE, or the net sales were more than €50 million **and** a more than €43 million balance sheet. All organizations also had to work with metadata, utilizing data governance, and measuring data quality. Finally, the organizations had to operate in the Netherlands but were not limited to the Netherlands. To generalize the findings to a broader range of organizations, the research should include all organizations, from small to large. Due to the time constraint this research could not include all organizations.

Third, in total nine interviews at six organizations have been held. The diverse sample of organizations operating in different industries creates a generalizable finding for the organizational view. However, in the best-case scenario, a larger sample size would be preferrable with multiple participants at different levels in the organization. This would make the findings even more valid. Due to time constraints nine interviews were held. It was also found that after nine interviews, not many more new codes were found during the coding process.

Fourth, due to the used research approach, exploratory qualitative research, the findings are influenced by the researchers' experience in using the research approach. The results are also influential to the subjectivity of the researcher. This has been tried to prevent by applying a proven method during the study, the Grounded Theory.

Finally, the findings can be improved even further by conducting a quantitative research approach with a greater sample size and using quantitative methods to get definitive results. The generalizability of the findings would be higher due to the sample size.

6. Conclusion

This research explored the concepts of metadata representation, data governance, and data quality. Specifically, how the representation of metadata could affect data governance and data quality in organizations. Therefore, the research question in this thesis was *"How could the representation of Metadata affect Data Governance and Data Quality in organizations?"*

In Chapter 2 Theoretical Background was found that there was insufficient literature on the (combined) concepts metadata representation, metadata & data governance, and metadata & data quality. Therefore, this study used an exploratory qualitative approach in answering the research question. The Grounded Theory method was followed to gather and analyze data. After analyzing the results from the interviews, a data structure was created (see Figure 2, Chapter 5). This data structure consists of nine themes that were aggregated into three dimensions. In Table 9 the dimensions and themes can be seen.

The given answers provided insight into the view of organizations on metadata representation in a data governance and data quality context.

- Metadata representation in the context of data governance involves presenting metadata in a structured and organized manner. This includes important details about data elements, definitions, quality standards, access rights, and compliance regulations. The purpose is to enable enterprises to better manage, govern, and control their data. Metadata provides valuable context to data, which helps in understanding and meeting compliance. It also ensures data quality and supports the establishment and enforcement of data governance rules and standards.
- Metadata representation in a data quality context refers to structured information about data, including its characteristics and attributes. Metadata in organizations helps in the understanding and assessment of data quality by providing information on data sources, verification processes, and business rules. It encourages the use of a single language, effective communication, and a shared understanding of data. Metadata connects technical elements and business goals, allowing for efficient data governance and data quality control.

Conclude, metadata representation facilitates effective data governance, helps compliance efforts, improves data comprehension, enables data quality management, and encourages communication and collaboration between stakeholders. Organizations that invest in well-structured metadata representation are better positioned to effectively manage and govern their data while maintaining high data quality standards.

Table 9 - Dimensions and themes

Dimension I – Improving understanding	**Dimension II – Data proficiency**	**Dimension III – Data Governance Lifecycle**
Understanding challenges	Technical skills	(Meta)data management importance
Business process improvement	Data tooling	Stages of data management/governance
	Data quality aspects	Effective data management
		Privacy in data management

6.1 Future research

The research used an exploratory research approach. This type of research is useful when an industry is understudied. Below, future research based on the propositions mentioned after the in-depth analysis of the themes is given. Finally, recommendations for future research are given based on the theoretical background aligned with this research.

Proposition 1: Understanding the challenges has a positive effect on the representation of metadata.

Future research can delve into the unique challenges that organizations encounter with metadata representation. Researchers can investigate how firms identify and overcome these challenges by conducting empirical studies or case studies. This study can shed light on the strategies, tactics, and technologies used by businesses to improve metadata representation and the influence on overall information management.

Proposition 2: Incorporating metadata positively affects business processes.

Future research can concentrate on specific businesses or domains to better understand the benefits of adding metadata into business processes. This research can look into how metadata integration affects various business processes such as decision-making, resource allocation, and performance evaluation. Researchers can analyze the actual benefits and problems associated with metadata integration in various organizational contexts by studying real-world scenarios and conducting quantitative studies.

Proposition 3: Technical skilled employees have a positive effect on the representation of metadata.

Further research might examine the specific skills, knowledge, and competences required for good metadata management to acquire a complete comprehension of the influence of technical skilled employees on metadata representation. This research could include surveying

professionals and conducting interviews to determine which skill sets contribute the most to metadata quality and representation. Furthermore, examining the impact of training programs and educational interventions on improving technical expertise can offer organizations significant information.

Proposition 4: Implementing specific data tools has a positive effect on the representation of metadata.

Further research can look into the adoption and usefulness of specific data tools for improving metadata representation. This could include comparing various data tools, rating their functions, and determining their impact on metadata quality and representation. Researchers can perform experiments or case studies to investigate the advantages and disadvantages of various tools, resulting in useful recommendations for organizations looking to improve their metadata management processes.

Proposition 5: High metadata quality has a positive effect on the representation of metadata.

Future research can focus on building comprehensive frameworks and approaches for analyzing metadata quality in order to better understand the relationship between information quality and its representation. This may include creating measurements, undertaking data audits, and assessing the influence of metadata quality on decision-making processes. Researchers can help to set standards and guidelines for good metadata management by developing accurate and meaningful measurements of metadata quality.

Proposition 6: Managing metadata has a positive effect on the representation of metadata.

Proposition 7: Having insight into all data governance/management stages have a positive effect on the representation of metadata.

Future research can investigate the relationships and interactions between these stages to provide insight into the positive effects of having an in-depth understanding of all data governance and management stages on metadata representation. This research can look into how metadata representation affects data governance processes and vice versa. Researchers can find best practices for aligning metadata management with broader data governance strategies by evaluating real-world examples and performing qualitative investigations.

Proposition 8: Managing data effectively has a positive effect on the representation of metadata.

Future research could look into data management as a key factor affecting metadata representation. This may include researching data management frameworks, policies, and practices in organizations to determine the most effective methods for ensuring data accuracy, consistency, and completeness. Researchers can provide practical suggestions for firms aiming

to maximize their metadata management efforts by researching the relationship between data management strategies and metadata representation.

Proposition 9: Insight into privacy with metadata has a positive effect on the representation of metadata.

Further research could explore the privacy issues related to metadata management strategies to gain insight into the impact of privacy considerations on metadata representation. This research may include investigating the legal and ethical implications of metadata use, conducting surveys or interviews with stakeholders, and reviewing privacy rules and guidelines. Researchers could provide recommendations for enterprises to manage privacy concerns while effectively managing metadata if they understand the privacy risks associated with metadata representation.

Finally, the themes in the data structure highlight the different aspects organizations have to acknowledge when they want to represent their metadata. All themes should be considered; however, a few themes were explicitly highlighted by participants for representing metadata effectively. These were technical skills, data tooling, and privacy. Another important aspect mentioned is the context in which the metadata is represented.

7. References

Aamot, S. (2022). *Indigenous data governance and sovereignty: a crosswalk comparison of four metadata schemas* [e University of North Carolina at Chapel Hill]. https://doi.org/https://doi.org/10.17615/qj9e-wx11

Abebe, M. A., Tekli, J., Getahun, F., Chbeir, R., & Tekli, G. (2020). Generic metadata representation framework for social-based event detection, description, and linkage. *Knowledge-Based Systems*, *188*, 104817. https://doi.org/10.1016/j.knosys.2019.06.025

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

Ackoff, R. L. (1989). From data to wisdom. *Journal of Applied Systems Analysis*, *16*(1), 3–9.

Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2019). A systematic literature review of data governance and cloud data governance. *Personal and Ubiquitous Computing*, *23*(5), 839–859. https://doi.org/10.1007/s00779-017-1104-3

Ashish, N., Dewan, P., & Toga, A. W. (2016). The GAAIN Entity Mapper: An Active-Learning System for Medical Data Mapping. *Frontiers in Neuroinformatics*, *9*. https://doi.org/10.3389/fninf.2015.00030

ATLAS.ti. (n.d.). *Coding Data - Basic Concepts - ATLAS.ti 22 Windows - Quick Tour*. Retrieved June 5, 2023, from https://doc.atlasti.com/QuicktourWin.v22/Codes/CodingDataBasicConcepts.html?highlight=grounded#how-groundedness-is-counted

Backman, K., & Kyngäs, H. A. (1999). Challenges of the grounded theory approach to a novice researcher. *Nursing & Health Sciences*, *1*(3), 147–153. https://doi.org/10.1046/j.1442-2018.1999.00019.x

Bansal, M., Chana, I., & Clarke, S. (2021). A Survey on IoT Big Data: Current Status, 13 V's Challenges, and Future Directions. *ACM Computing Surveys*, *53*(6). https://doi.org/10.1145/3419634

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM Computing Surveys*, *41*(3). https://doi.org/10.1145/1541880.1541883

Bearman, M. (2019). Focus on Methodology: Eliciting rich data: A practical approach to writing semi-structured interview schedules. *Focus on Health Professional Education: A Multi-Professional Journal*, *20*(3), 1–11. https://doi.org/10.11157/fohpe.v20i3.387

Becker, D., Jaster, J., & Kuperman, J. (2009). Flexible and Generic Data Quality Metadata Exchange. *ICIQ*, 31–45. http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202009/Papers/2-A.pdf

Begg, C., & Caira, T. (2012). Exploring the SME Quandary: Data Governance in Practise in the Small to Medium-Sized Enterprise Sector. *Electronic Journal of Information Systems Evaluation*, *15*(1), pp3-13. https://academic-publishing.org/index.php/ejise/article/view/237

Bergdahl, M., Ehling, M., Elvers, E., Földesi, E., Körner, T., Kron, A., Lohauß, P., Mag, K., Morais, V., & Nimmergut, A. (2007). Handbook on data quality assessment methods and tools. In M. Ehling & T. Körner (Eds.), *Ehling, Manfred Körner, Thomas*. EUROPEAN COMMISSION (EUROSTAT).

Berger, P. L., & Luckmann, T. (1967). *The social construction of reality: A treatise in the sociology of knowledge*. Anchor.

Bergmann, H., Mosiman, C., Saha, A., Haile, S., Livingood, W., Bushby, S., Fierro, G., Bender, J., Poplawski, M., Granderson, J., & Pritoni, M. (2020). *Semantic Interoperability to Enable Smart, Grid-Interactive Efficient Buildings*. https://doi.org/10.20357/B7S304

Bikauskaite, A., Gramaglia, L., Götzfried, A., & Linden, H. (2014). Better data quality through global data and metadata sharing. *European Conference on Quality in Official Statistics (Q2014)*, *3*, 5. https://www.q2014.at/fileadmin/user_upload/Global_data_and_metadata_sharing.pdf

Birkinshaw, J., Brannen, M. Y., & Tung, R. L. (2011). From a distance and generalizable to up close and grounded: Reclaiming a place for qualitative methods in international business research. *Journal of International Business Studies*, *42*(5), 573–581. https://doi.org/10.1057/jibs.2011.19

Boldyreva, E. (2018). *Cambridge Analytica: Ethics And Online Manipulation With Decision-Making Process*. 91–102. https://doi.org/10.15405/epsbs.2018.12.02.10

Brodie, M. L. (1980). Data quality in information systems. *Information and Management*, *3*(6), 245–258. https://doi.org/10.1016/0378-7206(80)90035-X

Bryant, A. (2002). Re-grounding grounded theory. *Journal of Information Technology Theory and Application (JITTA)*, *4*(1), 7. https://aisel.aisnet.org/jitta/vol4/iss1/7

Bryant, A. (2003). A constructive/ist response to Glaser. About Barney G. Glaser: constructivist grounded theory? Published in FQS 3 (3). *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, *4*(1).

Bryant, A., & Charmaz, K. (2007). Grounded theory in historical perspective: An epistemological account. In *The SAGE handbook of grounded theory* (pp. 31–57). https://doi.org/10.4135/9781848607941

Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.1819486

Brynjolfsson, E., & McElheran, K. (2016). Data in Action Data-Driven Decision Making in U.S. Manufacturing. *SSRN Electronic Journal*, Art. 2722502. https://doi.org/10.2139/ssrn.2722502

Bunge, M. (1977). *Treatise on basic philosophy: Ontology I: the furniture of the world* (Vol. 3). Kluwer Academic Publishers.

Bunge, M. (1979). *Treatise on basic philosophy: Ontology II: a world of systems* (Vol. 4). Kluwer Academic Publishers.

Burton-Jones, A., & Grange, C. (2013). From use to effective use: A representation theory perspective. *Information Systems Research*, *24*(3), 632–658. https://doi.org/10.1287/isre.1120.0444

Burton-Jones, A., Recker, J., Indulska, M., Green, P., & Weber, R. (2017). Assessing Representation Theory With a Framework for Pursuing Success and Failure. *MIS Quarterly*, *41*(4), 1307–1334. https://www.jstor.org/stable/26630295

Charmaz, K. (1990). 'Discovering'chronic illness: using grounded theory. *Social Science & Medicine*, *30*(11), 1161–1172. https://doi.org/10.1016/0277-9536(90)90256-R

Charmaz, K. (2000). Constructivist and objectivist grounded theory. In *Handbook of qualitative research* (Vol. 2, pp. 509–535).

Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis*. sage.

Charmaz, K. (2016). Shifting the grounds: Constructivist grounded theory methods. In *Developing grounded theory* (pp. 127–193). Routledge.

Charmaz, K., & Belgrave, L. (2012). Qualitative interviewing and grounded theory analysis. *The SAGE Handbook of Interview Research: The Complexity of the Craft*, *2*, 347–365. https://doi.org/10.4135/9781452218403

Childs, R. A., Ram, A., & Xu, Y. (2019). Combining dual scaling with semi-structured interviews to interpret rating differences. *Practical Assessment, Research, and Evaluation*, *14*(1), 11. https://doi.org/10.7275/wtdw-1c03

Cichy, C., & Rass, S. (2019). An Overview of Data Quality Frameworks. *IEEE Access*, *7*, 24634–24648. https://doi.org/10.1109/ACCESS.2019.2899751

Codd, E. F. (1970). A relational model of data for large shared data banks. *Communications of the ACM*, *13*(6), 377–387. https://doi.org/10.1145/357980.358007

Côrte-Real, N., Ruivo, P., & Oliveira, T. (2020). Leveraging internet of things and big data analytics initiatives in European and American firms: Is data quality a way to extract business value? *Information & Management*, *57*(1), 103141. https://doi.org/10.1016/J.IM.2019.01.003

Costin, A., & Eastman, C. (2019). Need for Interoperability to Enable Seamless Information Exchanges in Smart and Sustainable Urban Systems. *Journal of Computing in Civil Engineering*, *33*. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000824

Cuthill, M. (2002). Exploratory research: citizen participation, local government and sustainable development in Australia. *Sustainable Development*, *10*(2), 79–89. https://doi.org/10.1002/sd.185

Daas, P. J. H., & Ossen, S. J. L. (2011, May 20). Metadata quality evaluation of secondary data sources. *5th International Quality Conference*. http://www.cqm.rs/2011/cd/5iqc/pdf/096.pdf

Dai, C., Füllgrabe, A., Pfeuffer, J., Solovyeva, E. M., Deng, J., Moreno, P., Kamatchinathan, S., Kundu, D. J., George, N., & Fexova, S. (2021). A proteomics sample metadata representation for multiomics integration and big data analysis. *Nature Communications*, *12*(1), 5854. https://doi.org/10.1038/s41467-021-26111-3

DAMA International. (2017). *DAMA-DMBOK* (2nd ed.). Technics Publications LLC.

Daniel, C., Sinaci, A., Ouagne, D., Sadou, E., Declerck, G., Kalra, D., Charlet, J., Forsberg, K., Bain, L., & Mead, C. (2014). Standard-based EHR-enabled applications for clinical research and patient safety: CDISC–IHE QRPH–EHR4CR & SALUS collaboration. *AMIA Summits on Translational Science Proceedings*, *2014*, 19. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4419753/

Dasu, T. (2013). Data Glitches: Monsters in Your Data. In S. Sadiq (Ed.), *Handbook of Data Quality: Research and Practice* (pp. 163–178). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36257-6_8

Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, *84*(1), 98–107, 134. https://hbr.org/2006/01/competing-on-analytics

Davenport, T. H., Barth, P., & Bean, R. (2012, September 18). How 'Big Data' Is Different. *MIT Sloan Management Review*.

