A Business Value Perspective on Metadata Management

SUBMITTED IN PARTIAL FULLFILLMENT FOR THE DEGREE OF MASTER OF SCIENCE

H.R. van de Beek
12415359

MASTER INFORMATION STUDIES
Information Systems

FACULTY OF SCIENCE
UNIVERSITY OF AMSTERDAM

28-06-2019

1st Examiner
Prof. dr. Paul Groth
Informatics Institute,
University of Amsterdam

2nd Examiner
Dr. Benedikt Kratz
Technology Strategy & Transformation,
Deloitte Consulting B.V.

ABSTRACT

The relevance of metadata management for organizations is rising due to the ever increasing amount of data that is being generated, technological advancements, and new regulatory requirements. Nevertheless, literature on the value of metadata management is scarce. This is unfortunate, since decision makers need to understand the value of metadata management in order to justify investments in metadata management initiatives. To increase the understanding of the business value of metadata management, a study was performed with metadata management experts from various industries. The findings reveal that metadata management indirectly contributes to 11 business benefits through the working of 7 functions that are enabled by metadata management. The results of the study are combined in a novel framework.

KEYWORDS

Metadata, Management, Business, Benefits, Value, Framework

1 INTRODUCTION AND MOTIVATION

Organizations are generating, consuming and storing data to a growing extent [21], an effect that is amplified by the increasing availability of open data sources [52, 53]. Combined with advancements in efficient data storage technologies, this can lead to the creation of data lakes: large collections of data, that often lack sufficient contextual information about their contents, origin and purpose [4, 21]. This contextual information is metadata. If this metadata is not captured, knowledge about data assets becomes implicit knowledge only carried by the teams involved in acquiring and creating these data assets, potentially leading to unnecessary duplication of work, missed opportunities, losses in productivity and mishandling of data [8, 21]. It is not surprising that metadata management is named as a prerequisite for managing data as an asset, and that without metadata, organizations might not be able to manage their data at all [22]. The emerging recognition of the relevance of metadata management is illustrated by the 2018 Gartner prediction that enterprises will double their investments in metadata management by 2020 [9].

Moreover, data legislation such as Solvency II [11], BCBS239 [33] and GDPR [31] forces organizations to be in control of their data. Metadata management enables organizations to identify private or sensitive data and is essential to manage the entire life cycle of data in order to minimize the risk of exposure, hereby contributing to meeting such regulatory requirements [22, 31].

However, organizations often struggle to implement data management initiatives and, in particular, find it difficult to establish a strong business case for data management initiatives [14, 28, 30]. The same difficulty exists for metadata management, considering metadata management is a crucial component of data management [22]. An explanation for the difficulty in establishing a strong business case could be that metadata management does not lead to direct benefits [13, 26]. While the cost of a metadata management initiative is simple to calculate, it is more difficult to quantify the contribution of metadata management initiatives to the business once metadata has been created and made available [27]. Moreover, metadata created for a specific purpose could prove to be beneficial in a different context, which makes establishing the causal effect of metadata management initiatives on business performance difficult [21, 27]. In addition, metadata is mainly addressed from a technical angle, while its contribution to the business is rarely addressed in literature [38, 40]. As a result, a body of research on the assessment and improvement of metadata quality exists, while research on the value of metadata is scarce and scattered [4, 27, 40].

This is unfortunate, since decision makers need to understand the business value of metadata management in order to justify investments in metadata management initiatives [27, 36, 39]. Strikingly, the most critical strategic success factors for implementing managed metadata environments are having a clear value proposition and committed management support and sponsorship [49].

Considering the rising importance of metadata management, the lack of studies addressing its value and the difficulties in establishing the business case for metadata management, the following research question is posed:

What are the business benefits of metadata management and how are these business benefits enabled?

By demonstrating the potential business benefits of metadata management, and by visualizing the result in a framework, this study aims to increase the understanding of the business value of metadata management. To achieve this, a grounded theory inspired multi-case study using semi-structured interviews and a Delphi technique session has been conducted. Interviews were conducted with metadata-related software vendors, business owners and consultants and clients of multinational consultancy firm Deloitte.

In the next section, related work on the concepts that are central to this study is treated. Hereafter, the methodology of the study is discussed. Next, the findings are presented. Furthermore, the work is discussed and directions for future research are given, after which the paper is concluded.

2 RELATED WORK

This section delineates the three central concepts to this study: Metadata, metadata management and business benefits. Metadata and metadata management are defined from an business perspective. Next, business benefits are explained and a business benefit framework that is adopted in the current study is discussed. Moreover, the relationship between metadata, metadata management and business benefits is discussed on a conceptual level.

2.1 Metadata

A commonly used definition of metadata is "data about data" [14, 41, 53]. While this definition is simple, it is not particularly helpful towards understanding the concept and function of metadata [14]. To create a better understanding, the concept metadata is discussed in this section.

One of the earliest descriptions of metadata stems from a paper from 1968, in which the term is defined as follows:

> *"...the ability to associate explicitly with a data element a second data element which represents data "about" the first data element. This second data element we might term a "metadata element". Examples of such metadata elements are: an identifier, a domain "prescriptor" which specifies from what domain the values of the first element must be taken, an access code which limits the conditions under which the first data element can be accessed." - [2, p.26]*

This definition implies that there are two parts to metadata: the original data, and the metadata that "represents" data about the original data. The metadata has a function regarding the data it represents, such as identification and the enforcement of access control.