Davis, J., Mengersen, K., Bennett, S., & Mazerolle, L. (2014). Viewing systematic reviews and meta-analysis in social research through different lenses. *SpringerPlus*, *3*(1), 511. https://doi.org/10.1186/2193-1801-3-511

Deppenwiese, N., Duhm-Harbeck, P., Ingenerf, J., & Ulrich, H. (2019). MDRCupid: A Configurable Metadata Matching Toolbox. *MedInfo*, 88–92. https://doi.org/10.3233/SHTI190189

Dhillon, G. (2019, May 22). *The Importance Of A Data Governance Framework*. Forbes. https://www.forbes.com/sites/forbestechcouncil/2019/05/22/the-importance-of-a-data-governance-framework/?sh=5d162d093ee8

Dion, M. (2007). Metadata an integral part of Statistics Canada data quality framework. *Fourth International Conference on Agriculture Statistics, Beijing, China*. http://www.stats.gov.cn/english/ICAS/papers/P020071114297538592238.pdf

Dyché, J., & Levy, E. (2006). *Customer data integration: Reaching a single version of the truth* (Vol. 7). John Wiley & Sons.

Dyson, R. G., & Foster, M. J. (1982). The relationship of participation and effectiveness in strategic planning. *Strategic Management Journal - 1980 to 2009*, *3*(1), 77–88. https://doi.org/10.1002/smj.4250030107

Edmond, S. (2021, November). *25 Interesting facts about data science*. DataCamp. https://www.datacamp.com/blog/25-interesting-facts-about-data-science

Eichenlaub, N., Morgan, M., & Masak-Mida, I. (2021). *Undressing Fashion Metadata: Ryerson University Fashion Research Collection*. https://doi.org/10.32920/ryerson.14637945.v1

Eisenhardt, K. M. (1989). Making fast strategic decisions in high-velocity environments. *Academy of Management Journal*, *32*(3), 543–576. https://doi.org/10.5465/256434

Eisenhardt, K. M., Graebner, M. E., & Sonenshein, S. (2016). Grand challenges and inductive methods: Rigor without rigor mortis. *Academy of Management Journal*, *59*(4), 1113–1123. https://doi.org/10.5465/amj.2016.4004

Elberskirch, L., Binder, K., Riefler, N., Sofranko, A., Liebing, J., Minella, C. B., Mädler, L., Razum, M., van Thriel, C., Unfried, K., Schins, R. P. F., & Kraegeloh, A. (2022). Digital research data: from analysis of existing standards to a scientific foundation for a modular metadata schema in nanosafety. *Particle and Fibre Toxicology*, *19*(1), 1. https://doi.org/10.1186/s12989-021-00442-x

English, L. P. (2009). *Information Quality Applied: Best Practices for Improving Business Information, Processes and Systems* (1st ed.). Wiley. https://archive.org/details/informationquali0000engl

Esnaola-Gonzalez, I. (2021). Towards Publishing Ontology-Based Data Quality Metadata of Open Data. In M. Bramer & R. Ellis (Eds.), *Artificial Intelligence XXXVIII* (pp. 371–376). Springer International Publishing.

Evans, N., Fourie, L., & Price, J. (2012). Barriers to the effective deployment of information assets: The role of the executive manager. *Proceedings of the European Conference on Management, Leadership & Governance*, *7*, 162–169.

Fan, W., & Bifet, A. (2013). Mining Big Data: Current Status, and Forecast to the Future. *SIGKDD Explor. Newsl.*, *14*(2), 1–5. https://doi.org/10.1145/2481244.2481246

Fisher, T. (2009). *The data asset: How smart companies govern their data for business success*. John Wiley & Sons.

Forum Standaardisatie. (2013, October 4). *RDF | Forum Standaardisatie*. RDF | Forum Standaardisatie. https://www.forumstandaardisatie.nl/open-standaarden/rdf

Forum Standaardisatie. (2016a, November 15). *RDFa | Forum Standaardisatie*. Forum Standaardisatie. https://www.forumstandaardisatie.nl/open-standaarden/rdfa

Forum Standaardisatie. (2016b, November 15). *URI en IRI | Forum Standaardisatie*. Forum Standaardisatie. https://www.forumstandaardisatie.nl/open-standaarden/uri-en-iri

Forum standaardisatie. (2021, April 29). *RDFS | Forum standaardisatie*. Forum Standaardisatie. https://www.forumstandaardisatie.nl/open-standaarden/rdfs

Friese, S. (2016). *CAQDAS and grounded theory analysis* (MMG Working Papers Print). https://www.mmg.mpg.de/61762/wp-16-07

Garcia-Molina, H., Ullman, J., & Widom, J. (2008). *Database systems: the complete book* (2nd ed.). Pearson Education India.

Gartner, R. (2016). *Metadata*. Springer.

Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. (2017). Finding Theory–Method Fit: A Comparison of Three Qualitative Approaches to Theory Building. *Journal of Management Inquiry*, *27*(3), 284–300. https://doi.org/10.1177/1056492617706029

Gioia, D. A. (2004). A renaissance self: Prompting personal and professional revitalization. In P. J. Frost & R. E. Stablein (Eds.), *Renewing research practice: Scholars' journeys* (pp. 97–114). Stanford University Press Stanford, CA.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods*, *16*(1), 15–31. https://doi.org/10.1177/1094428112452151

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Aldine Transaction. https://books.google.nl/books?id=oUxEAQAAIAAJ

Goodchild, M. F. (2007). Beyond metadata: Towards user-centric description of data quality. *Proceedings, Spatial Data Quality 2007 International Symposium on Spatial Data Quality, June*, 13–15. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=d1db7fb23b854ef891753eb92a1527c4850255dc

Grover, V., Chiang, R. H. L., Liang, T.-P., & Zhang, D. (2018). Creating Strategic Business Value from Big Data Analytics: A Research Framework. *Journal of Management Information Systems*, *35*(2), 388–423. https://doi.org/10.1080/07421222.2018.1451951

Groves, M. (2022, October 7). *What is Data Modeling? Conceptual, Physical, Logical*. Couchbase, Inc. https://www.couchbase.com/blog/conceptual-physical-logical-data-models/

Guerra, E., & Fernandes, C. (2013). A qualitative and quantitative analysis on metadata-based frameworks usage. *Computational Science and Its Applications–ICCSA 2013: 13th International Conference, Ho Chi Minh City, Vietnam, June 24-27, 2013, Proceedings, Part II 13*, 375–390.

Hagmann, J. (2013). Information governance–beyond the buzz. *Records Management Journal*, *23*(3), 228–240. https://doi.org/https://doi.org/10.1108/RMJ-04-2013-0008

Harley, K., & Cooper, R. (2021). Information Integrity: Are We There Yet? *ACM Comput. Surv.*, *54*(2). https://doi.org/10.1145/3436817

Haupt, M. (2018, June 14). *"Data is the New Oil" — A Ludicrous Proposition - Project 2030*. Medium. https://medium.com/project-2030/data-is-the-new-oil-a-ludicrous-proposition-1d91bba4f294

Huang, G., Yuan, M., Li, C., & Sun, Q. (2017). Research on ontology generation and evaluation method in oil field based on the MDR. *Journal of Computational Methods in Sciences and Engineering*, *17*, 665–676. https://doi.org/10.3233/JCM-170751

IEEE. (2020). IEEE IC Big Data Governance and Metadata Management: Standards Roadmap. *IEEE IC Big Data Governance and Metadata Management: Standards Roadmap*, 1–62. https://ieeexplore.ieee.org/servlet/opac?punumber=9133345

International Organization for Standardization (ISO). (2023). *Data quality — Part 2: Vocabulary (ISO Standard No. 8000-2:2022)*. https://www.iso.org/standard/85032.html

Ishwarappa, & Anuradha, J. (2015). A Brief Introduction on Big Data 5Vs Characteristics and Hadoop Technology. *Procedia Computer Science*, *48*, 319–324. https://doi.org/10.1016/j.procs.2015.04.188

Jeong, S., Kim, H. H., Park, Y. R., & Kim, J. H. (2014). Clinical Data Element Ontology for Unified Indexing and Retrieval of Data Elements across Multiple Metadata Registries. *Healthc Inform Res*, *20*(4), 295–303. https://doi.org/10.4258/hir.2014.20.4.295

Kanakia, H., Shenoy, G., & Shah, J. (2019). Cambridge Analytica: A case study. *Indian Journal of Science and Technology*, *12*(29), 1–5. https://doi.org/10.17485/ijst/2019/v12i29/146977

Karkošková, S. (2023). Data Governance Model To Enhance Data Quality In Financial Institutions. *Information Systems Management*, *40*(1), 90–110. https://doi.org/10.1080/10580530.2022.2042628

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

Kock-Schoppenhauer, A.-K., Bruland, P., Kadioglu, D., Brammen, D., Ulrich, H., Kulbe, K., Duhm-Harbeck, P., & Ingenerf, J. (2019). Scientific Challenge in eHealth: MAPPATHON, a Metadata Mapping Challenge. In *MEDINFO 2019: Health and Wellbeing e-Networks for All* (pp. 1516–1517). IOS Press. https://doi.org/10.3233/SHTI190512

Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, *6*(70), 1.

Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review*, *24*(4), 691–710. https://doi.org/10.5465/amr.1999.2553248

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2010, December 12). Big Data, Analytics and the Path From Insights to Value. *MIT Sloan Management Review*.

Li, Z., Wen, J., Zhang, X., Wu, C., Li, Z., & Liu, L. (2012). ClinData Express – A Metadata Driven Clinical Research Data Management System for Secondary Use of Clinical Data. *AMIA Annual Symposium Proceedings*, *2012*, 552. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3540469/

Liu, Q., Feng, G., Tayi, G. K., & Tian, J. (2021). Managing Data Quality of the Data Warehouse: A Chance-Constrained Programming Approach. *Information Systems Frontiers*, *23*(2), 375–389. https://doi.org/10.1007/s10796-019-09963-5

Locke, K. (2001). Grounded theory in management research. In R. Thorpe & M. Easterby-Smith (Eds.), *Grounded Theory in Management Research*. Sage.

Löpprich, M., Jones, J., Meinecke, M.-C., Goldschmidt, H., & Knaup, P. (2014). A Reference Data Model of a Metadata Registry Preserving Semantics and Representations of Data Elements. *Studies in Health Technology and Informatics*, *205*, 368–372. https://doi.org/10.3233/978-1-61499-432-9-368

Loshin, D. (2015). Enterprise Data Architecture and Data Governance: Use Metadata to Get to the Starting Gate. *Konowledge Integrity Inc*. https://knowledge-integrity.com/blog2/wp-content/uploads/2015/12/Enterprise-Data-Architecture-and-Data-Governance-final.pdf

Luborsky, M. R., & Rubinstein, R. L. (1995). Sampling in qualitative research: Rationale, issues, and methods. *Research on Aging*, *17*(1), 89–113. https://doi.org/10.1177/0164027595171005

Mandal, A. K., Sarkar, A., & Debnath, N. C. (2016). Context driven metadata representation for SaaS. *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, 826–831. https://doi.org/10.1109/INDIN.2016.7819274

Maumet, C., Auer, T., Bowring, A., Chen, G., Das, S., Flandin, G., Ghosh, S., Glatard, T., Gorgolewski, K. J., Helmer, K. G., Jenkinson, M., Keator, D. B., Nichols, B. N., Poline, J.-B., Reynolds, R., Sochat, V., Turner, J., & Nichols, T. E. (2016). Sharing brain mapping statistical results with the neuroimaging data model. *Scientific Data*, *3*(1), 160102. https://doi.org/10.1038/sdata.2016.102

Mayer, R. E. (2003). The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction*, *13*(2), 125–139. https://doi.org/https://doi.org/10.1016/S0959-4752(02)00016-6

Mcafee, A., & Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, *90*(10), 60–66, 68, 128. https://hbr.org/2012/10/big-data-the-management-revolution

Melo, D., Rodrigues, I. P., & Varagnolo, D. (2021). A strategy for archives metadata representation on CIDOC-CRM and knowledge discovery. *Semantic Web*, *Preprint*, 1–32. https://doi.org/10.3233/SW-222798

Mills, A. J., Durepos, G., & Wiebe, E. (2009). *Encyclopedia of case study research*. Sage publications.

Ministerie van Economische Zaken en Klimaat. (2021, August 4). *Mkb of grote onderneming?* Rijksdienst Voor Ondernemend Nederland. https://www.rvo.nl/subsidies-financiering/tvl/mkb-grote-onderneming

Moges, H.-T., Vlasselaer, V. Van, Lemahieu, W., & Baesens, B. (2016). Determining the use of data quality metadata (DQM) for decision making purposes and its impact on decision outcomes — An exploratory study. *Decision Support Systems*, *83*, 32–46. https://doi.org/https://doi.org/10.1016/j.dss.2015.12.006

Morse, J. M. (2016). Tussles, tensions, and resolutions. In *Developing grounded theory* (pp. 13–22). Routledge.

Myrseth, P., Stang, J., & Dalberg, V. (2011). A data quality framework applied to e-government metadata: A prerequsite to establish governance of interoperable e-services. *2011 International Conference on E-Business and E-Government (ICEE)*, 1–4. https://doi.org/10.1109/ICEBEG.2011.5881298

Ngouongo, S. M. N., Löbe, M., & Stausberg, J. (2013). The ISO/IEC 11179 norm for metadata registries: Does it cover healthcare standards in empirical research? *Journal of Biomedical Informatics*, *46*(2), 318–327. https://doi.org/https://doi.org/10.1016/j.jbi.2012.11.008

NICOLESCU, I. A. (2019). From big data to data: The value behind Metadata Governance in Statistics. *Economic Convergence in European Union*, *309*. http://store.ectap.ro/suplimente/Theoretical_&_Applied_Economics_2019_Special_Issue_Summer.pdf#page=309

NISO. (2017). Understanding metadata. *National Information Standards Organization (NISO)*. https://www.academia.edu/download/82117494/Understanding_20Metadata.pdf

Otto, B. (2013). On the Evolution of Data Governance in Firms: The Case of Johnson & Johnson Consumer Products North America. In S. Sadiq (Ed.), *Handbook of Data Quality: Research and Practice* (pp. 93–118). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-36257-6_5

Panian, Z. (2010). *Some practical experiences in data governance*. *38*, 150–157. https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=173626431ed114b232c1dab8ae60a005a0d593dc

Papež, V., & Mouček, R. (2017). Applying an Archetype-Based Approach to Electroencephalography/Event-Related Potential Experiments in the EEGBase Resource. *Frontiers in Neuroinformatics*, *11*. https://doi.org/10.3389/fninf.2017.00024

Park, J.-R., & Tosaka, Y. (2010). Metadata Quality Control in Digital Repositories and Collections: Criteria, Semantics, and Mechanisms. *Cataloging & Classification Quarterly*, *48*(8), 696–715. https://doi.org/10.1080/01639374.2010.508711

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

Pomerantz, J. (2015). *Metadata*. MIT Press.

Price, R., & Shanks, G. (2011). The impact of data quality tags on decision-making outcomes and process. *Journal of the Association for Information Systems*, *12*(4), 1. https://doi.org/10.17705/1jais.00264

Ramakrishnan, R., Gehrke, J., & Gehrke, J. (2003). *Database management systems* (Vol. 3). McGraw-Hill New York.

Robinson, O. C. (2014). Sampling in Interview-Based Qualitative Research: A Theoretical and Practical Guide. *Qualitative Research in Psychology*, *11*(1), 25–41. https://doi.org/10.1080/14780887.2013.801543

Rogers, S., & Thompson, K. (2012). Outperforming in a Data-Rich, Hyper-Connected World. *IBM Center for Applied Insights, IBM Corporation, Armonk, NY*. https://dokumen.tips/documents/outperforming-in-a-data-rich-hyper-connected-a-sitescene-a-custom-a-userfiles.html?page=1

Rytsar, Y., Voloshynovskiy, S., & Pun, T. (2003). Metadata representation for semantic-based multimedia security and management. *On The Move to Meaningful Internet Systems 2003: OTM 2003 Workshops: OTM Confederated International Workshops, HCI-SWWA, IPW, JTRES, WORM, WMS, and WRSM 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings*, 769–778.

Salkind, N. J. (2010). *Encyclopedia of Research Design*. SAGE Publications, Inc. https://doi.org/10.4135/9781412961288

Schippers, M. C., & Rus, D. C. (2021). Optimizing decision-making processes in times of COVID-19: using reflexivity to counteract information-processing failures. *Frontiers in Psychology*, *12*, 650525. https://doi.org/10.3389/fpsyg.2021.650525

Schutz, A. (1972). *The phenomenology of the social world*. Northwestern university press.

Shankaranarayanan, G., EVEN, A., & Sussman, S. (2006). The role of process metadata and data quality perceptions in decision making: An empirical framework and investigation. *International Journal of Information Technology and Management - IJITM*, *17*. https://www.researchgate.net/profile/Ganesan-Shankaranarayanan/publication/255607113_The_role_of_process_metadata_and_data_quality_perceptions_in_decision_making_An_empirical_framework_and_investigation/links/540d970b0cf2df04e754d2b5/The-role-of-process-metadata-and-data-quality-perceptions-in-decision-making-An-empirical-framework-and-investigation.pdf

Shankaranarayanan, G., & Zhu, B. (2021). Enhancing decision-making with data quality metadata. *Journal of Systems and Information Technology*, *23*(2), 199–217. https://doi.org/10.1108/JSIT-08-2020-0153

Shankaranarayanan, G., Zhu, B., & Cai, Y. (2008). Decision Support with Data Quality Metadata. *ICIQ*, 281–295. http://mitiq.mit.edu/ICIQ/Documents/IQ%20Conference%202008/Papers/3C-1%20Shankaranarayanan%20et%20al.pdf

Shmueli, G., Bruce, P. C., Yahav, I., Patel, N. R., & Lichtendahl Jr, K. C. (2017). *Data mining for business analytics: concepts, techniques, and applications in R*. John Wiley & Sons.