Within the business domain, metadata is usually categorized in three categories: business, technical and operational metadata [7, 22]. *Business metadata* describes information about the content and condition of the data it relates to. Examples are definitions, business rules, applicable algorithms and ownership information. *Technical metadata* describes technical data details. Per example, in which systems data is stored and how the data flows through different systems. *Operational metadata* provides detailed information on the accessing and processing of data, such as access logs, modification history and system response times. Notably, these categories relate more to the place where the metadata originates from than to where it is used. The distinction between the categories is not strict when it comes to usage [22, 41]. The business metadata category illustrates that metadata does not only concern technical system details and operational outputs, since business metadata is concerned with the administration of more human understandable information about data assets.

Essentially, metadata helps to answer five questions about data in an organization: What do we have? What does it mean? Where is it? How did it get there? and how do we get it? [45]. Stvilia et al. [42] put emphasis on the "how did it get there" question by stating that metadata ideally explicates who, or what, did something to data, when, where, how, and why.

A slightly adapted definition of metadata established by Greenberg [18, p.1876] covers metadata as discussed in this section well:

Metadata is data about an object that supports functions associated with the designated object

The term "structured data" from the original definition was changed to "data", since business metadata includes unstructured textual descriptions, such as definitions and textual descriptions of data sets. The "object" in this definition can be any information bearing entity, which implies that metadata can also describe objects other than data, such as systems, algorithms and business rules. Furthermore, this definition implies that it possible to do something with metadata, since it enables functions related to the object it is associated with.

While creating a single comprehensive definition for metadata remains a subject for philosophical discussions, this definition is deemed sufficiently complete and concise, and will therefore be adopted in this study.

2.2 Metadata management

While metadata is defined in the previous section, it remains unclear how metadata is acquired, created or curated. In order to provide an answer, various perspectives on metadata management are discussed in this section.

According to Huner, Otto and Osterle [24], the goal of metadata management is to have comprehensive, unambiguous and understandable metadata, and to achieve this, knowledge on different business objects emerging from various departments should be identified, discussed, consolidated and finally recorded. Business experts from these departments should be involved in this process. This definition indicates that metadata requires a management approach that encompasses various departments and involves business users in the process.

Actions	**Area of Governance**	**Decision Domain**
Define	Roles and responsibilities	Metadata
Implement	Policies	
Monitor	Processes and Procedures	
	Standards	
	Strategy	
	Technologies	
	Guidelines	
	Requirements	

Table 1: Metadata management activities, which are a combination of an *action* plus an *area of governance* plus the *decision domain* metadata (adopted from [1])

Thangarathinam et al. [46] state that metadata should be identified and defined through repeatable processes, captured in a standardized manner and governed through approved processes in order to ensure correct, available and credible metadata that eventually leads to business benefits. Interestingly, this definition states that metadata should not only be defined, but also identified, hereby implying that metadata management does not only concern the creation of new metadata, but also identification of existing metadata within the organization. This is confirmed by Roszkiewicz [34], who states that the first step towards creating a corporate metadata taxonomy should be identifying all systems that contain metadata. Hence, metadata is not only an output of metadata management, but also an input.

The Data Management Association defines metadata management as *"planning, implementation, and control activities to enable access to high quality, integrated metadata"* [22, p.419]. The activities referred to in this definition are: (1) Defining a metadata strategy, (2) understanding the metadata requirements from a business and a technical perspective, (3) defining a metadata architecture by creating a metadata model, applying metadata standards and managing the storage of metadata, (4) creating and maintaining metadata, including the delivery, integration and distribution of it, and finally (5) querying, reporting and analyzing the metadata. The activities imply that metadata management is also concerned with making the metadata accessible and extracting insights from the metadata. Moreover, the decision domains of metadata management encompass strategic decision making, technical implementation and operational use. Hereby, this definition implies that metadata management is a holistic approach concerned with all decisions and activities concerning metadata.

A recent literature review on data governance activities by Alhassan, Sammon and Daly [1] consolidates information found in 61 scientific and practice-oriented sources and concludes that all data governance activities can be defined as an *action* plus an *area of governance* plus a *decision domain*. One of these decision domains is metadata. Therefore, all *actions* plus *areas of governance* in the *decision domain* of metadata are considered to be metadata management activities. An overview of these metadata management activities can be found in Table 1. It could be argued that the terms

"management" and "governance" differ. Sadiq et al. [35, p.96] argue that data management concerns making decisions and implementing them, whereas data governance determines which decisions are to be made and who should be involved in the decision making process. However, Alhassan, Sammon and Daly [1] include *implement* as an *action*, hereby encompassing both the decision domains of data management and data governance.

Since no widely accepted definition for metadata management was found in academic literature nor the practitioners' community, a definition for metadata management based on the different perspectives discussed in this section is posed:

Metadata management is an organization-wide approach concerned with defining, implementing and monitoring all metadata related initiatives, with the goal of achieving accessible fit for purpose metadata that leads to business benefits

This definition emphasizes that metadata management is not an initiative reserved for the technical department only and should be approached from an organization-wide perspective: Metadata management is concerned with all initiatives that create metadata, improve metadata, enhance the accessibility of metadata or that utilize metadata to enable business benefits.