Simon, H. A. (1996). The sciences of the artificial. In *MIT Press Cambridge* (3rd ed.). The MIT Press.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333–339. https://doi.org/10.1016/J.JBUSRES.2019.07.039

Song, T.-M., Park, H.-A., & Jin, D.-L. (2014). Development of Health Information Search Engine Based on Metadata and Ontology. *Healthc Inform Res*, *20*(2), 88–98. https://doi.org/10.4258/hir.2014.20.2.88

Specka, X., Gärtner, P., Hoffmann, C., Svoboda, N., Stecker, M., Einspanier, U., Senkler, K., Zoarder, M. A. M., & Heinrich, U. (2019). The BonaRes metadata schema for geospatial soil-agricultural research data – Merging INSPIRE and DataCite metadata schemes. *Computers & Geosciences*, *132*, 33–41. https://doi.org/https://doi.org/10.1016/j.cageo.2019.07.005

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques:Techniques and Procedures for Developing Grounded Theory*. Citeseer.

Streb, C. K. (2010). Exploratory Case Study. In A. J. Mills, G. Durepos, & E. Wiebe (Eds.), *Encyclopedia of Case Study Research* (pp. 372–374). Sage.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data Quality in Context. *Commun. ACM*, *40*(5), 103–110. https://doi.org/10.1145/253769.253804

Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A framework for information quality assessment. *Journal of the American Society for Information Science and Technology*, *58*(12), 1720–1733. https://doi.org/https://doi.org/10.1002/asi.20652

Sundarraj, M., & Rajkamal, M. N. (2019). Data governance in smart factory: Effective metadata management. *Int. J. Adv. Res. Ideas Innov. Technol*, *5*(3), 798–804. https://www.ijariit.com/manuscripts/v5i3/V5I3-1233.pdf

Taylor, P. J., Catalano, G., & Walker, D. R. F. (2002). Exploratory Analysis of the World City Network. *Urban Studies*, *39*(13), 2377–2394. https://doi.org/10.1080/0042098022000027013

Timmerman, Y., & Bronselaer, A. (2019). Measuring data quality in information systems research. *Decision Support Systems*, *126*, 113138. https://doi.org/10.1016/J.DSS.2019.113138

Torraco, R. J. (2005). Writing Integrative Literature Reviews: Guidelines and Examples. *Human Resource Development Review*, *4*(3), 356–367. https://doi.org/10.1177/1534484305278283

Trani, L., Atkinson, M., Bailo, D., Paciello, R., & Filgueira, R. (2018). Establishing Core Concepts for Information-Powered Collaborations. *Future Generation Computer Systems*, *89*, 421–437. https://doi.org/https://doi.org/10.1016/j.future.2018.07.005

Ulrich, H., Kock-Schoppenhauer, A.-K., Deppenwiese, N., Gött, R., Kern, J., Lablans, M., Majeed, R. W., Stöhr, M. R., Stausberg, J., & Varghese, J. (2022). Understanding the nature of metadata: systematic review. *Journal of Medical Internet Research*, *24*(1), e25440. https://doi.org/10.2196/25440

University of Cambridge. (n.d.). *Choosing Formats*. Retrieved June 7, 2023, from https://www.data.cam.ac.uk/data-management-guide/creating-your-data/choosing-formats

van Helvoirt, S., & Weigand, H. (2015). Operationalizing Data Governance via Multi-level Metadata Management. In M. Janssen, M. Mäntymäki, J. Hidders, B. Klievink, W. Lamersdorf, B. van Loenen, & A. Zuiderwijk (Eds.), *Open and Big Data Management and Innovation* (pp. 160–172). Springer International Publishing.

Verbitskiy, Y., & Yeoh, W. (2011). *Data quality management in a business intelligence environment : from the lens of metadata*. https://dro.deakin.edu.au/articles/conference_contribution/Data_quality_management_in_a_business_intelligence_environment_from_the_lens_of_metadata/20996557

Viljoen, S. (2021). A Relational Theory of Data Governance Feature. *Yale Law Journal*, *131*(2), 573–654. https://heinonline.org/HOL/P?h=hein.journals/ylr131&i=595

Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. *White Paper, IDC*, *14*, 1–14. https://www.admin-magazine.com/HPC/content/download/5604/49345/file/IDC_BigData_whitepaper_final.pdf

Wand, Y., & Wang, R. Y. (1996). Anchoring data quality dimensions in ontological foundations. *Communications of the ACM*, *39*(11), 86–95. https://doi.org/10.1145/240455.240479

Wand, Y., & Weber, R. (1990). An ontological model of an information system. *IEEE Transactions on Software Engineering*, *16*(11), 1282–1292. https://doi.org/10.1109/32.60316

Wand, Y., & Weber, R. (1993). On the ontological expressiveness of information systems analysis and design grammars. *Information Systems Journal*, *3*(4), 217–237. https://doi.org/10.1111/j.1365-2575.1993.tb00127.x

Wand, Y., & Weber, R. (1995). On the deep structure of information systems. *Information Systems Journal*, *5*(3), 203–223. https://doi.org/10.1111/j.1365-2575.1995.tb00108.x

Weber, R. (1997). *Ontological foundations of information systems*. Melbourne, Vic. : Coopers & Lybrand and the Accounting Association of Australia and New Zealand.

Weick, K. E. (1979). *The social psychology of organizing* (2nd ed.). McGraw-Hill, Inc.

Wende, K. (2007). A model for data governance–Organising accountabilities for data quality management. *ACIS 2007 Proceedings*, *80*. https://aisel.aisnet.org/acis2007/80

Wolpers, M., Niemann, K., & Prause, C. R. (2009). Metadata representation of real-world objects for architectural education. *2009 Ninth IEEE International Conference on Advanced Learning Technologies*, 465–467. https://doi.org/10.1109/ICALT.2009.22

Wong, G., Greenhalgh, T., Westhorp, G., Buckingham, J., & Pawson, R. (2013). RAMESES publication standards: meta-narrative reviews. *BMC Medicine*, *11*(1), 20. https://doi.org/10.1186/1741-7015-11-20

Woodward, R. L., & Masters, G. (1989). Calibration and data quality of the long-period SRO/ASRO networks, 1977 to 1980. *Bulletin of the Seismological Society of America*, *79*(6), 1972–1983. https://doi.org/10.1785/BSSA0790061972

Yamazaki, H., Slingsby, B. T., Takahashi, M., Hayashi, Y., Sugimori, H., & Nakayama, T. (2009). Characteristics of qualitative studies in influential journals of general medicine: a critical review. *Bioscience Trends*, *3*(6).

Yan, H., Wang, J., & Zhou, Y. (2022). Ontology-Based Metadata Model Design of Data Governance System. In Y. Tan & Y. Shi (Eds.), *Data Mining and Big Data* (pp. 330–342). Springer Nature Singapore.

Yen, Y. N., Weng, K. H., & Huang, H. Y. (2013). STUDY ON INFORMATION MANAGEMENT FOR THE CONSERVATION OF TRADITIONAL CHINESE ARCHITECTURAL HERITAGE–3D MODELLING AND METADATA REPRESENTATION. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *2*, 331–336. https://doi.org/10.5194/isprsannals-II-5-W1-331-2013

Yin, R. K. (2018). *Case study research and applications : design and methods / Robert K. Yin.* (Sixth edition.). SAGE.

Yourdon, E., & Constantine, L. L. (1979). Structured design. Fundamentals of a discipline of computer program and systems design. *Englewood Cliffs: Yourdon Press*.

Yu, S., Qing, Q., Zhang, C., Shehzad, A., Oatley, G., & Xia, F. (2021). Data-Driven Decision-Making in COVID-19 Response: A Survey. In *IEEE Transactions on Computational Social Systems* (Vol. 8, Issue 4, pp. 989–1002). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/TCSS.2021.3075955

Zhang, R., Indulska, M., & Sadiq, S. (2019). Discovering Data Quality Problems. *Business & Information Systems Engineering*, *61*(5), 575–593. https://doi.org/10.1007/s12599-019-00608-0

Zikopoulos, P., & Eaton, C. (2011). *Understanding Big Data: Analytics For Enterprise Class Hadoop and streaming Data* (1st ed.). McGraw-Hill Education.

8. Appendices

8.1 Appendix I: Design of the literature review

The literature review can be designed in different ways. According to Snyder (2019), three main types of methods are often used in literature reviews.

(i) Systematic review:
Used to create summaries of studies in an organized, clear, and reproducible way. The objective is to identify all empirical evidence that matches certain criteria. Biases can be prevented, through the use of structured methods and considering all evidence when reviewing an article (Davis et al., 2014).

(ii) semi-systematic: review:
Often used for subjects that have been visualized differently and have been studied by a broad group of disciplines. Conduct when a systematic review process is not feasible (Wong et al., 2013).

(iii) integrative review:
similar to the semi-systematic review but serves a different purpose. The main objective of an integrative review is to evaluate, analyze, and merge existing literature on a research topic to identify new theoretical frameworks and perspectives (Torraco, 2005).

This literature review follows the semi-systematic review method. The semi-systematic review is used because a systematic review would not be feasible in the given time of this study. The integrative review method has not been selected due to the nature of the review for this thesis.

Database

There have been used two databases to find articles. The first is the WorldCat Discovery (WorldCat) database. This database is provided by Tilburg University as their main search catalog to find articles and books. In WorldCat it is possible to filter and select multiple criteria e.g., only books and articles that have been peer-reviewed or the range of years they were published in. The second database used is Google Scholar. This is a database created by Google to search for articles and books in multiple libraries in the world matched with given search terms. When searching in both databases it is possible to use a query structure to find more specific results. In both databases the results were sorted on 'Best Match'/'Relevance'. Within the WorldCat database the criteria 'Article, chapter' and 'Books' were selected. In Google Scholar the 'include citations' option was unchecked.

The following key concepts have been searched in databases to find articles with information to answer the concepts; (i) Metadata Representation; (ii) Metadata & Data Quality; (iii) Metadata & Data Governance. See **Error! Reference source not found.** for a schematic overview.

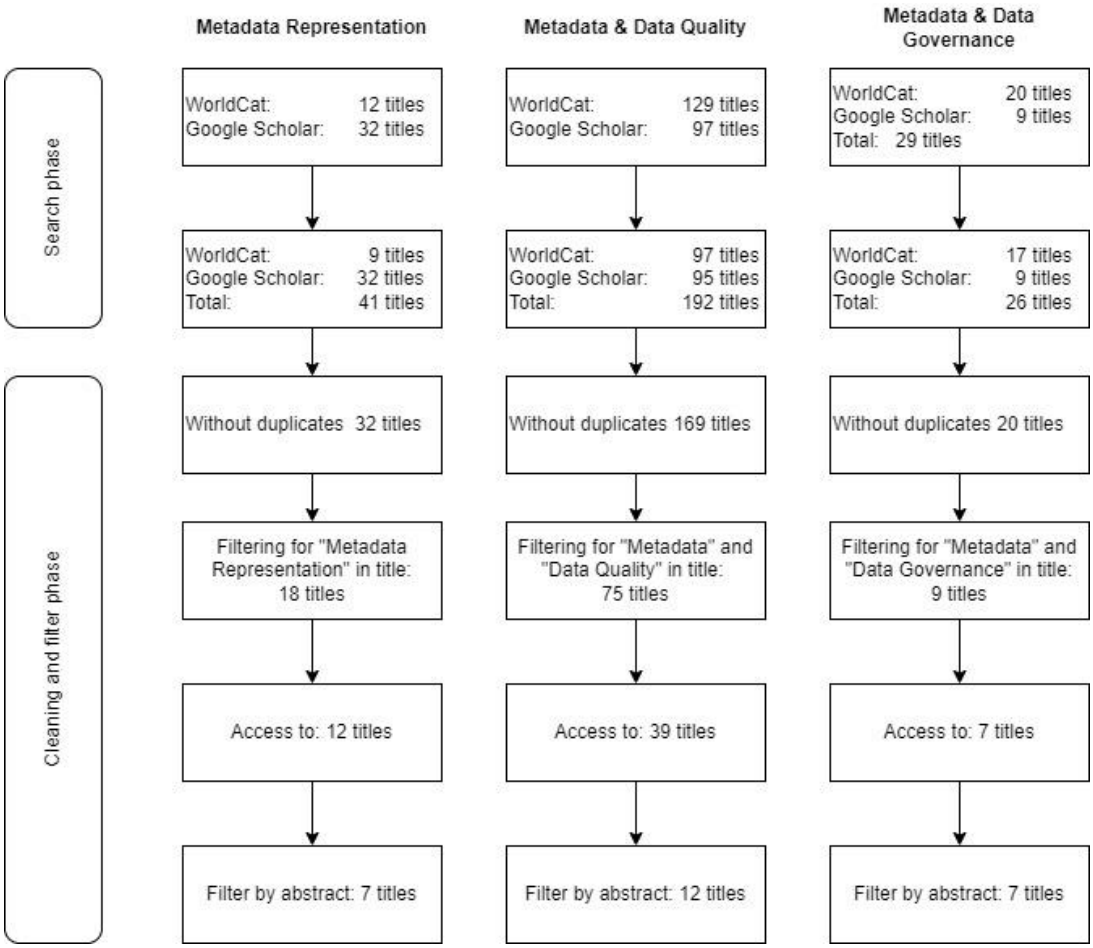

Figure 3 - Schematic overview of search for literature.

(i) Metadata Representation

Searching for the key term Metadata Representation in the two databases with the following queries:

- WorldCat: ti: "Metadata representation";
- Google Scholar: allintitle: "Metadata Representation".

In WorldCat selected the option for "in possession of library" "Tilburg University Library". This resulted in twelve titles for the WorldCat and 32 titles in the Google Scholar database. All titles were copied in Microsoft Excel (Excel) in a 'title' column and their year of publication in the next column. Both database results were copied in separate Excel tabs. By using Excel, it was possible to filter for duplicates. This was done by taking the next steps:

1. In the WorldCat titles were three duplicates found which in total make nine titles. There were no duplicates in the Google Scholar database titles resulting in 32 titles. Combining all articles in one-tab results in 41 titles;
2. Automatically deduplicating these titles through Excel four were found. This makes 37 unique titles. However, since Excel cannot find similarities but is restricted to 'if and only if equal to' it won't filter the same titles with differences and/or extra spaces. Manually going through the list where all titles were sorted in alphabetical order resulted in five duplicates. Subtracting these five titles from the 37 titles results in 32 titles.
3. Since a time, constraint is involved a selection has to be made of these 32 titles. This is done by filtering the titles for the keywords "Metadata Representation" by using the IF, ISNUMBER and SEARCH functions. It is important to mention that Excel in these functions takes all combinations into account, therefore it is not CASE sensitive. The results gave eighteen titles. Since the duplicates were done manually there is another search for duplicates. This time there were no duplicates found.
4. Filtering for articles I do have access to, articles/books and more duplicates. Resulted in three titles that were not an article/book or not full text. Three titles I had no access to. This makes six titles that can be subtracted from the eighteen titles. Which gives a total of twelve titles.
5. Reading the abstracts to filter the articles. Five articles were excluded from the selected articles. In total seven titles were selected based on their abstract.

(ii) Metadata & Data Quality

Searching for the key terms Metadata and Data Quality in the two databases with the following queries:

- WorldCat: ti: "Metadata" AND "Data Quality";
- Google Scholar: allintitle: "Metadata" "Data Quality".

This resulted in 129 titles for the WorldCat and 97 titles in the Google Scholar database. All titles were copied in Microsoft Excel (Excel) in a 'title' column and their year of publication in the next column. Both database results were copied in separate Excel tabs. By using Excel, it was possible to filter for duplicates. This was done by taking the next steps:

1. In the WorldCat titles were 32 duplicates found which in total make 97 titles. There were found two duplicates in the Google Scholar database titles resulting in 95 titles. Combining all articles in one-tab results in 192 titles;
2. Automatically deduplicating these titles through Excel none were found. However, since Excel cannot find similarities but is restricted to 'if and only if equal to' it won't filter the same titles with differences and/or extra spaces. Manually going through the list where

all titles were sorted in alphabetical order resulted in 23 duplicates. Subtracting these titles from the 192 results in 169 titles.

3. Since a time, constraint is involved a selection has to be made of these 169 titles. This is done by filtering the titles for the keywords "Metadata" and "Data Quality" by using the IF, ISNUMBER and SEARCH functions. It is important to mention that Excel in these functions takes all combinations into account, therefore it is not CASE sensitive. The results gave 102 titles. Since the duplicates were done manually there is another search for duplicates. This time there were 27 duplicates found. Which resulted in 75 titles.
4. Filtering for articles I do have access to, articles/books and more duplicates. Resulted in five titles that were not an article/book or not full text. Thirty-one titles I had no access to. This makes 36 titles that can be subtracted from the 75 titles. Which gives a total of 39 titles.
5. Reading the abstracts to filter the articles. Found one duplicate article with a slightly different published name. One article was an editorial issue. In total 27 titles were not selected based on their abstract and form. This gives a total of twelve titles.

(iii) Metadata & Data Governance

Searching for the key terms Metadata and Data Quality in the two databases with the following queries:

- WorldCat: ti: "Metadata" AND "Data Governance";
- Google Scholar: allintitle: "Metadata" "Data governance".

This resulted in twenty titles for the WorldCat and nine titles in the Google Scholar database. All titles were copied in Microsoft Excel (Excel) in a 'title' column and their year of publication in the next column. Both database results were copied in separate Excel tabs. By using Excel, it was possible to filter for duplicates. This was done by taking the next steps:

1. In the WorldCat titles were three duplicates found which in total make seventeen titles. There were no duplicates found in the Google Scholar database titles resulting in nine titles. Combining all articles in one-tab results in 26 titles;
2. Automatically deduplicating these titles through Excel there were none found. However, since Excel cannot find similarities but is restricted to 'if and only if equal to' it won't filter the same titles with differences and/or extra spaces. Manually going through the list where all titles were sorted in alphabetical order resulted in six duplicates. Subtracting these titles from the 26 results in twenty titles.
3. Using the same steps as above filtering for articles with the keywords "Metadata" and "Data Governance" by using the IF, ISNUMBER and SEARCH functions. It is important to mention that Excel in these functions takes all combinations into account, therefore it is not CASE sensitive. The results gave nine titles.

4. Filtering for articles I do have access to, articles/books and more duplicates. Resulted in two titles that I had no access to. This makes two titles that can be subtracted from the nine titles. Which gives a total of seven titles.
5. Reading the abstracts to filter the articles. All articles were considered to be useful for this thesis.

Founded literature

The table below provides the literature from the three main subjects.

Table 10 - Literature after selection

Metadata Representation	**Metadata & Data Quality**	**Metadata & Data Governance**
(Abebe et al., 2020; Dai et al., 2021; Mandal et al., 2016; Melo et al., 2021; Rytsar et al., 2003; Wolpers et al., 2009; Yen et al., 2013)	(Becker et al., 2009; Bikauskaite et al., 2014; Daas & Ossen, 2011; Dion, 2007; Esnaola-Gonzalez, 2021; Goodchild, 2007; Moges et al., 2016; Myrseth et al., 2011; Shankaranarayanan et al., 2006, 2008; Shankaranarayanan & Zhu, 2021; Verbitskiy & Yeoh, 2011)	(Aamot, 2022; IEEE, 2020; Loshin, 2015; NICOLESCU, 2019; Sundarraj & Rajkamal, 2019; van Helvoirt & Weigand, 2015; Yan et al., 2022)

8.2 Appendix II: Interview structure

Introduction
Introducing myself

Purpose
Studies have found that there's still a lack of Metadata standards. The purpose of this interview will contribute to creating insight into Metadata in an organizational context; the representation of Metadata in organizations; the effect of Metadata representation on Data Governance and Data Quality. The research question is: *"How could the representation of Metadata affect Data Governance and Data Quality in organizations?"*

Benefits
The interviewee will get the results of this study. The study's results can be of interest in finding what Metadata representation methods are used most often in organizations. The results will outline the effect of Metadata Representation on Data Governance and Data Quality.

Confidentiality
All information shared is confidential and is only used for this study. The interviewee can get full anonymity, there will be no names mentioned.

Time
The interview time will be around one hour.

Recording of the interview
Asked before the interview if the interview could be recorded through audio.

Interview Questions

Personal information

A1. What is your current position in the organization?
A2. How long have you been working in the organization?
A3. In which (types of) positions have you been working?
A4. To what extent do you have to deal with Metadata in your current role? (range 1-6)
A5. What is your educational background?

Company information

A6. To what extent do you see Metadata as a beneficial asset to your organization? (range 1-6)
A7. To what extent would you rate your organization to be data-driven? (range 1-6)
A8. How does this rate compare to your competitors? (range 1-6)
A9. Do you feel your organization makes sufficient use of potential information? (range 1-6)
A10. What is the organizational structure of the organization?
A11. Does the information flow through the whole supply chain or is it more centralized?

General Metadata related question

B1. How would you define Metadata?
The definition used in this study;
Using the definition of DAMA International (2017*): "Metadata includes information about technology and business processes, data rules and constraints, and logical and physical data structures. It describes the data itself (e.g., databases, data elements, data models), the concepts*

the data represents (e.g., business processes, application systems, software code, technology infrastructure), and the connections (relationships) between the data and concepts"

B2. Do you think this definition captures everything?
B3. How important is Metadata in your organization? Could you rate this? (range 1-6)
B4. What are your thoughts about how Metadata has evolved over time?

Research Question specific
Metadata Representation in the organization

B5. What is your view of Metadata Representation?
B6. What Metadata Representation methods do you know?
B7. What Metadata Representation methods are used in the organization?

Metadata Representation in Data Governance

B8. What is your definition of Data Governance?
Definition in this study; *"Data governance specifies a cross-functional framework for managing data as a strategic enterprise asset. In doing so, data governance specifies decision rights and accountabilities for an organization's decision-making about its data. Furthermore, data governance formalizes data policies, standards, and procedures and monitors compliance"* (Abraham et al., 2019)
B9. Do you think this definition captures everything?
B10. What is your view on the role and importance of metadata in Data Governance?
B11. How could Metadata be represented in Data Governance?
B12. Do you have Metadata Representation techniques specifically for Data Governance in the organization?
B13. Are there relevant cases you might think of sharing?

Metadata Representation in Data Quality

B14. What is your definition of Data Quality?
Definition for this study; Data Quality is the right output of Information and it is usually evaluated in terms of its dimensions: (i) Accuracy; (ii) Reliability/consistency; (iii) Timeliness (currency); (iv) Completeness (Wand & Wang, 1996).
B15. Do you think this definition captures everything?
B16. What is your view on the role and importance of metadata in Data Quality?
B17. How could Metadata be represented in Data Quality?
B18. Do you have Metadata Representation techniques specifically for Data Quality in the organization?
B19. Are there relevant cases you might think of sharing?

Finalizing the interview

C1. Is there something you have missed during this interview or have come up with during the interview?
C2. Is there anything else you would like to share?
C3. Do you have any questions?

Rational for interview questions

Table 11 – Rational for interview questions

Question	Rational	Source
A1 A2 A3 A4 A5	These questions are used to gather information about the participant, related to their job position, and have an open starting point	(Charmaz & Belgrave, 2012)
A6 A7 A8 A9	These questions were asked to capture the participants view on (meta)data in the company. Since multiple participants in the organization can be interviewed. This rating would be interesting to compare.	(Childs et al., 2019)
A10 A11	These questions create an overview of the organizational structure and end the introductory section.	(Bearman, 2019; Charmaz & Belgrave, 2012)
B1 B2 B3 B4 B5 B6 B7 B8 B9 B10 B11 B12 B13 B14 B15 B16 B17 B18 B19	These questions were asked to capture the definition of the participants on the concepts and introduce the definitions used in this study. The questions about the exploration of metadata representation were more specific to the research question. The questions are open-ended, allowing the participant to respond freely without any predetermined direction or suggested answers.	(Charmaz & Belgrave, 2012; Childs et al., 2019)
C1 C2 C3	The questions at the end are used to transition the conversation back to normal.	(Charmaz & Belgrave, 2012)

8.3 Appendix III: The four-point approach to qualitative sampling

The four-point approach to qualitative sampling, adopted from Robinson (2014, p. 26).

Table 12 - The four-point approach to qualitative sampling from Robinson (2014, p. 26)

	Name	**Definition**	**Key decisional issues**
Point 1	Define a sample universe	Establish a sample universe, specifically by way of a set of inclusion and/or exclusion criteria.	Homogeneity vs. heterogeneity, inclusion and exclusion criteria
Point 2	Decide on a sample size	Choose a sample size or sample size range, by taking into account what is ideal and what is practical.	Idiographic (small) vs. nomothetic (large)
Point 3	Devise a sample strategy	Select a purposive sampling strategy to specify categories of person to be included in the sample.	Stratified, cell, quota, theoretical strategies
Point 4	Source the sample	Recruit participants from the target population.	Incentives vs. no incentives, snowball sampling varieties, advertising

8.4 Appendix IV: Sample universe

The sample universe in a graphical view. The circles are not scaled on real world sizes for the amount of organizations.

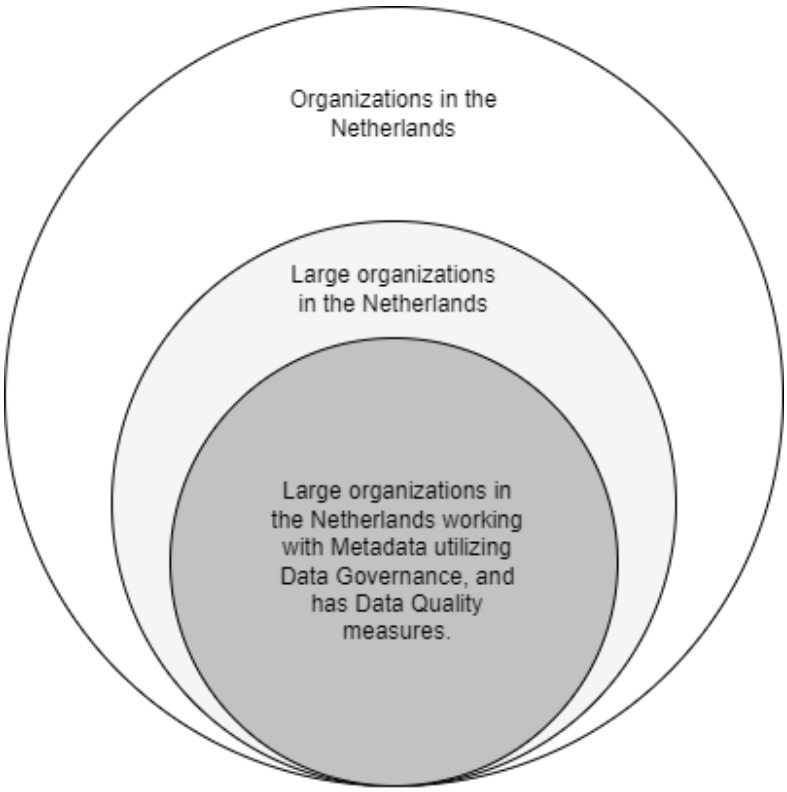

Figure 4 - Sample universe of this research

8.5 Appendix V: Interviews

Due to the confidential information given by the participants, the interviews are not publicly available. Contact the author of this thesis for more information about the interviews.

8.6 Appendix VI: Deleted category codes

These are the deleted category codes with their sub-codes.

Table 13 - Deleted category codes

Category code	Sub-code
Referencing	- Forgetting things - Mentioning/Referring to Something - Reference data
Research methodology	- Importance - Importance of Guidelines - Importance of planning - Importance of Schema Design - Primary keys - Problem solving - Process improvement - Process orientation - Structured approach
Attention to detail	- Importance - Importance of Guidelines - Importance of planning - Importance of Schema Design - Primary keys

8.7 Appendix VII: Citations in result chapter

Metadata general

Lack of clarity (mentioned by 6)

"What is it? What? What does it mean to them and where is the value and why? And and and we we at least then do not have that kind of meta data management strategy defined yet. And I think it's important to either and and it doesn't have to be a strategy, but it can be just what are your guidelines around metadata? What are you know we we we have to start doing and documenting uh. The approach causes we right now we're kind of we think we know what we're doing. Then we're doing it right when uh, it may not necessarily be correct we you know we we should be looking we should be following standards and guidelines. As regards metadata management." – A3.MD

"Well, the other way around you are required that you have your data well described, all relationships have good ownership, then you can only monitor your data quality properly, adjust it

well and get it into your system from the beginning, because that is often where People are. Just do something and then check data quality afterwards, are you going to flange again? No, you. Should actually be done in advance. Know exactly what your definitions are, so you have that system you use to fill in something, that's already. It is and I find that meta data with then the breadth of metadata that I have. Well, that definition that you just used that If you have all that right, then you can influence your data quality well. And, it is almost a kind of resultant." – E.MD

Recognition/Awareness (mentioned by 7)

"Well, I think it's very important with us In the organization, but then I'm mainly about more the definitions of Some data. Where we often use a lot of different names for the same thing, so It's more about the vocabulary around a not so much data, but also just even machine lines. We have a lot of different names for that and that's already metadata, so. What do you call? A production line and what do you call a machine that puts a label on a bottle? One says labeler, the other says labeling machine the other, says Etiqa. And actually they all mean the same thing and That's all meta data. Things are already going wrong at that level, which makes it very difficult. Each time you have to switch between different samples. As it were. And, you see that across the board. That there is no unambiguous definition there. English, Dutch mixed together. So yes, yes, it's a bit of chaos." – E.MD

"Yes, because it's just not recognized. Yes, every time they run into it, but then taking the action remains a fight." – E.MD

"Yeah, I think there is more recognition of it. Mesh data as. Being important as being something that should be captured. Even the term. Helped along by. My good old Mr. Zuckerberg is renaming his company. Yeah, but that does make a difference that I think there is way more recognition of Metadata as a concept and its importance. I don't think it's there yet. I think it's going to continue to grow; I really do. It's an interesting angle to actually pick up. Yeah, I think it's really developing. I think there's going to be and meeting Someone Like You with this kind of focus. I can see where you will be in a few years' time so." – A1.MD

"Yes, and the awareness, so how important they think it is to say yes how important is metadata for your organization. if You ask me is Of course. 6 nothing is more important than almost nothing more important than that. That is especially when something breaks or when you have to do impact analysis. If you really need to go back and look at your data to see what's actually happening. And those are efforts in companies that take so much time and effort. And that actually always happens when it's under pressure, because then either something is broken or

you go. Want to do a project? Well, he wanted durations et cetera and then you're super happy that you have that metadata, so It's just you the right archiving of what's happening in your Company, then it's not even archiving. Yes, it's constantly archiving, but it's just keeping an eye on what's going on huh? Data should not be a Black box, you just have to know exactly what is happening with all the key data you have. Not all the data you have, that's nonsense, but all the key data you have just know what happens to that??" – A2.MD

"We live and breathe my I live and breathe metadata, and when I say metadata more on the business metadata than more on the technical metadata. Right? Right. My team do a lot on the metadata side on on the technical side, but I think that's why. Metadata needs to actually be broken down. In my personal view, people don't understand what we mean when we say metadata. It's data on data, OK? I don't care. It's data on data. Why is that an? Issue. What are you actually trying to do? Where does that add value? I think for me you need to start breaking it down and and saying OK from a business metadata perspective, this is what we mean from a technical metadata. This is what we mean. So, we live and breathe this we're obviously implementing. And also having a data governance tool to help us operationalize. Our processes around mess business and technical metadata. And we we. Use the tooling. We use the whole metadata management side. To ensure or to help [COMPANY NAME] to stay in control. It's all about that. They know where your audit trail. It's an audit trail on your data. We do see in. I do see that on the technical level of metadata is probably the easiest part because yeah, but the business side of the metadata is where you really need. To help people understand what that means and what's the value. And that business metadata can be from also standards rules. We we forget. That piece of it and that's. Where a lot of the effort needs to come in." – A3.MD

"I think without without knowing where your data is, how do you know how you're going to dispose it? And like that's again looking from from the kind of the very technical uh view and then on the on the privacy piece and and the the the how do you classify your data then I see that kind of at that middle middle layer, the kind of the business layer. So, I yeah, I. I'm a strong believer. It's it's very important, but it's it's again, it's it's awareness that needs to happen and. It's not just for technical eyes, metadata management. They should be used across an. Enterprise; they should be able to see the value." – A3.MD

Data Management (mentioned by 9)

"Yes and the awareness, so how important they think it is to say yes how important is metadata for your organization. if You ask me is Of course. 6 nothing is more important than almost nothing more important than that. That is especially when something breaks or when you have to do impact analysis. If you really need to go back and look at your data to see what's actually happening. And those are efforts in companies that take so much time and effort. And that actually always happens when it's under pressure, because then either something is broken or you go. Want to do a project? Well, he wanted durations et cetera and then you're super happy that you have that metadata, so It's just you the right archiving of what's happening in your Company, then it's not even archiving. Yes, it's constantly archiving, but it's just keeping an eye on what's going on huh? Data should not be a Black box, you just have to know exactly what is happening with all the key data you have. Not all the data you have, that's nonsense, but all the key data you have just know what happens to that??" – A2.MD