2.3 Business benefits and metadata management

To attribute business benefits to metadata management, it is essential to delineate the concept of business benefits and understand the connection between metadata management and business benefits on a conceptual level.

Business benefits in the IT domain have been approached from a tangible perspective, where they are quantified as measurable performance improvements such as time and money saved [3, 43]. However, Brynjolffson and Hitt [5] argue that tangible business benefits only explain a part of the business benefits that can be achieved through IT. In line with this statement, business benefits have also been approached from a qualitative perspective, where organizational performance and intangible business benefits such as decision making and information accessibility are emphasized [5, 29, 37]. Tangible and intangible IT business benefits affect different levels of the organization: tangible business benefits of IT are realized at the tactical and operational level, while intangible business benefits occur at the strategic level [25].

To structure business benefits, a highly-cited business benefit framework developed by Shang and Seddon [37] is adopted in this study. This framework covers tangible and intangible business benefits and considers the notion that business benefits can occur at different organizational levels. Furthermore, it has been developed to structure and explain business benefits in the IT domain, which makes it relevant to the goal this study. Moreover, it has recently been adopted to structure the business benefits of related topics, such as enterprise data models [32] and data governance [48], which increases the external validity of the framework. Shang and Seddon [37] distinguish between operational, managerial, strategic, IT infrastructure and organizational business benefits. These categories are individually explained in section 4.3, accompanied with the business benefits corresponding to the categories that were discovered in this study.

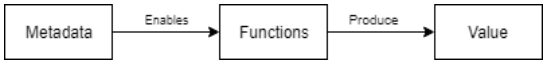


Figure 1: Diagram depicting the indirect relationship between metadata and value (adopted from [27])

The question as to how metadata management enables business benefits remains. According to Lytras and Sicilia [26, 27], the assessment of the value of metadata is a challenging affair since it is dependent on the wide variety of *functions* the metadata can enable. This statement implies that metadata itself does not directly create value, since the value is created through the *functions* that are enabled by metadata. Interestingly, this finding is in line with a study that addressed the value of Enterprise Architecture (EA), which concluded that the claimed benefits were *"too distant to clearly explain how these benefits can be achieved"* and therefore focused on benefit enablers: *"factors which could be clearly seen as EA outcomes, and that in turn are known to have the potential to deliver benefits"* [44, p.144]. The benefit enablers in this example are similar to the functions enabled by metadata. The existence of this indirect relationship between metadata and business benefits is confirmed in literature. Per example, Stock and Winter [40] assess the value of metadata by studying the implementation of a business glossary, which could be seen as a *function* enabled by business metadata. Another example is given by Bilalli et al. [4], who argue that metadata can enhance end-user business understanding by presenting information about the source of the data and the journey it has undertaken to arrive at the current state or location, which is called lineage. In this case, metadata enables the *function* lineage which leads to the *business benefit* business understanding. This indirect relationship is depicted in Figure 1 and considered in the current study to explain how metadata management can enable organizations to achieve business benefits.

Synthesizing the aforementioned insights, the following definition for the business benefits of metadata management is posed:

The tangible and intangible positive outcomes for an organization, enabled by functions that could be clearly seen as metadata management outcomes.

3 METHODOLOGY

Considering the limited availability of existing work on the business benefits of metadata management, the current study demands an inductive and exploratory approach [12]. A methodology that supports such approach while also stressing academic rigor, is the Gioia methodology [17]. This methodology is a variation on the more traditional grounded theory approach and stresses concept discovery and theory building.

The study was conducted during an internship at Deloitte: a multinational professional services network that provided access to a network of experts which could be approached for interviews and expert opinions. This setting fits the Gioia methodology, since it encourages investigating multiple cases by indicating that to understand a phenomenon, it is best to consider a broad array of perspectives [16]. Moreover, the preferred method of data collection is through semi-structured interviews [17]. While adopting a multi-case approach could negatively impact the internal validity of this

study, opting for a single-case approach at a firm that exclusively performs case-based work would defy the goal of a single-case study. Moreover, a multi-case approach does positively impact the external validity of the findings.

To allow for the discovery of new insights, the Gioia methodology advises to postpone judgment about the conclusions of the existing literature. Nonetheless, the Gioia methodology also stresses that knowing, and not knowing, should be finely balanced to prevent reinventing the wheel while attempting to discover new concepts [17]. Furthermore, it is encouraged to use the method flexibly to create a fit with the aim of the study [16]. Therefore, the concepts metadata, metadata management and their conceptual relationship to business benefits were investigated through a literature review prior to conducting the interviews.

3.1 Data collection

To establish a theoretical foundation and prepare for the interviews, a literature review following the guidelines of Wolfswinkel, Furtmueller & Wilderom [51] was conducted. This literature review method aims to create an accurate concept-centric review, which helps to synthesize the existing perspectives of a variety of authors concerning a concept [50]. All considerations regarding the literature review can be found in Appendix A. Appendix B contains the conceptual framework that was established based on the literature review. Next, semi-structured interviews were conducted with experts that had multiple years of professional experience with metadata related projects. Twelve interviews were planned based on the finding that theoretical saturation, meaning that additional interviews lead to no new information for the researcher, generally occurs within twelve interviews [20]. Two interviewees decided to give an interview together, resulting in a final number of 11 interviews and a total of 12 interviewees. Interviews were conducted in person and through meeting software (e.g. Skype™, Zoom™). Diplomacy, discretion and transparency are fundamental in a study involving informants [17]. Therefore, interviewees were informed about the goal of the study, the structure of the interview and asked for consent to record and process the interview with the promise to anonymize the results.