"I think that's much more important nowadays, because I think you just have a lot more data at all. At the exponential it just goes up, so it's kind of out of control actually. Then you can often no longer see the forest for the trees and that is what metadata is for. Become more important and. Because there is much more, you also have a lot more meta data, but you try the. Structure the in hold. I think it's especially important too late there, so it's just becoming more and more important." – B2.MD

"Well, we could we we we have done it. But we can get better at it, and I think where metadata management also helps is around the GDPR and PII. It helps you identify. And helps the not only

business but also technical people. Technical you know, kind of developers or the IT type of business analysts or engineers? It helps for them to identify what is PII data, what, where and what. What can be GDPR related? I think that's underestimated the power of what metadata management as well can do for an organization. Also, it for me metadata management plays also a very important role in data retention and disposal." -A3.MD

"True, correct and and that has of course also had its repercussions in metadata management, Because you so. It also makes no sense to think to have the ambition that you have all the metadata. In one fell swoop you have to bring in, model and bring it underground then you don't solve a problem. So you really have to look at where are the pain points? Where can meter let management? Making a difference or solving a problem?" – F.MD

Metadata type (mentioned by 7)

"We live and breathe my I live and breathe metadata, and when I say metadata more on the business metadata than more on the technical metadata. Right? Right. My team do a lot on the metadata side on on the technical side, but I think that's why. Metadata needs to actually be broken down. In my personal view, people don't understand what we mean when we say metadata. It's data on data, OK? I don't care. It's data on data. Why is that an? Issue. What are you actually trying to do? Where does that add value? I think for me you need to start breaking it down and and saying OK from a business metadata perspective, this is what we mean from a technical metadata. This is what we mean. So, we live and breathe this we're obviously implementing. And also having a data governance tool to help us operationalize. Our processes around mess business and technical metadata. And we we. Use the tooling. We use the whole metadata management side. To ensure or to help [COMPANY NAME] to stay in control. It's all about that. They know where your audit trail. It's an audit trail on your data. We do see in. I do see that on the technical level of metadata is probably the easiest part because yeah, but the business side of the metadata is where you really need. To help people understand what that means and what's the value. And that business metadata can be from also standards rules. We we forget. That piece of it and that's. Where a lot of the effort needs to come in." – A3.MD

"Let's get back to metadata. I figured it out. So we have very brief definition standing in. The conceptual framework stands for data about data. And then explained in line with DMBOK, a distinction is made between different types of metadata business Metadata concerns meaning and state quality of data and aspects that relate to data governance. A few examples of technical metadata concern technical storage of data and information systems. The examples and operational metadata concerns details related to processes and access to data in systems. in 3 so in 3 parts yes." – B1.MD

"Well, that's about how many characters is it? A date field, so varchar I know a lot what that kind of thing That was Metadata that's it. Name. Maybe description, then it became a bit more difficult, because hey, do you have room for that in your data warehouse? That was about the question and the rest of those d ie fields, which was somewhere like a label to your data, so then that was metadata, just like your computer files that have a few certain number of things, That's your metadata and that's how I thought about it, actually I think. 10 years ago. And so it's broader field than that. That's only from the last 5 years or so that there are a lot more aspects involved and that's really a discipline that actually requires a chief data officer in a company. Yes, That's really only from the last 5 years 4 years. Well, I've been working for a long time 5 years that you can see that there is a need for it, But the definitions and how that works exactly that we all struggle with that as a company. And still, so we come from far like. If you're just starting a new business, you can do well from the start. And what you see is that a lot of companies did it that way now and uh metadata, That's Just those labels and that's IT. And that for myself that has also grown in the past 10 years as a yes that something has to be done about that. And especially focused on data quality to be very honest that, I can see that. More and

more as a common thread? It has to be enabling to get and keep your data quality right." – E.MD

"For business yes, and well okay, now let's get to the definition of metadata. yes, look, I think that's regardless of the the Technology and it. Say The time in which we? Asking such a question: the meaning of the word metadata has never changed. It's just descriptive data, isn't it? It comes from the Greek meta is above it, it's all you do. What's underneath? Defines so It is It is data about data. And with that definition, you've covered everything, so it can have a lot of manifestations. Data model is metadata error logs from one from a system are also metadata. Within certain definitions, operational metadata. If you look at it. The scope of Van mijn product within [COMPANY NAME], then you are talking about the collibra platform. Of course, they also have their own collibra, and they have also achieved the success that they have achieved in the past 15 years by choosing a focus area very specifically. And That's okay, Let's go. A generic model. Offer for business users, but also technical users. With a free. Wide generic scope. But not everything fits in easily, does it? I would say, for example, the operational metadata. Collibra pays less attention to that. That doesn't fit so easily into the model either. They really do look at. The metadata that. Have meaning for the business. And that the links to the world of. The technical metadata. Yes and the business metadata. Often becomes. given the it. It often becomes the starting point for business metadata. Considered huh? The business glossary, eh, so say The dictionary of the business. How, how, how, how. Do you call your things and what are the definitions? And then you're going to link that to, okay. What are the different forms in which such a piece of business metadata occurs in the applications? Because you have applications from many different software packages and vendors. Both of the Shelf or In the Cloud, on Prem build itself and eventually all of those. Have underlying data storage and data models. And I think that's kind of the focus of that these days. Now write everything you find important in terms of business metadata and link that to it. Where can I find that in the organization? It is in 5 different business applications. It's called anyway, you can you slap it, then can. You can see what what means, you can also see what that some things get or have different names in different systems. So linking business metadata and technical metadata is, I think, the focus of what I'm doing now. And that in turn expanded with things like data quality tools. Measuring Quality on Certain. Objects or entities what you want to call it? Assets, but also reports key performance indicators datasets, which is also something a lot. Attention has received especially the last 5, 6, 7 years. And I think. A bit like an answer to that. The fact that the time of very clear Enterprise data warehouses that. Which is a bit over, because there are too many. So data is. Almost impossible to put in a model anymore. So you get a huge proliferation, especially the cloud platforms of all kinds of data, so there's no one knows what data you can trust and and that has led to a movement or a current that says Van. Well, then I'm going to designate certain datasets under governance, describe them, give them warships and then I call them trusted datasets. It's actually a species. You have to say if you compare to the time when you still had a very authoritative and reliable Enterprise data warehouse. Is this actually one? Because now you have to say. Okay, we have. 7000 datasets in 5 data lakes and we don't know where anymore. We can find the reliable data? For for customer or for sales orders. Or for products. Then you have. Then trust the datasets is actually a kind of band-aid on the wounds of okay, I'm going to at least make sure that I can trust the data I want and that I'm going to bring that lower limit. Yes, But I think it's 1 1 1 step back with the earlier days when it was clearer and you needed a lot less governance. And what do you actually do? Said is Okay, We have. 20 source systems, We have an Enterprise data warehouse. We have to learn that once, eh? Of course, that's going to change over time, but those changes were pretty small and all that data was transformed that the central model and you could report on it. So ready all that data and that did warehouse was. Trusted by definition. But those times are over, it's just not feasible." – F.MD

Data type (mentioned by 5)

"Well, that's about how many characters is it? A date field, so varchar I know a lot what that kind of thing That was Metadata that's it. Name. Maybe discription, then it became a bit more difficult, because hey, do you have room for that in your data warehouse? That was about the question and the rest of those d ie fields, which was somewhere like a label to your data, so then that was metadata, just like your computer files that have a few certain number of things, That's your metadata and that's how I thought about it, actually I think. 10 years ago. And so it's broader field than that. That's only from the last 5 years or so that there are a lot more aspects involved and that's really a discipline that actually requires a chief data officer in a company. Yes, That's really only from the last 5 years 4 years. Well, I've been working for a long time 5 years that you can see that there is a need for it, But the definitions and how that works exactly that we all struggle with that as a company. And still, so we come from far like. If you're just starting a new business, you can do well from the start. And what you see is that a lot of companies did it that way now and uh metadata, That's Just those labels and that's IT. And that for myself that has also grown in the past 10 years as a yes that something has to be done about that. And especially focused on data quality to be very honest that, I can see that. More and more as a common thread? It has to be enabling to get and keep your data quality right." – E.MD

"Well, we could we we we have done it. But we can get better at it, and I think where metadata management also helps is around the GDPR and PII. It helps you identify. And helps the not only business but also technical people. Technical you know, kind of developers or the IT type of business analysts or engineers? It helps for them to identify what is PII data, what, where and what. What can be GDPR related? I think that's underestimated the power of what metadata management as well can do for an organization. Also, it for me metadata management plays also a very important role in data retention and disposal." – A3.MD

"I think. We started quite well, because we could also start clean and we hired a data warehouse of the day. I think we got off to a good start in that sense. I think you have two. Do you think you see 3 changes? The first is simply if definitions of things just change, so for example first we had orders and then deliveries where the difference is. That an order is an order, only one in it would switch to deliveries Because People then you can still add it to your order, so you've got your order, oh, I'm still doing something in basket under of that sort of thing. Those are mega restructurings in the data, so those are just pure that model whole. Very much needs tweaking. So That's the data itself which is the metadata itself that I'm changing. What we also started with a bit more is. So which one really those? Lineage you say to pull over time from exactly how what happens to data. But I think what 1 big is, is the whole GDPR story and that we just do a lot too Because also Because we have simply become a lot bigger as a company and have matured that we are increasingly critical. Being sharper about who I have access to, then. Most People can't see personal information, you. And fewer and fewer People can see, for example, purchase prices. All of that sort of thing. That's a piece of security. As in from a business point of view and a piece of privacy and just ethical for example." – C.MD

Metadata Representation

Understanding the data (mentioned by 5)

"Just as well and also to a certain extent as a business business user so to in metadata Of course yes descriptive and should actually be available to a wide audience. Yes, you also want to keep it understandable, so we stay away from that technique." – A2.MDR

"And I think. Yes, that extra action is really tricky and I also don't think it's something everyone should see constantly, but Only if there is a need for it, Only then is a step to a tool that they already don't know very well anyway. Already quite a cavee." – E.MDR

"Yes, yes, Only Without real AI you could think of this too, huh? That it's just a very accessible man.search through your own catalog. Which, of course, is the goal of Collibra. But what I'm saying, You have to go to a web page, right? You have to know what you're looking for, right? Just like Google, you need to know which terms you type in, which results you are going to click on and you would prefer to go the extra mile." – E.MDR

(Meta)Data Management (mentioned by 8)

"Yes, you've probably worked with it too, but let's face it, for me collibra is just metadata. For me, it's especially important to get value out of that meter of Let, so you can really put a lot of information into it and also record about that data. But it all has to bring value, so. I would make it as simple as possible and so. Yes, keep it as intuitive as possible to display information about the number, so. You just have to be able to explain it all properly to. The business, really? It's as simple as possible, and. Especially if so. That that is a bit easier to interpret, what the data is about and what you can and cannot do with it. Kind of yes." – B2.MDR

"Well you, like I said, we use collibra, That's a place where you can get data in by having multiple People working on it, so It's collaboration library what does that stand for? It's a separate tool, so People have to get back in there. They have to understand that, even if it's very user-friendly as far as I'm concerned. A lot of People are scared of it So what I would like is that it happens very close to the consumption of data, so that you can see over something with your mouse. There are also all kinds of nice solutions from collibra, but they have not been implemented with us and That is also very difficult, because If you move your mouse over a number, yes, what does the user want to see? So with some artificial intelligence from Hey. This person who works in finance, so they might want to have the finance definition of these fields. That would be the ideal image for me that you can just there, well, augmented reality behind the Overlay something Show something that If you wanted it, that you see it in the place where you consume it, the data immediately something about it or on the other hand, the place where you input something and we use it. SAP Well, I don't know. Have you ever experienced SAP up close? They dried, that's really from those very old fill-in fields in a fill-in form that really looks like those In the years. 90 has been developed and I think that's true. But there you would actually be again and there you also have all kinds of search functions. But they are going to look for the software Supplier input and you actually want it to show hey, This is an important field for that department with Pietje Jantje Klaasje, so if you enter something wrong here, then it has problems with it. You would prefer to have that much closer to the data there as well. And yes, it's the idea of collibra of yours. You open the screen next to it with collibra and you type something there. In, that's already too many actions that that, you just see that with all the People who are in surgery. Which it does. We could take advantage of it to look it up once in a while. They don't use it to go anywhere else in a tool unless they really feel the pain of quality of data they have. Somewhere they have to generate that they are consumers of it themselves. There you see that

they go to collibra before to get a definition. To search or to. Want to know a little more background? But there are only a few, a handful in the organization." – E.MDR

"Obviously, you can. You can use tooling to help, but I think it's more of of doing roadshows and awareness on on what, what metadata actually is and the importance. And actually, having specific use cases and how metadata having the right metadata captured as also within you know, how does a you if within a use case, how does minute, why metadata management is so important and how it supports and what's the value you're going to get. That's how I see it as from a use case basis, you can do it from two approaches. You can do use case basis and then show why it's important. Make it part of the the use case show the value. And then that slowly I think helps also them with the whole awareness around it and get people to adapt to this way of thinking to kind of change. We need to start getting the organization. To be very also. Data literate, right? And I think being data literate, metadata management can play a part in that to support being data literate within your organization." – A3.MDR

"Yes, here we are of course also dependent on the tool choice, eh? And behold, within [COMPANY NAME], metadata management is also practiced by other groups of People. Also with other tools, yes look at azure, purview. That's also a catalog though. That's also a piece of metadata. In in Hana Cloud the database of SAP they also have their own libraries and dictionaries and metadata. So there's always metadata everywhere, but the platform I'm the plot oil for is the only one that has the purpose of being Enterprise Wide. a so. Complete possible picture of Van of all data in In the enterprise. In in a model actually. So yes the. The forms of representation then depend on what collibra offers and That is. Yes, Maybe is. Interesting to mention that the In the backend collibra still has a relational database. That's PostGress. That's what I think is now an open source SQL database, so It's In the back In the backend there are still tables with rows and columns linked together by form keys and Primary keys. So it's very complicated model. And there, too, they occasionally run into limitations. And I have. Also sometimes with a number of product People talked about collibra, who also hinted that they are certainly looking at collibra. Possible switch to a graph database. So a graph database has a very different concept of modeling and storage, where you focus much more on the connections between entities. And not so much on. The the tables and relational models that that could have the advantage of that. The visualization would also make the diagramming nicer and easier, but storage technical and model technical also has other challenges. So I'll see where collibra comes in. In principle, it is irrelevant for the end user of the cloud solution how the data is modeled in storage, because that is also what it becomes, so to speak. Hidden from it. The users of the tool, so you are actually only exposed to the frontend. So you can get the data in and out and you can see it in the outer shell. But that is of course another abstraction layer of the underlying model that collibra has chosen." – F.MDR

Organization: Old VS New (mentioned by 7)

"Yes in my Visions expresses that as much as possible on a logical note. Level want to do. Because then love? You an understandable, so you want to look, if you like it then. If you already have. That technical mess gets, then you get. All those things with those underscores. And I know a lot what the system depends on how that device is, because That often makes sense how a certain yes certain source system works with it. And a Salesforce uses a completely different structure than one than an SAP, for example. Often they are also based on other Languages, etcetera. You don't see any of that at your metadata level, because That's not interesting at all. For that level, it makes sense. Logical structure you do have logical names so that it is immediately or almost immediately clear what is already there, at least a good direction is guided. And look, eventually I did. You obviously have a description of that or a definition of it. What exactly makes you? Says what you mean by it, but that makes it much more understandable. And especially if you like that business, as he wants so badly, is that that business becomes more involved in that data world. Then you have to keep logical and then it is also much easier to say look If you want to see your data flow to your reporting. Or or or want to do the whole processing, I'll take here you see metadata here, so I don't know collibra views also nice do you actually see that flow already? And if that flow also consists of relatively understandable language, it is much easier to read for Everyone who has metadata We must be reasonable, must be able to appeal to a wide wide audience." – A2.MDR

"Well you, like I said, we use collibra, That's a place where you can get data in by having multiple People working on it, so It's collaboration library what does that stand for? It's a separate tool, so People have to get back in there. They have to understand that, even if it's very user-friendly as far as I'm concerned. A lot of People are scared of it So what I would like is that it happens very close to the consumption of data, so that you can see over something with your mouse. There are also all kinds of nice solutions from collibra, but they have not been implemented with us and That is also very difficult, because If you move your mouse over a number, yes, what does the user want to see? So with some artificial intelligence from Hey. This person who works in finance, so they might want to have the finance definition of these fields. That would be the ideal image for me that you can just there, well, augmented reality behind the Overlay something Show something that If you wanted it, that you see it in the place where you consume it, the data immediately something about it or on the other hand, the place where you input something and we use it. SAP Well, I don't know. Have you ever experienced SAP up close? They dried, that's really from those very old fill-in fields in a fill-in form that really looks like those In the years. 90 has been developed and I think that's true. But there you would actually be again and there you also have all kinds of search functions. But they are going to look for the software Supplier input and you actually want it to show hey, This is an important field for that department with Pietje Jantje Klaasje, so if you enter something wrong here, then it has problems with it. You would prefer to have that much closer to the data there as well. And yes, it's the idea of collibra of yours. You open the screen next to it with collibra and you type something there. In, that's already too many actions that that, you just see that with all the People who are in surgery.