Semi-structured interviews use an interview protocol with a prearranged set of open-ended questions and allow the researcher to ask additional in-depth questions on topics emerging from the dialogue [10]. Following the guidelines by Castillo-Montoya [6], an inquiry-based interview protocol aligned with the research question was developed. To improve the quality of the interview protocol, feedback on the initial version was requested, after which the interview protocol was piloted, before continuing with the actual interviews.

The interview was deliberately split in two parts with regard to postponing judgment about the conclusions of the existing literature prescribed by the Gioia methodology [17]. In the first part of the interview, interviewees were inquired about metadata management and its value without providing any context. This open way of questioning was adopted to increase the chance of discovering new insights [6, 17]. The second part of the interview leveraged the theoretical foundation established earlier in the study. In this part, the five business benefit dimensions of Shang and Seddon [37] were explained to the interviewees to support their understanding of business benefits and to help them in rationalizing how metadata management can possibly contribute to the different categories of business benefits. To be fully exhaustive and prevent bias, interviewees were also asked if they could think of any business benefits of metadata management that could not be categorized in the existing five dimensions.

The interview protocol can be found in Appendix D. An anonymized list with information about the interviewees and the duration of the interviews can be found in Appendix 5.

3.2 Data analysis and validation

All interview were recorded and fully transcribed. The transcriptions were coded and analyzed in three rounds as prescribed by the Gioia methodology [17]. Qualitative data analysis tool Atlas.ti™ was used for coding and analysis. In the first round, open coding based on informant-centric terms was applied, resulting in 125 unique first-order codes. The open codes were initially assessed to exclude duplicates and irrelevant codes and then split into two networks of functions (51 codes assigned 414 times) and business benefits (60 codes assigned 587 times). Next, the first-order codes within the two networks were assessed on similarity and interdependency in order to sort them into second-order themes. As a result, 53 business benefit codes remained, which were assigned to 20 second-order code groups. For the functions, 42 codes remained, which were divided into 11 second-order code groups. To distill aggregate dimensions, the second-order code groups where thoroughly analyzed by manually revisiting all codes within each group. Codes were reassigned, second-order code groups were renamed and finally aggregate dimensions were distilled, resulting in 13 business benefits and 7 functions. The business benefits were then categorized along Shang and Seddons [37] five business benefit dimensions.

To increase the validity of the results, a Delphi technique session was organized [23]. During the Delphi technique session, 6 experts, of which 4 participated in the interviews and 2 were new participants, were asked to discuss the functions, business benefits and their relationship until a consensus among the participants was reached. Functions and and business benefits could be extended, removed, renamed or validated during the session. During the session, two business benefits were removed and multiple functions and business benefits were renamed. A common technique using in a Delphi session is the ranking of concepts [23]. In the case of this study, this yielded no consensus, which is possibly attributable to the multi-case design with different actors who all had unique needs and desires. Furthermore, no consensus on coupling specific functions to specific business benefits was reached. After the Delphi session, the codes were reassessed to see whether the outcome of the Delphi session could be justified in the coding scheme.

Information about the participants of the Delphi session and a summary of the results can be found in Appendix E. Furthermore, a visual representation of all intermediate steps of the coding process, accompanied with tables representing the final coding scheme, including frequency of code assignment and occurrence in the interviews, can be found in Appendix F.