Which it does. We could take advantage of it to look it up once in a while. They don't use it to go anywhere else in a tool unless they really feel the pain of quality of data they have. Somewhere they have to generate that they are consumers of it themselves. There you see that they go to collibra before to get a definition. To search or to. Want to know a little more background? But there are only a few, a handful in the organization." – E.MDR

"Obviously, you can. You can use tooling to help, but I think it's more of of doing roadshows and awareness on on what, what metadata actually is and the importance. And actually, having specific use cases and how metadata having the right metadata captured as also within you know, how does a you if within a use case, how does minute, why metadata management is so important and how it supports and what's the value you're going to get. That's how I see it as from a use case basis, you can do it from two approaches. You can do use case basis and then show why it's important. Make it part of the the use case show the value. And then that slowly I think helps also them with the whole awareness around it and get people to adapt to this way of thinking to kind of change. We need to start getting the organization. To be very also. Data literate, right? And I think being data literate, metadata management can play a part in that to support being data literate within your organization."- A3.MDR

"But I was teaching a course. Is that here? Yes, well, so we look too. You're in a club that's inside now that actually helps us set up tool. But that's then collibra tool, I know. Does that say anything where? You actually, it actually helps to unambiguously record that metadata and also to make a link with architectural models. Yes, so That's basically what you do, but the basics are still there. Definition okay, what are we talking about definition are there synonymous. Where does it come from which source? What is the core source? Yes, I know that, but is independent for me. From what for? Language or applications actually that is actually the core and. Then you prefer to choose one or another tools to get that together." – B1.MDR

Data Management: Tools (mentioned by 8)

"Well you, like I said, we use collibra, That's a place where you can get data in by having multiple People working on it, so It's collaboration library what does that stand for? It's a separate tool, so People have to get back in there. They have to understand that, even if it's very user-friendly as far as I'm concerned. A lot of People are scared of it So what I would like is that it happens very close to the consumption of data, so that you can see over something with your mouse. There are also all kinds of nice solutions from collibra, but they have not been implemented with us and That is also very difficult, because If you move your mouse over a number, yes, what does the user want to see? So with some artificial intelligence from Hey. This person who works in finance, so they might want to have the finance definition of these fields. That would be the ideal image for me that you can just there, well, augmented reality behind the Overlay something Show something that If you wanted it, that you see it in the place where you consume it, the data immediately something about it or on the other hand, the place where you input something

and we use it. SAP Well, I don't know. Have you ever experienced SAP up close? They dried, that's really from those very old fill-in fields in a fill-in form that really looks like those In the years. 90 has been developed and I think that's true. But there you would actually be again and there you also have all kinds of search functions. But they are going to look for the software Supplier input and you actually want it to show hey, This is an important field for that department with Pietje Jantje Klaasje, so if you enter something wrong here, then it has problems with it. You would prefer to have that much closer to the data there as well. And yes, it's the idea of collibra of yours. You open the screen next to it with collibra and you type something there. In, that's already too many actions that that, you just see that with all the People who are in surgery. Which it does. We could take advantage of it to look it up once in a while. They don't use it to go anywhere else in a tool unless they really feel the pain of quality of data they have. Somewhere they have to generate that they are consumers of it themselves. There you see that they go to collibra before to get a definition. To search or to. Want to know a little more background? But there are only a few, a handful in the organization." – E.MDR

"But I was teaching a course. Is that here? Yes, well, so we look too. You're in a club that's inside now that actually helps us set up tool. But that's then collibra tool, I know. Does that say anything where? You actually, it actually helps to unambiguously record that metadata and also to make a link with architectural models. Yes, so That's basically what you do, but the basics are still there. Definition okay, what are we talking about definition are there synonymous. Where does it come from which source? What is the core source? Yes, I know that, but is independent for me. From what for? Language or applications actually that is actually the core and. Then you prefer to choose one or another tools to get that together." – B1.MDR

"Yeah, we try. We are trying to establish a single point, a single hub within the organization, to do that, and our tool for that is Collibra. So, market-leading, you know, well recognized every bank, every regulatory, almost every regulated industry has it, and that that becomes somewhere that we can anybody in the company can type in [COMPANY NAME].Collibra.com internally within the company, and they go there, right? I want to find this thing, and I want to know who's the owner, or I want to know, you know, which systems it's in, or so that's. That's what we're trying to do to establish that within the company. Thit that answer the question?" – A1.MDR

"We definitely do, and I think that's also part of our problem. So, as I mentioned, within our digital organization, we have a software engineering software development chunk. It is the biggest part, and they are developing this next Gen. digital architecture and they're using best-of-breed tools for different parts of the architecture. And so, we've got Salesforce, we've got Pega, we've got various different things, and they're setting up an API catalog and APIs to allow these different systems to talk to each other. That API catalog is not in Collibra, which is somewhere else." – A1.MDR

"Yes, I think so. Two forms of representation are those that will not disappear any time soon. On the one hand, that's Natural list overviews, isn't it? So just yes sheets with with lists that are also in Excel, but fit with rows and columns. Two-dimensionally, everyone understands that and everyone still cooperates. The other appearance is more visualized. Then you have to think about diagram. Often also Network diagram behind the graphs or diagrams that at least visualize connections, because that is just easier for the human brain to read. The limitation with visualization is Of course always volume, which only fits x number of entities on your screen. And even humans are not visually equipped, for example to read a barcode or a QR code. That human eye can't be the machine so if you present a diagram. With yes, more than 15. Let's say, up to 20 forms and and and connections. Then your brain is actually already at its Max more you can't take in, so You can always zoom in on something small in visualization and as soon as It goes as soon as you start to get more information, it quickly becomes too big. To be able to visualize, then you have to. Are you going to aggregate again or are you going to go? Slices and Dice s. So, so It's certainly not easy and I have to say, I don't think the collibra is very strong in that. I don't think their diagramming functionality is super strong. It's solid and and consistent me, but I don't find it very attractive or very nice or modern looking. I've seen other tools that can do that nicely, not necessarily pointing along, but a different diagram. So representation forms, yes, capturing. Of course, it also depends on how you have mutually agreed on the model, eh. So what you call in collibra asset model, huh? But in. In more generic terms, you can actually call it the kind of meta data model, so the data model of your meta data and depending on how complex you want to make it and how fine-grained. Also determines how the data is stored and therefore must also be represented. The problem is. Always that you have to find the right balance between. Yes full consistency and and hand coverage. On the one hand and on the other, avoiding too much complexity and too much layering. Yes, because then the average end user is soon yes, the way. Lost, then it becomes too complicated. While the model is still technically based. Can't you ultimately achieve your goal with that? So yes, I think choose an effective purposeful model. Then yes the storage, that's just. Also prompted by the tooling you use for this and the representation is therefore yes again, either search and filter for specific information and be able to connect that with other information, supported by diagramming." – F.MDR

Metadata Representation in Data Governance
Understanding (mentioned by 6)

"What is each role? The role, eh? That's part of the role anyway, so you only have a limited number of positions in the company that have anything to do with data management, so I'm Alone. But you have a lot of People have a role like a data steward or a. Data owner that is. No features, that's roles as well. There you canbe yes, so exactly that. How does that work Together?" – B1.MDRDG

"yes well so it. That's coming slowly, it's slowly becoming another step, yes. Yep project was very manual. First that pulls was really thinking out with each other. What are we talking about? We have one more thing you want to say there, We also have. There have been consultants who have said, okay, but you can just buy a data dictionary from us. So then you can walk through that thick sea. If you say the one, you've got it. There must also be that true consultancy too There is at some point so more In the insurance side there has also been supervision those that People were behind and they had to take care of very quickly. Time was, and then you can't buy it anymore. And put it down. We have very strong there, I have. Also had some discussions about it. It's about the People who have to do it and the People who work with it. But it does take time, but they also have to go through it to make it their own. So you can put it next to it and it does work, but it's a bit more of a yes not lived through by the business and that. There we have. So I can add that. So people are in the process that early, what are we doing? Yes, that's really still full of technical. Really connect?" – B1.MDRDG

"Think dat data governance just seems like it. Policy, which writes about how to store data. Or yes, what you need to have described about your data to comply with. Certain certain requirements. Yes I know is, I think just yes, the policy of one of the organization just about your data. I think that data management is more adhering to policy." – B2.MDRDG

Organizational management (mentioned by 8)

"Yes Maybe that cross-functional is also interesting, because It is more cross-team with us. Data warehouse, but also all those production systems that. So it is. Pretty hard I think to organize this kind of over a whole. How and what and and per team People can have efforts. But to keep an overview or throughout the company that find, That's quite a tricky one, I think only we know we. I'm sure you can be better at that." – C.MDRDG

"OK. Uh, yeah. So, from me from my DAMA-DMBOK. So, the, operational manage the operational managing of the data governing of the. So, if I look at what our data governance team do is probably the best way of describing it rather than me trying to actually pin it down so that they are running the. The monthly sessions or the frequent sessions get the various ownership owners together to discuss data, topics, definitions, et cetera, et cetera. That's the governance part, actually. The operational management part of managing. I should come up with a better term for that. Are you going to give me the version in a second? So, they will work on the policies, standards, and frameworks. They work on ownership and responsibility. They work with the owners, and they also help to get people trained on Dharma so they have an understanding of it. Yeah. So, it's. The functional part of data management. Yeah, it's not that I'm not happy. With my answer. But not that." – A1.MDRDG

"Yes, basis describes how you actually deal with data and how the responsibility lies for it. I think that's the crux of it. Kind of what I told you. How does that structure work? What is a structure within the company? Who is responsible for it? The data. Mandatory is yes, how do you set that

up? And if there is a conflict? Yes, how do you deal with that? I think that is indeed suitcase." – B1.MDRDG

"But I guess. It would be very useful though. It kind of being able to visualize who has access to what and which one. What is the sensitive data We have and who has access to it? And that that's also something, nice just kind of ready-made is clearly visible, because then you can also check and check people more often and things like that." – C.MDRDG

"Yes, everything that is crucial for your company, how do you ensure that you have the right data for that, which also meets the right processes and procedures and protocols within you? You inside you? Company and that the right People are responsible for it. That is of course also crucial, that responsibility" – A2.MDRDG

"Yes, I think metadata is very important, because if you don't have that, you can't do it. Yes, you can't be compliant either, I guess. Yes you can be kind of familiar with SQL yes, If you just say select star of the table and If you have that as a query and you're always going to pull in your data like that, then you just grab everything. And if there are changes, you just take them, but If you just give your metadata from. Well, I like this and this column. Then he only picks up the data you want, that you have defined and whatever you have agreed with. Yes with that. DLA or SLA. So That's it. Yes, you can't do without metadata, because. So dates are coming. To comply with that, that's just necessary, yes." – D.MDRDG

"Yes, I think it is. I did open it. Just very ordinary, very demanding. That she has to do certain things? Must give away about your data, so I have a whole section on privacy, especially me. Done and to comply. The requirements from the Dutch Data Protection Authority, then you simply have to have certain things stored. Just about your data, so you just need certain metadata. Available to comply with certain judgments . So I think it's very much to do with that." – B2.MDRDG

"I've never been that fond of the term, I must say. Because it is? Not personally, but because I notice it. A little less now, but years ago they were often seen as oh bureaucracy, you know. Well Because governance that is often mentioned in the same breath as security, compliance, governance you know, so more the. The auditors who come along with the booklet with all the rules you have to comply with, so it had that aftertaste in the beginning. But now that's a bit off. And and behold, one knows. I do, by the way, as something positive. I think, I think it has everything to do with it. Engagement, why do I choose this word Because. Governance can also be a bit of a one. Are you saying that one? Empty term words because. It's a very promising term that suggests you have it completely under control. Hey, If you control it? And controlling something implies that you. People who take accountability. And, that's in practice, right? Still quite difficult and complex, because Let's be like you. Having that discussion. We still yes almost weekly too. At [COMPANYN NAME] of okay it is even possible to designate a data for your customer data, so then that, you would say that there is a person. Hey with a real physical person with a first name and a last name within [COMPANYN NAME] who is on the payroll and

who is then appointed as a data owner. Well what does that mean? Yes, first of all, that person must feel like taking on that role, because. Yes, he's not going to be there? Extra for paid. It is. No function, it's a role. And then he needs to know what kind of roles and responsibilities go with them. I want to say that I. On Saturday night at 3 a.m. if a system crashes, that is the role of the data over. You know, so It's not obviously the solution, but I think I should actually aim for that. That you make the organization aware of the fact that one must be engaged to manage data as well as possible, including. Setting standards, managing quality ensure that you have your affairs under control and in order. that you know where the data is, how it is defined. Yes, and then you also have to appoint some People who want to take on a responsible and accountable role. But in practice, thosepeople are actually going to solve something for you are more the points of contact. And you say. You you your oak points and organization. Who help you orientate? In the organization to to. To be able to link the data to. Organization, eh? So they say. Once data has to. Following the process, so data follows process. You can see that this is also getting more attention today, the day that you have to start with the business processes and only then look at data. And then people often say okay, those People and those departments that create the data in the process, so create or change it. Or use, those are the People wherever you are. Governance has to lay down responsibilities and and I think that's reasonable, so that for me that's governance, the engagement and and and awareness aspect, yes." – F.MDRDG

Data Management and Tooling

"UM. Uh, the other, the other bit that they're doing is also the tooling. Is it mentioning that? And it's a little bit of a catch because what happens sometimes is data governance becomes all about the tooling, and they have got caught in that trap. Hey, we've, we've got Collibra. It's our data governance tool, and it's all about establishing the tool and getting the tool connected to different systems and getting people on board with the system. The tool, when actually it's the other way around, it's data governance. And then using the tool to help it whereas opposed to hey, we've got a tool, and we've got to set up the tool. And then when we set up the tool, we'll do the day's government so, but they are responsible for setting up Collibra, and so there's a technical part. Of that as. So, if you want to capture the lineage from various systems, then connect Collibra to those systems to gather it. And yeah. I think it's OK not to call it out specifically, but it does turn into that." – A1.MDRDG

"No, You have to have a way to capture you well, and then I just said yes. Those kinds of things are of course just bottom line the important ones that you really have to record, but roles and responsibilities all those kinds of things you want to keep track of somewhere in a tooling or something, because otherwise you can't facilitate governance." – E.MDRDG

"Maybe monitoring. Monitoring is another one. The operational part of it. Yeah, that's tricky because it also comes with data governance. What also happens there, similar to the tooling problem? We have a bunch of data governance analysts and specialists. And the business

departments end up saying, oh, you're the data governance person; you do it well. No, we have the, we'll lay out the policy and the standards, but you as the business are responsible for your data, and we'll give you the tooling to show you what your data is and for you to actually be in control of it, but. That's not up to us. So monitoring is a tricky one because. Yeah, we can. Say, perhaps, how many systems there are or data elements there are that don't have. But yeah, you can see. Where I'm going, yeah." – A1.MDRDG

"yes, I think that's if you look at the structure and if you look then and if you look at it. The media data say. But the information about data itself. That that, think that's very important, because I often think that the information is from something like that itself, that that's often the sensitive thing about something, you should know. We have a customer If it's just a d, just how that customer has and where they live data. Is what is more sensitive, say, and also that if you have a structure that you can easily distinguish there. Yes, so that you can say, well, you may know about that customer, for example, because you may know a part, is a family city where he lives, is a family composition. But how that real hot in which which house number That is, then again secured, while that kind of thing. I think, That's very bad. Yes, you have to have that structured. You have to be clear about that. If you ever want to be able to govern there, I guess." – C.MDRDG

"No If it is indeed If it really wants to have visual, then it is that lineage tool, so just say how, how? Everything is connected? So it is from the From the physical to the conceptual that. That actually it? It the greatest value that say But the man. Visual that collibra brings and has Of course just of any asset and you can see you very easily. Which department does certain assets belong to in which period? So you've clustered that very nicely. I think she orders it very nicely that it brings that the most value and where you feel It is, pretty well. You have to click pretty far because they have correct information to come. But all information is in just on a certain asset page. A particular element either of the dataset or of the data agreement. But it's not really in pictures, it always is. A text, I think." – B2.MDRDG

"But I I think it depends on what we mean again by. But what what to what? I think it depends on what level of metadata they would want to describe, right? Yeah. Yeah. So, if you're cataloguing, if you're cataloguing data, then yeah, I think you should be. Obviously, the best thing to do is is look for some look for a tool that can help on the cataloguing. That can be represented in visualizations for people to actually understand. I think it's good for methods that you would have. Of what is meant by your metadata. And and that should become available. So I would use tooling as much as you can and have those tooling integrated to the cross applications when when it makes sense. And very much linked to me think also your kind. Of maybe your business intelligent tooling your BI tooling." – A3.MDRDG

Privacy (mentioned by 4)

"yes, I think that's if you look at the structure and if you look then and if you look at it. The media data say. But the information about data itself. That that, think that's very important, because I often think that the information is from something like that itself, that that's often the sensitive thing about something, you should know. We have a customer If it's just a d, just how that customer has and where they live data. Is what is more sensitive, say, and also that if you have a structure that you can easily distinguish there. Yes, so that you can say, well, you may know about that customer, for example, because you may know a part, is a family city where he lives, is a family composition. But how that real hot in which which house number That is, then again secured, while that kind of thing. I think, That's very bad. Yes, you have to have that structured. You have to be clear about that. If you ever want to be able to govern there, I guess." – C.MDRDG

"But I guess. It would be very useful though. It kind of being able to visualize who has access to what and which one. What is the sensitive data We have and who has access to it? And that that's also something, nice just kind of ready-made is clearly visible, because then you can also check and check people more often and things like that." – C.MDRDG

"Yes, I think it is. I did open it. Just very ordinary, very demanding. That she has to do certain things? Must give away about your data, so I have a whole section on privacy, especially me. Done and to comply. The requirements from the Dutch Data Protection Authority, then you simply have to have certain things stored. Just about your data, so you just need certain metadata. Available to comply with certain judgments . So I think it's very much to do with that." – B2.MDRDG

"In the data governance context. Yes the good ones, because myself yes, We have more to do with the source with the end users that say but then the managers think I think they are not very interested, at least some. What I call the technical metadata. But yes, that is stored in a certain way that Everyone can read that, because in principle the data itself is not so sensitive. Then the data itself so information about the data I can from Everyone by principle. If you say Yes, I'm missing a table, say yes yes, just look in the metadata. Yes, it's not there. And then yes, then they can make a request to well, or the manager or so of yes, I want to train those new data, then you can eventually own or work Because then add for example. Yes, that process is a bit tight, but not all at once of oh, I'm missing a column or can you just add it? In your query, yes." – D.MDRDG

"With the right level of it, it's making data available. To the right, people at the right time for the right purpose. And ensuring that the data is secure." – A3.MDRDG

Data Management and Quality (mentioned by 7)

"Yes, good question. Yes, I had looked it up beforehand too, but. As a municipality, remembering it was just everything that has to do with the data flow. Well, that you kind of manage that and

quality. Tries to keep high and also try to comply with all the rules that come with it. Yes, I think that's how briefly it is. Short description." – D.MDRDG

"yes well, like I, yes, actually would, I said, Well, so what do we have? We've been very busy from the starting point, actually from what we have a that people are with to capture Together with the data stewards In the parts. Basically the definitions of each yes, then you have one. We have a dictionary or a glossary where you are actually at business term level. How do they talk about data trying to hold on? We have at a level below that dictionary level. That's what we call it Dutch. Sometimes we really have more the. Data elements captured. And with data elements, all additional information, for example that quality requirements that she must meet. We also have a whole framework with the methodologies. What is critical for the company for the business units, so critical data and. And if so, what don't we call it? It's not not critical data, but it's the data and the critical data and especially the policies and frameworks that really focus on that critical data. How do you deal with that? And also for all those, so that critical data is really recorded on the basis of a number of data quality dimensions that must meet. Well, that's actually all very, very much worked out in detail as well. Yes, for me that's actually all of that, so that's close to data quality, but it's actually metadata about the data. And what are we doing now? So in that. Data tools to link that with the physical, so really technically link with the physical data In the sources at the actually third level at the physical data level. So then you actually get that linked together and then you can see pretty quickly. Okay, what is? Well critical, and what does it do? Also with that data dictionary. Linked to this is also a control framework. So if what is critical has? Actually minimal, do you want at least a check on it or Maybe. More capture, preferably automated that has checks carried out, so. So the framework so the meta, so part of this is more data quality, But the thing to base that on is all metadata focused on yes that metadata needed to do. To be able to steer that? So what was your question was." – B1.MDRDG

"Yes, we are. We are now working on indeed just with logical Quality rules for a piece of the logical data model looking from okay, We have established data quality rules before. For yes, in this case it is. I keep losing it, I often hear waiting, I stood here in front of me. So We're working on procurement, so basically procurement and then with suppliers, so that part of Van procurement. We are working on a logical data model there and there are also certain logical quality rules on it. So we now logically want those logical quality rules to attribute that data. In the logical data model, so that we know what they all have an impact on. In the end, you want to show those out of those who will have those Quality Rules themselves those physical rules. On the data itself eventually, but we are still working on a logical level to tie thattogether." – A2.MDRDG

"Well, I can hardly imagine that you can do effective data governance without doing something about management. Thinks so is their requirement. But the footprint you need to do that. Have, you can vary. For example, I can go through I have. Yes 5 ago Worked at a smaller financial

institution. It was a fund. And asset manager, literally an asset manager, but. They did Yes fund management and they actually had a free time. Small data landscape, not many systems. About 15 to 20 systems. A number of data models and then a number of data flows and integration points. We still suffered from inconsistent definitions, quality issues, et cetera and actually did time to scratch a piece of governance set up quite small-scale with stakeholders, so we set up a data board. Are often also called a data governance Board purely to start the conversation about where are the problems? How can we? Better share our information and make it available. And then We Go Away started capturing a glossary. Whether that was called, then I believe. Anyway, the words catalog and glossary are also often used a bit interchangeably, but it also depends a bit on where you are. The context. So that catalog was basically a glossary of information about the definition. Where that data was located In the landscape. The Golden source eh so what is the the the system where that? Data is born. Any data cultural or standards that apply to that object of engagement. And, That was the list of no more than 120 objects actually, and also something along those lines and then once that too. 1 1 1. Responsible person were assigned, then you saw. That yes, very good synergy arose between the department to share knowledge and to ensure that you get more control and order in that data. So there it was. The proportion of metadata needed to achieve governance in a first iteration of maturity wasn't even that high, was it? We didn't have a car withdifficult connectors that pull in all the metadata and put it in model. That is actually Advanced users, which we also do with collibra at [COMPANYN NAME]. But in If It's Still In Its Infancy, you can actually get very far with very simple means. I have also seen certain solutions in sharepoint at very large companies that were actually linked to Excel files that were under governance and change management. To certain referential data and metadata. To manage and the single version of the truth, as it often so beautiful. Is that called you those in order? Gets, so it's possible. You sometimes don't even need very difficult resources and not so much to let go to one. Small goal to achieve." – F.MDRDG

"Yes, we are. We are now working on indeed just with logical Quality rules for a piece of the logical data model looking from okay, We have established data quality rules before. For yes, in this case it is. I keep losing it, I often hear waiting, I stood here in front of me. So We're working on procurement, so basically procurement and then with suppliers, so that part of Van procurement. We are working on a logical data model there and there are also certain logical quality rules on it. So we now logically want those logical quality rules to attribute that data. In the logical data model, so that we know what they all have an impact on. In the end, you want to show those out of those who will have those Quality Rules themselves those physical rules. On the data itself eventually, but we are still working on a logical level to tie thattogether." – A2.MDRDG

Metadata Management (mentioned by 8)

"Well, I can hardly imagine that you can do effective data governance without doing something about management. Thinks so is their requirement. But the footprint you need to do that. Have,

you can vary. For example, I can go through I have. Yes 5 ago Worked at a smaller financial institution. It was a fund. And asset manager, literally an asset manager, but. They did Yes fund management and they actually had a free time. Small data landscape, not many systems. About 15 to 20 systems. A number of data models and then a number of data flows and integration points. We still suffered from inconsistent definitions, quality issues, et cetera and actually did time to scratch a piece of governance set up quite small-scale with stakeholders, so we set up a data board. Are often also called a data governance Board purely to start the conversation about where are the problems? How can we? Better share our information and make it available. And then We Go Away started capturing a glossary. Whether that was called, then I believe. Anyway, the words catalog and glossary are also often used a bit interchangeably, but it also depends a bit on where you are. The context. So that catalog was basically a glossary of information about the definition. Where that data was located In the landscape. The Golden source eh so what is the the the system where that? Data is born. Any data cultural or standards that apply to that object of engagement. And, That was the list of no more than 120 objects actually, and also something along those lines and then once that too. 1 1 1. Responsible person were assigned, then you saw. That yes, very good synergy arose between the department to share knowledge and to ensure that you get more control and order in that data. So there it was. The proportion of metadata needed to achieve governance in a first iteration of maturity wasn't even that high, was it? We didn't have a car withdifficult connectors that pull in all the metadata and put it in model. That is actually Advanced users, which we also do with collibra at [COMPANYN NAME]. But in If It's Still In Its Infancy, you can actually get very far with very simple means. I have also seen certain solutions in sharepoint at very large companies that were actually linked to Excel files that were under governance and change management. To certain referential data and metadata. To manage and the single version of the truth, as it often so beautiful. Is that called you those in order? Gets, so it's possible. You sometimes don't even need very difficult resources and not so much to let go to one. Small goal to achieve." – F.MDRDG

"I think that's that, so you certainly have that for your critical data you have to get that really right, otherwise you can't steer on that. Can you too? Data quality yes, can you just not monitor and measure data quality? For the other data, it is particularly important to actually be able to provide the right dataset to the right person. Give with correct definitions on it." – B1.MDRDG

"Yes, I think metadata is very important, because if you don't have that, you can't do it. Yes, you can't be compliant either, I guess. Yes you can be kind of familiar with SQL yes, If you just say select star of the table and If you have that as a query and you're always going to pull in your data like that, then you just grab everything. And if there are changes, you just take them, but If you just give your metadata from. Well, I like this and this column. Then he only picks up the data you want, that you have defined and whatever you have agreed with. Yes with that. DLA or SLA. So That's it. Yes, you can't do without metadata, because. So dates are coming. To comply with that, that's just necessary, yes." – D.MDRDG

"Yes, yes, so crucial: Without metadata, you just can't apply your data governance properly, not or not properly? Yes. How else do you know the proportions?" – A2.MDRDG

Metadata Representation in Data Quality Understanding (mentioned by 8)

"Well and People who approach your such dataset can see what the quality is. From that set. And if that one, for example, eh? So if you put it very nicely on it, then you know how often is that refreshed? How, how new is it? Is this checked by someone or not on that sort of thing you get Of course that's actually the next step." – B1.MDRDQ

"Well, so it just belongs together. Yes, so you have Without that yes, you actually have, you need to know what yes what about what data elements do you just have? Those just need to be fixed with definition in a Dictionary and glossary Yes, you just need that to make the agreements. To be able to make. If that works. Really very much on the on the data on the data dictionary so data element level in base and now since recently also much more on the dataset. That would be me further added dataset Nobody. – B1.MDRDQ

"Yes, yes, so well, both I just have harder rules that you can set in a system to check and therefore indeed, because if your person just wants to look at the right, it makes sense somewhere that thanks to metadata you can get a lot more feeling at once to be able to look with the logical eye." – C.MDRDQ

"Yes, I I'm. This is, yeah, and and I'm all on the business side rather than on the technical side. Because I think that's sometimes missed in in some in many organizations it becomes too technical and people forget to actually understand, OK, what is it? What are we actually trying to do?" A3.MDRDQ

"I think it it then having metadata is an enabler to show where your data quality may be. But data quality, if it's if data quality is measuring actual data metadata is not. It's it's not, it's not actual data, it's fields a schema as it's it's it's something so. From a bit from maybe from. A business rule perspective, it helps, but it's an enabler, I think for data quality." – A3.MDRDQ

"For from the business, if you. Want to look at the metadata. Also has been a glossary of terms. Do we see do we call terms a metadata it's business metadata. I know what this means. Then that also becomes very important. So, and again, it helps I think for for a common language to be established. If you have the right level of of. If you know exactly how you want to capture your metadata that you'll do it from the top down or a bottom-up approach. That you can show you know and start to get that enterprise wide, harmonized uh, view on on what what your data is." – A3.MDRDQ

Managing quality of data (mentioned by 9)

"I think two things, I think. That the structure is very clear, so that you don't have to look up or clean a lot yourself that it's very stratford, how you put it together. Belonging together and

therefore a first one, eh? That it's just so well modeled in a clear way that it makes sense to follow, but to recover. But that's also things like, for example, very clear descriptions of fields you have of them. I know, this field is exactly this and this this and then that can't be long enough to make it very clear. And the other thing is simply the reliability of. The dates themselves so If I If I have a If I look at how order did I do yesterday? That's the right number, so to speak. Yes because both, because both really aren't trivial. Both really not easy to do." – C.MDRDQ

"I think that's very important to know whether you can trust something in data or not. I actually all end up trusting the Van de data in the end, because If you know that the data quality is too low, then you don't know if you. Or your analyses that you? It's right, isn't it? I think it is very important for a user that you know that your data quality is high, so that you have the metadata about your data. Where that, so she knows it's that quality very high. Low is," – B2.MDRDQ

"Yes, I think I have. I have just answered the current question. I thank now look also that previous question can answer. Representation of metadata and data quote context. Yes, of course, it starts with business Rolls. Yes, a functional logic and and standards including conventions and certain Mandatory. Values that sort of thing. Sometimes code values in reference data huh? You can only choose from these 5 values and otherwise have. Do you have a problem? So, I think quite. Very linked to to. You should actually display it so that it is for the. End user also very fast. This can be translated into daily practice. And, and that is sometimes quite difficult, because. If you put in a system like collibra and That's not collibra's fault. That have every tool that problem once you pull this within the walls of one. Application then you are also somewhat trapped in that application, aren't you? So and and People are people anyway, for example People who only have a SAP system werken Okay, they want their. Seeing things in an SAP system, huh? They want to be there, they are used to those screens, to that layout, to that search function, to the representation layer. Also the UX, huh? The user experience. If they then have to look for something in a collibra system completely different look and feel, different experience, different That is often a threshold and you can't do that easily. Brushing away, other than you? Still, it should take care of that. The use of such a tool as well, if only to look something up. Facilitates a lot possible, but also includes somewhat standard in your processes, because otherwise goes. It's not going to be that, is it? Huh, that? We all suffer from that like you, don't we? Your work is done in a certain system and That is today. Do you already need 30 systems to do your normal work? If someone takes you to another sharepoint or a teams environment or another application where you have to log in again where you then have to look again. Okay, where do thoseyou're standing for what, huh? At some point. At the moment, there is also a saturation point in People to have to learn something else, so you always suffer from that. Yes." – F.MDRDQ

"So, we still use Collibra. So, we go from business to conceptual and then down to logical. We can have data quality rules at a business term level, and we can then have them at a logical level, and then, of course, we apply them at a physical level, so we use Collibra mostly for the top 2 layers. And then, for the actual application of DQ rules, we were using a tool called Trillium,

and we're now using the Atacama. And so yeah, that then that's this Metadata captured throughout." – A1.MDRDQ

"And the metadata example there is well, how to do your present data quality? Is it? Is it one out of 10, or is it 99%? Or you know. We do that. That's a challenge that's really difficult because if we just present one number, then, well, how do you trust that number? So, we actually break it down into the concepts, the data concepts. So, like the lease contracts and the customer and then we do it per entity as well. So, Brazil. Contract 75%, right? And then how do you get to that 70%? Is it an average of? And so yeah, so that's, that's where the metadata comes into play with." – A1.MDRDQ

"Basis data element level what actually data really stop, data element level, But we. What you've always had is that that's actually kind of because of those toolsets. In the end, is there anyone? Does someone send a dataset outside or use it and it often contains data elements from multiple parties. So we also have one now. Dataset owner actually has governance, but he is actually responsible for data quality. For the delivery of that set and the. What is it used for? And that is based on the data quality behind it, You have to know on that set, can you what is the purpose of that dataset? Because that is actually part of the requested data quality." – B1.MDRDQ

"yes, look, I can get me, so to speak. Having my data scattered for all loose Excel files that all just don't quite match or just over and then all say something different, but it's accurate and can be reliable and things like that and can be complete. Only that you that it then Maybe not quite logical how you then combine data and If you want to come to an insight, say Because you have it all scattered, or that kind of thing. And So I think it's in one place and all. All the structures how everything tells itself to each other is all links on so all. It is very clear that that is also a very big, very big important thing to keep quality good, otherwise you will make a mistake faster, so it is nice that your coron. Data is correct, but it still is. For business, you make a mistake." – C.MDRDQ

"Yes, that's the one, isn't it? Important because you measure You physically though, But do that. You with a. My logical reasoning Why you want that and so you first just say, I want this data to be ok for such and such and that reason that I am going to measure in this and this way. Then you say, you put those rules on top of those on that physical data. But your physical data is of course presented by that metadata. Then you can say well, this part of my metadata model is data quality checked, so that's okay, so I can do that. Can I give a check mark of this data has been verified and it is correct so it is all Related. So yes, it is, it's very important, yes. At that level, you want to see where your reliability data is from." - A2.MDRDQ

"Well, well, what's the execution, so to speak, so how are you going to make sure that the data quality is higher or stays high? Get better or something, that You have to clean. Yes, you have to. Yes, you have to make your own processes for that. You may get In the assignment of yes make sure it is good, but In practice you also have to define that yourself. From, how do we go?