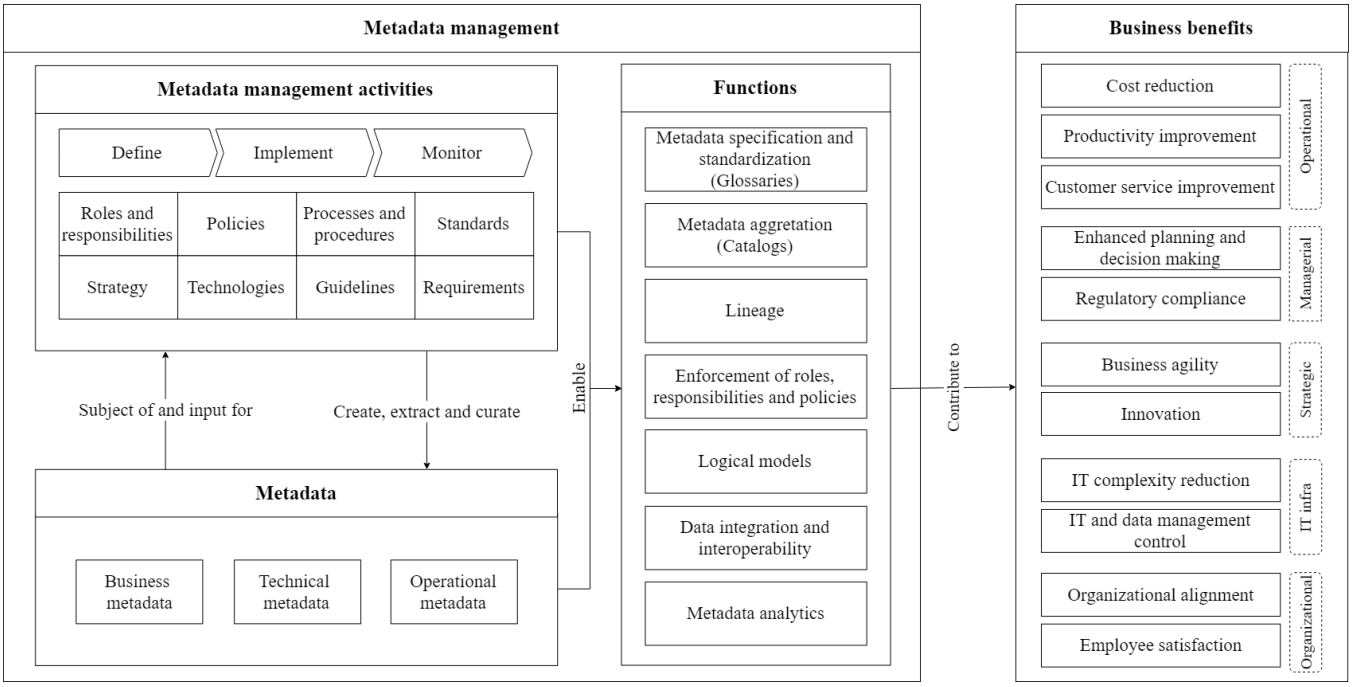


Figure 2: Metadata management benefit framework

4 RESULTS

This section presents the results of this study. The results are visualized in the form of a framework (see Figure 2). This section starts by elaborating the structure of the framework. Hereafter, the contents of the framework are discussed. Hereby, this section provides an answer to the question: *What are the business benefits of metadata management and how can they be achieved?*

4.1 Metadata management benefit framework

The structure of the metadata management benefit framework was grounded in literature and completed based on the data gathered through the interviews. The framework consists two large blocks: metadata management and business benefits. Metadata management contains metadata management activities, metadata and functions. Metadata management activities are the activities of Alhassan, Sammon and Daly. [1], which embody all activities related to metadata management initiatives. Metadata concerns the actual metadata that is managed and contains the three metadata categories that are used in business environments, which are business, technical and operational metadata [7, 22]. Metadata and metadata management activities have a bidirectional relationship because metadata is the subject of, and input for metadata management activities, and metadata management activities aim to create, extract and curate metadata to achieve accessible fit for purpose metadata. Section 2 concluded that the relationship between metadata and business benefits is indirect, since metadata enables functions that eventually lead to business benefits [27, 44]. However, metadata has to be created, extracted, curated and utilized to enable functions. Therefore, metadata management activities and metadata together enable functions in the framework. The functions are part of metadata management, since enabling the functions is also a metadata management initiative that is affected by metadata management activities. Therefore, metadata management activities, metadata and functions all belong to overarching block of metadata management. The functions that are the outcome of metadata management contribute to business benefits. The business benefits are structured according to the five business benefit dimensions of Shang and Seddon [37]. Hereafter, the 7 functions that are enabled by metadata management will be elaborated. Next, the 11 business benefits where the functions contribute to will be discussed.

4.2 Functions

Analysis of the interviews confirms that metadata can enable new functions, or increase the performance of existing functions when it is managed and utilized. The functions that were found in this study either consist of metadata or are enabled by metadata, making them clear outcomes of metadata management. Essentially, the functions are the things that are enabled when metadata is utilized after it has been created, extracted and curated. Table 7 depicts the codes that were assigned to each function, including information about the frequency of assignment and occurrence in interviews.

Metadata specification and standardization (Glossaries). Definitions for commonly used business terms, data elements and standard measures can be specified, standardized and captured as metadata [7]. A repository that contains this metadata is referred to by interviewees as a glossary. The glossary is a reference tool that enables a common understanding about the business terms, data elements and measures that are defined in it.

Metadata aggregation (Catalogs). Available metadata about data assets and systems can be aggregated and published in a central location. Interviewees refer to these aggregations of metadata as catalogs. Employees can use catalogs to document and discover all known data assets in the organization, based on metadata. Catalogs function as search-engines that contain information about the available data assets and systems in an organization with varying levels of detail, depending on the available metadata.

Lineage. Metadata can be tracked to reveals the complete journey of data from the original source to the final target destination [4, 7, 15]. Interviewees refer to this function as lineage. Various types of lineage were reported. Backward or reverse lineage starts at a point of data usage (e.g. a report, a warehouse) and reveals where the data has originated from and what has happened to it along the way (e.g. transformation logic, cleansing, merging, applied algorithms) to arrive at its current state. Vice versa, impact analysis or forward lineage starts at a data source and reveals what other systems depend on this source. In other words, it reveals what else will be impacted if a change is made to a source. Lineage can be revealed with various levels of granularity. For example, how systems interact on a metadata level (e.g. table a and b > algorithm c > report d), which was referred to as horizontal lineage and logical lineage. However, at the most detailed level, lineage can reveal the exact transformation details of data in each part of its journey from source system to target location, which was referred to as physical lineage.

Last, vertical lineage or semantical lineage reveals how a business requirement is translated towards the data model. For example, how a performance indicator is calculated, which data entities and attributes are used for that, in which tables these are stored, and in which databases these tables can be found. Interviewees commented that this type of lineage is usually recorded in the glossary that was discussed earlier.