Maybe we'll run into things like oh, this didn't really realize it was all active. Yes, Maybe we should make something for that? In completeness, for example, empty tables and think, Oh, that's not right." – D.MDRDQ

"So yes, in terms of process? Yes, We are busy a lot, That is more Perhaps from the automation idea In the beginning. When we were building there were raids and there wasn't much yet. Data In the. But then we ended up just building those pipelines and then we just started throwing everything in like that. We really had data quality checks as well. Yes, That was In the beginning that's okay, but eventually there are also just users and those brands we all call that to a nomaly or something or things that don't look good and then you realize oh, that has to be time. The Quality checks also have to be built and some are very clear and some less so. Yes, we are in the process. Now try to focus a little more on that. In the beginning it's more of oh, We have to bring in sources and then after that we also have quality and also increase. But yes, Without data you can't say anything, but so. First promise of okay, this project will go. Yes, that means being able to unlock sources. So could be. Well, if you have a few sources in, then you can also focus on that. Work to automate or improve existing processes. Yes and That is also faster, goes example computing power and if so." – D.MDRDQ

"Yes, I think data quality should always be linked to business processes that need data and your data quality should be fit for purpose for that process. And then you often have to deal with the fact that you have a process to create the data. Think, for example, of an office book. This often happens in customer sales, service or customer organisation. And in addition, you often need that customer information to provide insight. In which markets will you supply and your planning processes? So the planning process is a completely different business process, but it depends on the data that is created somewhere. So your data quality isn't just in your silo, so me. I serve customers, so my customer data such as addresses I know, what should be right? No, you also have to make sure that the data for adjacent processes is also correct, so data quality always has the dimension of the consumer and. The producer who must have very clear insight into what the other person needs to manage the data well for you or to use the data properly." – E.MDRDQ

"It's also very broad. OK. Again, I think it's it's trying. to fit for purpose. Data needs to be fit for purpose. There needs to be a common. There should be commonalities. There should be enterprise wide. Uhm. Standards around quality around data in order to ensure that, yeah, everybody is speaking the same language. So, I think data quality plays a very big part in ensuring that that that we are have a common language across the business. Across any business. And it really needs to be coming from. You know from from business rule type driven, what are the business rules around that data and and. Translate that into kind of monitoring of of of the data." – A3.MDRDQ

"I think you think it's always a relative term. Is, But that that general? Definition could be that you? Make sure that the data meets the standards that you propose yourself. Those can be

different focus. Have, huh? You always have those. Data Quality Dimensions, eh, which you probably know too, huh? So? Accuracy, currency enn de completeness. Integrity well that kind of dimension and I think it always has two Axes. First of all, you can't bring everything under quality management, but you have to look at what are your key or critical data elements, eh? The dirt At all that really needs to be brought under control or that can cause problems if you don't. VAT. Number on your customer file, then you can't have an invoice. Send well, then you have a problem. So such a simple check through the number Or something that that's obvious, but so I guess we're looking at where. What does the what do the processes need? Make a selection of the data element that you want to bring under quality management and. We then look. According to which dimensions you want to exercise that control and also look at which layers of process and technology are already there. Certain validations take place or business rules are applied. Because often is. It's not that hard to? Check the data quality? Or the monitors and report on it, but to it. Correcting is difficult and having to prevent is also very difficult, because you never know in which layer. Of the whole process. Or the technology is anyone? Is capable of either or. Is authorized to adjust certain field values in a certain orientation or in. An interface or. There are a lot of hidden logics left. Which is very hard too. To be found is yes." – F.MDRDQ

"Yes, look, it obviously has common ground with it. Your metadata model, eh? The business and governance metadata information around business tools and data quality tools that you link to that. What do you want to apply those Rolls to? Logical and physical. And of course it has the advantage that you can set it up very once and reuse it many times, so you can play a data quality role. Managing and setting up. To then technically check in different physical systems, but you have determined at a point how. You do that? That does have the advantage that you are not going to do it in 20 places in 20 different ways, so I think the added value is mainly in. The central approach. Making it transparent and being able to link it to your underlying data. World so the models and the systems." - F.MDRDQ

Insights from data (mentioned by 9)

"And the metadata example there is well, how to do your present data quality? Is it? Is it one out of 10, or is it 99%? Or you know. We do that. That's a challenge that's really difficult because if we just present one number, then, well, how do you trust that number? So, we actually break it down into the concepts, the data concepts. So, like the lease contracts and the customer and then we do it per entity as well. So, Brazil. Contract 75%, right? And then how do you get to that 70%? Is it an average of? And so yeah, so that's, that's where the metadata comes into play with." – A1.MDRDQ

"Well, there are tools available within the section that have that. We don't have at least yet central. A central tool to do it. A moment later asks, but there are only going to be. There are, There are all kinds of Controls and main frameworks, with also data about that data are available,

so yes so that. And, we have there. That depends a bit. The part of the company off. So that's for one we have. That is now growing closer together. Well, almost what I said you have that that that that that quality consultation that in terms of time is discussed, so the one has a very nice one. Cockpit that quality cockpit, in which actually? Comes in where we can just see what data has been checked, how many times it's been checked that sort of thing, so It is. Yes, currently still built in different tools. The intention is that everything actually comes together. At some point." – B1.MDRDQ

"It's also very broad. OK. Again, I think it's it's trying. to fit for purpose. Data needs to be fit for purpose. There needs to be a common. There should be commonalities. There should be enterprise wide. Uhm. Standards around quality around data in order to ensure that, yeah, everybody is speaking the same language. So, I think data quality plays a very big part in ensuring that that that we are have a common language across the business. Across any business. And it really needs to be coming from. You know from from business rule type driven, what are the business rules around that data and and. Translate that into kind of monitoring of of of the data." – A3.MDRDQ

"For from the business, if you. Want to look at the metadata. Also has been a glossary of terms. Do we see do we call terms a metadata it's business metadata. I know what this means. Then that also becomes very important. So, and again, it helps I think for for a common language to be established. If you have the right level of of. If you know exactly how you want to capture your metadata that you'll do it from the top down or a bottom-up approach. That you can show you know and start to get that enterprise wide, harmonized uh, view on on what what your data is." – A3.MDRDQ

"We're we; we want to show the results of the tests of the DQ checks in COLLIBRA. But at the moment, we're using power bi. So, we take and extract the tool and pump it into the Power BI and then build Dashboards." – A1.MDRDQ

"Yes, well, the data or data in all kinds of monitoring, dashboards and insights that you can have about it, so That's actually the next step, so of course you can also steer on that meta data on how it actually is." – B1.MDRDQ

"Yes subjectively I would say, the extent to which data is useful. If you have high data quality, you can simply get more and more value out of it. And if the data quality is low unusable and now it has no purpose to have data. Say so black and white and? Yes, the higher the data quality, the better in principle, but. Yes and beyond that you are also just dependent on the source, so sometimes you go from. Yes, sometimes you also have crap in, crap out, so to speak. And then bad data comes in, then it is also just difficult to clean it. But if you want just? Also in a good way that data gets also the format can sometimes also help, then you can also do much more with it. So yes." – D.MDRDQ

"And you can represent that in kind of a in the visualization on a sort of lineage and visualization where you can, when a user can see exactly where DQ is being measured. And then it also shows the score. Of that and on the visualization, it shows that score." – A3.MDRDQ

"So that's very important, but we have not yet done is to put an alert you can we uh. It would be. The process the next step on the maturity level. Would be when, when, when that monitoring goes below that certain threshold. So, you know that those fields, tables, columns. Uh are are going below the quality of of what? Your completeness check is, then an alert would be sent out to to a Data steward. Or to a data owner to tell them and that there is an issue with your data and maturity. Again, level is ideally this should also be going up to the reporting line. If you have global reports report owner should know whether or not this report can be trusted. If there is. If he can see that he or she can see. That the the. The data quality score is good. Then he can they can these report owners can. Be confident on on their certification of their reports. So plays a very important role in in that sense." – A3.MDRDQ

Data Quality measure (mentioned by 8)

"Yes, I think I have. I have just answered the current question. I thank now look also that previous question can answer. Representation of metadata and data quote context. Yes, of course, it starts with business Rolls. Yes, a functional logic and and standards including conventions and certain Mandatory. Values that sort of thing. Sometimes code values in reference data huh? You can only choose from these 5 values and otherwise have. Do you have a problem? So, I think quite. Very linked to to. You should actually display it so that it is for the. End user also very fast. This can be translated into daily practice. And, and that is sometimes quite difficult, because. If you put in a system like collibra and That's not collibra's fault. That have every tool that problem once you pull this within the walls of one. Application then you are also somewhat trapped in that application, aren't you? So and and People are people anyway, for example People who only have a SAP system werken Okay, they want their. Seeing things in an SAP system, huh? They want to be there, they are used to those screens, to that layout, to that search function, to the representation layer. Also the UX, huh? The user experience. If they then have to look for something in a collibra system completely different look and feel, different experience, different That is often a threshold and you can't do that easily. Brushing away, other than you? Still, it should take care of that. The use of such a tool as well, if only to look something up. Facilitates a lot possible, but also includes somewhat standard in your processes, because otherwise goes. It's not going to be that, is it? Huh, that? We all suffer from that like you, don't we? Your work is done in a certain system and That is today. Do you already need 30 systems to do your normal work? If someone takes you to another sharepoint or a teams environment or another application where you have to log in again where you then have to look again. Okay, where do thoseyou're standing for what, huh? At some point. At the

moment, there is also a saturation point in People to have to learn something else, so you always suffer from that. Yes." – F.MDRDQ

"Yes, that's the one, isn't it? Important because you measure You physically though, But do that. You with a. My logical reasoning Why you want that and so you first just say, I want this data to be ok for such and such and that reason that I am going to measure in this and this way. Then you say, you put those rules on top of those on that physical data. But your physical data is of course presented by that metadata. Then you can say well, this part of my metadata model is data quality checked, so that's okay, so I can do that. Can I give a check mark of this data has been verified and it is correct so it is all Related. So yes, it is, it's very important, yes. At that level, you want to see where your reliability data is from." – A2.MDRDQ

"Data quality is yes the both completeness and accuracy of. You like your data. Is that that yes is that that data meets your expectations. And that it is correct? So, look carefully then. You expect it to be filled in Of course very important. Yes, you can have a table in which all data is correct, if only 10% is filled in. Yes, I still have. Nothing at all. You have the reliability yourself and that of course also depends on it. It's good If the 95% is reliable, it should be 100 or therein between. You just have the the notation itself of knocks, we use the right currencies, that kind of thing so it There are quite a lot of dimensions to the quality and normally you do of course by applying certain rules. Do you think certain logical data, quality rules? And I'm going to measure preferably as close. Possibly at the. source or whether it is actually correct." – A2.MDRDQ

8.8 Appendix VIII: Categories and themes to data structure

In the tables below can be seen the five high over categories of the different themes with their codes from the results. The themes were:

(i) Understanding;
(ii) (Meta)Data management;
(iii) Organization;
(iv) Data management tools; and
(v) Privacy.

The codes have been colorized for categorization purposes.

Table 14 - Category: Understanding

Theme	Participant ID	Codes
Understanding the data	A1; A2; A3; B1; B2; E	Clarity/Understanding: Confusion
Understanding	A1; B1; B2; C; D; E	Clarity/Understanding: Uncertainty
Understanding	A1; A3; B1; B2; C; D; E; F	Clarity/Understanding: Ambiguity
Lack of clarity	A3; B1; B2; C; D; E	Clarity/Understanding: Unclear
		Clarity/Understanding: Difficulty understanding
		Technology: Difficulty
		Personal and Organizational Development: Challenges
		Data Management: Data literacy
		Clarity/Understanding: Clarifying
		Business Management: Business glossary
		Data Management: Data definition
		Technology: Technical Jargon
		Technology: Technical constraints

Table 15 - Category: (Meta)data management

Theme	Participant ID	Codes
(Meta)Data Management	A1; A2; A3; B2; C; D; E ; F	Data Management: Data management
Metadata Management	A1; A2; A3; B1; C; D; E; F	Data Management: Metadata Management
Data Management and Quality	A1; A2; A3; B1; C; D; F	Data Management: Data modeling
Data Management	A1; A2; A3; B1; B2; C; D; E; F	Data Management: Data Quality
		Data Management: Importance of correct data definitions
		Data Management: Metadata cataloging
		Data Management: Data governance
		Technology: Technical requirements
		Data Quality: Data quality assurance
		Technology: Metadata
		Data Management: Data analysis
		Technology: Technical expertise
		Business Management: Business processes

		Personal and Organizational Development: Teamwork Personal and Organizational Development: Awareness Business Management: Business adoption

Table 16 - Category: Organization

Theme	Participant ID	Codes
Organization: Old VS New system Organizational management Recognition/Awareness	A1; A2; A3; B1; C; D; E; A1; A2; B1; B2; C; D; E; F A1; A2; A3; B1; B2; C; E	Technology: Organization Personal and Organizational Development: Awareness Personal and Organizational Development: Accountability Business Management: Organizational structure Business Management: Organization governance Data Management: Importance of metadata Clarity/Understanding: Clarifying Data Management: Recognition of metadata Personal and Organizational Development: Efficiency Business Management: Workflow processes Business Management: Business processes Technology: Legacy systems Data Management: Data capture Personal and Organizational Development: Compliance Business Management: Risk management

Table 17 - Category: Data management tools

Theme	Participant ID	Codes
Data Management and Tooling Managing quality of data Insights from data Data Quality measure Data Management: Tools Metadata type Data type	A1; A2; A3; B1; C; D; E A1; A2; A3; B1; B2; C; D; E; F A1; A2; A3; B1; B2; C; D; E; F; A1; A2 ;A3; B1; C; D; E; F A1; A2; B1; B2; C; D; E; F A1; A2; A3; B1; C; E ;F A2; A3; C; E; F	Technology: Programming Technology: Technical expertise Technology: Technology Data Management: Data cleaning Technology: AI Technology: APIs Data Management: Data modeling Data Management: Data capture Technology: Documentation Technology: Automation Technology: Integration Technology: Representation Technology: Metadata Technology: Technical metadata Data Management: Data privacy

		Technology: Tooling Data Management: Data Management Tools Data Quality: Data Quality Control Data Management: Data analysis Data Management: Data visualization Technology: Analysis
		Data Management: Data classification Data Management: Importance of accurate data Data Quality: Data quality assurance Data Management: Data processing Data Management: Importance of data quality Data Quality: Data quality assessment Data Quality: Validity Data Quality: Accuracy Data Quality: Consistency Data Quality: Reliability Data Management: Data Quality
		Data Management: Data management Business Management: Business processes Business Management: Organization governance Data Management: Data governance Technology: Organization Data Management: Metadata Management Technology: Hierarchy Data Management: Data privacy Business Management: Business metadata

Table 18 - Category: Privacy

Theme	Participant ID	Codes
Privacy	A3; B2; C; D	Data Management: Data privacy Technology: Privacy

Table 19 - From first order concepts to dimensions

1st order	aggregated	2nd order	dimension
Difficulty and confusion with understanding	Challenges and difficulties in understanding, including confusion with understanding, challenges in data literacy, and technical language.	Understanding challenges	Improving understanding
Challenges in data literacy	Importance of data management and governance, including the importance of data definitions, (meta)data management, awareness and adoption in the business, and recognizing metadata for effective data governance/management.	Business process improvement	

Importance of data definitions	Focus on improving business processes, capturing data in legacy systems, and complying with regulations.	Data management importance	
Technical language	Emphasis on the technical skills required for effective data management.	Stages of data management/governance	Data proficiency
The importance of (meta)data management	Recognition of different stages in data management/governance.	Effective data management	
Technical requirements to ensure data quality	Importance of data tooling in aiding data management.	Privacy in data management	
Business Processes & Teamwork	Focus on different aspects of data quality.	Technical skills	
Awareness and adoption in the business	Managing data effectively through data management practices.	Data tooling	Data Governance Lifecycle
Recognizing metadata and creating awareness for effective data governance/management	Consideration of privacy in data management.	Data quality aspects	
Improving business procesess			
Capturing data in legacy systems			
Complying with regulation			
Technical skills are required			
Different stages in data management/governance			
Data tooling helps with data management			
Focusing on different aspects of data quality			
Managing data effective through data management			
Privacy of data in data managent			