Enforce roles, responsibilities and policies. Interviewees considered metadata valuable in the process of designing roles, responsibilities and policies related to data assets. First of all, in order to make deliberate choices on who should be responsible or accountable for a data asset, or what policies should apply to it, it helps if proper information on the data asset is available (e.g. definition, source). At the same time, roles, responsibilities and policies that apply to data assets can be stored as metadata. For example, metadata can be used to store information about who is accountable for the contents of a data asset (e.g. ownership), who are responsible for keeping the data up to date (e.g. data stewards), and what policies apply to the data (e.g. access restrictions, regulatory constraints). This information is helpful

Logical models. Interviewees illustrated that metadata is used to logically model relationships between data assets. For example, when data is not named consistently across different systems (e.g. customer, client, buyer), metadata can be used to link these data elements. These logical models were also referred to as semantic frameworks, taxonomies and ontologies. Interviewees commented that logical relationships are usually communicated through catalogs and glossaries.

Data integration and interoperability. The availability of proper metadata is considered to enhance the capacity for data integration and to increase the interoperability of systems, sometimes even enabling automated data exchange. Interviewees indicated that to enable the translation of data from a source system to a consumable format for a target system, detailed technical metadata is required to understand differences between the data from the source system and the data from a target system (e.g. data formats, table structures).

Metadata analytics. Considering metadata is a type of data, it can used for analysis. Lineage is a specific type of metadata analysis, as it tracks technical and operational metadata to reveal how data flows through systems. In addition, several other types of metadata analysis exist. For example, operational metadata (e.g. access and update logs) can be analyzed to reveal usage statistics and update information for data assets and systems. Interviewees indicated that such statistics can be valuable data quality indicators. Moreover, interviewees suggested that when an organization starts assigning roles, responsibilities and policies to data assets and systems, metadata can be analyzed to reveal which data assets and systems have no roles, responsibilities and policies assigned.

4.3 Business benefits

This section explains the business benefits that can be achieved by the functions that are enabled by metadata management. Where possible, specific functions are coupled to the business benefits. The business benefits are structured along Shang and Seddon's five business benefit categories and follow the order of the benefits in that framework [37]. It was chosen to do so, since no consensus was reached about the importance of individual benefits.

Table 8 contains an overview of the codes that were assigned to each business benefit, how frequently each individual code was assigned, and in how many interviews codes belonging to each business benefit were mentioned.

4.3.1 *Operational business benefit category.*

Operational business benefits have an impact on *"day-to-day activities that involve acquiring and consuming resources"* [37, p.278]. Operational activities acquire and transform resources, such as data, into products, services or results. *Operational cost reduction*, *productivity improvement* and *customer service improvement* were identified as operational business benefits in this study. These business benefits will be elaborated in the following paragraphs.

Operational cost reduction. This category refers to operational cost reductions that can be expressed in amount of money saved. Interviewees suggested that large and complex organizations often lose the overview of data assets and systems that are being paid for, which results in redundancy. Providing insights in what data assets and systems are available in the organization is an essential step towards decreasing unnecessary operational costs caused by redundancy. As interviewee 11a stated:

> *I think one big successes that we have already made, is again, trying to put things in one place and creating transparency about external data sources. It's about knowing what data do you have and where do you have them. [..] We have already been able to save the company €1.3 million by doing that.*

The findings suggest that catalogs can provide an overview of all data assets and systems that are available across the organization. Furthermore, glossaries provide insights in the contents and the

functionality of these data assets and systems. When the aggregated metadata is analyzed, the result can reveal redundant data assets and non-optimal system configurations. Moreover, interviewees implied that less employees are required for data migration and integration related tasks as a result of the improved data integration and interoperability function. Interviewee 1 declared:

> *With good metadata management in place [..] you don't need 10 users sitting at the terminals anymore, maybe you only need 4 guys.*

Productivity improvement. Productivity improvements imply that employees can do more work using the same resources. Productivity improvements can eventually lead to cost reductions. However, they differ from cost reductions in that they are expressed in time saved instead of money saved. Interviewees expressed that having proper metadata available can reduce time spent on searching for data. This is achieved through a variety of functions. For example, catalogs provide a searchable overview of all available data assets and systems, glossaries provide insights in the meaning of existing data assets, logical models make it simpler to discover related data assets and clear roles, responsibilities and policies make it insightful for what purpose data assets can be used and who should be approached to get access. Interviewee 11a expressed:

> *We have a lot of data in our company and we want to make use of it. But, if you don't even know what data do you have, where they are, who can access it, then you basically don't know where to start. You can explore, but it's blindly exploring, it takes a lot of time, a lot of effort, it's very inefficient.*

Interestingly, most tasks that interviewees referred to were existing data related tasks, such as mapping lineage, data integration, creating business intelligence reports, and doing data analysis. The findings indicate that these tasks require less manual work, are less prone to errors and can generally be performed more efficiently when proper metadata is available. In the following quote, interviewee 4 the effort required for mapping lineage when no proper metadata is available:

> *The fact that [lineage] information is not available leads to a lot of extra work with internal projects. When you start a project, you are spending the first 3 weeks on impact analysis, if not more. What happens when I adjust this system? Preferably, that kind of information is collected quicker and made reusable.*

Furthermore, interviewees suggested that having an overview of reusable data assets in a catalog can lead to a reduction in unnecessary overwork. The findings indicate that all functions contribute to productivity improvements.

Customer service improvement. Interviewees indicated that the availability of proper metadata can result in improved customer service by enabling quicker resolution of customer related issues and better service towards customers. The functions lineage, catalogs and having a glossary with clearly defined business definitions were most attributed to this category. Customer service improvement is interrelated with other business benefits such as *productivity improvement*, *IT complexity reduction* and *innovation*, since these have the potential to speed up issue resolution or delivery to customers. Interviewee 7 explained how having lineage and ownership information available contributed to customer service improvement:

> *We had pretty good business metadata at our disposal [..] and when people rang the customer service about issues on the website, we could very quickly locate, on the basis of that business metadata, where things had gone wrong and who should be contacted to resolve the issue. [..] In other words, we had a business form of lineage available. Website, process and information were made insightful. That helped to quickly achieve improvements.*

4.3.2 *Managerial business benefit category.*

Managerial business benefits have an impact on activities that are concerned with *"allocating and controlling the firm's resources, monitoring of operations and supporting of business strategic decisions"* [37, p.278]. Business benefits in the managerial category ensure that the organization can perform its operations smoothly. *Enhanced planning and decision making* and *regulatory compliance* are the managerial business benefits that were identified in this study.

Enhanced planning and decision making. The findings suggest that catalogs and glossaries make data in the organization generally more accessible and understandable, which enhances the ability to use data to support decision making processes.

Furthermore, interviewees expressed that to manage resources adequately, or do any type of portfolio-steering, an overview of available resources is required. Catalogs, glossaries and logical relationships can help in providing this overview. For example, if a product is being sold under different brand names in different countries, this information can be defined in a glossary. The sales data for these products can then be logically related to get an overview of the total sales. Metadata about the resulting data set can be documented in the glossary to make the data set reusable for future decision making.

Moreover, metadata can provide contextual insights in the meaning of data on which decision makers have to base their decisions. Interviewees indicated that a glossary can aid in understanding the definition of data that is being presented. Moreover, lineage was attributed to providing transparency about how the data that is being presented arrived at its current state, which makes the data more trustable. Last, metadata analysis can reveal contextual information about data assets that is valuable to decision making. For example, data assets that have not been updated for a long time, might not be reliable enough to base decisions on.

The findings indicate that glossaries, catalogs, lineage, logical relationships and metadata analytics together can lead to more substantiated data-driven decision making. Interviewee 6 illustrated how having clearly defined definitions in a glossary contributes to decision making:

> *When I'm looking for how many sales I have done in a certain region, it is not clear to me what a region is. Is that a state? Is that a council? Is that a city? Is that an area? What is it? A region may be different in different continents and in different countries. So without clear definitions, or granularity of definitions, it is very, very difficult. That could lead to getting wrong data, in which I then need to trust in order to make business decisions which may be vital to the company.*

Regulatory compliance. Interviewees reported that metadata managemend is required to become compliant with new data regulations. For example, GDPR requires organizations to record what customer data they have, where it is stored, who has access to it and what it is used for [31]. Moreover, BCBS239 forces organizations in the financial domain to have clear definitions for all measures in risk reports and requires them to be able to show these measures were "aggregated", which implies they have to show the full

lineage from the source systems to the eventual report [33]. Following the requirements of these regulations, the functions lineage, catalogs, glossaries, and enforcement of roles, responsibilities and policies contribute to becoming regulatory compliant. Furthermore, interviewees indicated that automating data related tasks reduces regulatory risk by eliminating the chance of human errors. Interviewee 10 clearly illustrates how lineage and roles, responsibilities and policies contribute to regulatory compliance:

> *From different angles, such as the GDPR and BCBS239, we are required to show our lineage. [..] We deliver critical reports to the regulator and their response is: can you show me where that data came from? What path it has travelled within your organization? But perhaps also whether it is external data? What have you done with that? What transformations have you applied to it? Who are the owners of the data, who has signed off? That is all metadata. You name it, and regulators require us to do it.*

4.3.3 *Strategic business benefit category.*

Strategic business benefits have an impact on strategic future planning and the overall direction of an organization. Activities on the strategic level are concerned with *"long-range planning regarding high-level decisions, such as business merging and acquisition, marketing competition, product planning, customer retention and capital sourcing"* [37, p.278]. Strategic business benefits that were identified in this study are *business agility* and *innovation.*

Business agility. Business agility entails the ability to adequately respond to unpredictable internal and external changes [47]. Changes mentioned by the interviewees where metadata management can contribute to are the need to integrate new external data, the need to merge data and systems during the merging and acquisition of organizations, the need to open up parts of the organizations' data and systems to external services, and changing regulations that require the delivery of new data or reports. Moreover, when an internal change has to be made to systems, being able to oversee how that change impacts existing systems through lineage was commonly expressed an ability that improves such processes. Interviewee 11a explained that having metadata available is helpful when rapid changes have to be implemented:

> *To do things quickly, we need to have metadata in place that gives the analyst power to know exactly what they're doing, what the business concepts mean, what the IT systems mean, how they are all linked to each other. That for me is one of the big advantages.*

The findings suggest that all functions contribute to business agility. A common view was that whatever changes have to be made to data assets or systems in the organization, all additional information that is available helps.

Innovation. Interviewees expressed that having a clear overview of what data is available, what that data means and what it can be used for, can improve the ability to use that data to innovate. Similar to the planning and decision making category, interviewees suggested that having an overview of existing products and customers could enable more quicker and more informed decision making regarding product innovation for new and existing customers. Another suggestion given was that having clearly defined roles and responsibilities regarding data assets will increase data awareness and understanding in the organization, which could encourage employees to use data to innovate. Hereby, the functions catalogs, glossaries and enforcement of roles, responsibilities and policies contribute to innovation

The following quote by interviewee 3 indicates that knowing what data is available and what the data means, which is enabled by glossaries and catalogs, contributes to becoming more innovative:

> *Time to market and your capacity for innovation as an organization will increase enormously if people in the organization immediately know what data they can use. They will be able to discover patterns in certain data faster, [..] and if they then also know what the data means, they can make new combinations with it, they can create new products, deliver new services. So on the innovation and time to market side I see great benefits of metadata management.*

4.3.4 *IT Infrastructure business benefit category.*

IT infrastructure business benefits have an impact on shareable and reusable IT resources that form that basis for business applications [37, p.279]. *IT complexity reduction* and *IT and data management control* are the IT infrastructure benefits identified in this study.

IT complexity reduction. Interviewees suggested that when a considerable number of IT systems are connected, the IT landscape can become complex to understand and to manage. Catalogs can provide insights in what data assets and systems exist, glossaries in what their functionality is, logical relationships in how they are related to each other, lineage in how data flows through them and metadata analytics in how they are used. This enables a reduction in IT complexity in two ways. First, it reduces the perceived IT complexity by providing detailed insights in all data assets and systems. Second, it enables the IT department to make informed decisions regarding the decommissioning or merging of systems. For example, unused reports could be shut off and systems with overlapping functionality could be merged. This could decrease physical IT complexity. The following quote by interviewee 10 illustrates that lineage is crucial to IT complexity reduction:

> *We need to turn off a platform, but we have no idea which systems are still depending on it. It is important? is it business critical or not? It takes a lot of effort.*

Interviewee 11b commented how aggregating and analyzing metadata can be used to aid physical IT complexity reductions:

> *You can make implicit knowledge explicit, and then you can analyze it, and realize that there is a lot of duplication, there's a lot of overlap, some things are still running because nobody has paid attention. [..] You see that with reporting, [thousands of] reports are being sent out and then there's probably one user. Some reports are just going to mailboxes, and just laying there, and there's nobody using it. While it's taking up your infrastructure, it's taking up your storage. There's so much cost in there.*

IT and data management control. Interviewees commented that clear roles, responsibilities and policies combined with an overview of all data assets and systems through glossaries and catalogs contribute to assuring that data assets and systems are properly maintained and only used with legitimate authorization. On the contrary, interviewees illustrated that not having a clear overview of data assets and systems and their functionality, combined with weak authorization policies, can lead to shadow IT. Shadow IT refers to systems that were never authorized or systems of which the functionality is unknown. Interviewee 10 indicated:

> *Because we do not have that technical metadata under control, one of our core customer systems has shadow IT. At one point we started*

counting how many loose interfaces it has. We stopped counting at 1500. We estimate that there are probably 5000 to 10000. We just stopped counting and we have no idea what those interfaces do.

Glossaries, catalogs, roles, responsibilities and policies and lineage enable organizations to get an overview of running IT projects, who is working on these projects and insight in what systems are impacted by these projects. Interviewees suggested that this enables more efficient IT project management. Interviewee 4 commented:

Often you see that programs and projects get in each others way. They are waiting for each other because they both need the same thing, try to change the same system, or try to adjust the same process. Usually, the overview of who is doing what is missing. [..] isn't it better to get it the first time right? Such decisions are of course often made at management level, and to be able to do that well, you just have to have a good overview of what you have, and who is doing what.

4.3.5 *Organizational business benefit category.*

Organizational business benefits are achieved through IT when it contributes to *"focus, cohesion, learning and execution of its chosen strategies"* [37, p.279] within the organization. This benefit dimension has a strong focus on employees, how they perform their work, and how they function within the organization. *Organizational alignment* and *employee satisfaction* are the organizational benefits that were identified in this study.

Organizational alignment. Interviewees argued that a glossary enables a common understanding about all objects that are defined in the glossary. This common understanding contributes to effective communication and collaboration between different departments. Moreover, interviewees reported that having clear definitions for data assets in a glossary enhances human understanding of data. Interviewee 1 commented:

At least it will help build, a common language, and a common understanding of each others business. I can assure you, that some people don't understand what is the other department is doing. [..] By just looking at the data, they don't understand, if there are no proper definitions.

Furthermore, lineage was attributed to enhancing transparency about how data flows through the organization. This could help people to understand that data collected or created in one part of the organization is required to support processes in another part of the organization. Interviewee 4 reported:

You can see that data is created somewhere by a business process, is eventually consumed somewhere else by other processes. So you can see as a callcenter employee: I collect e-mail addresses here, but that data will later be used by marketing, and by billing, and by all sorts of other departments. Then you can understand the value of your work.

Moreover, interviewees reported that catalogs can enhance data sharing across organizational silo's.

Employee satisfaction. Interviewees expressed that it is frustrating for employees if data is difficult to find. Furthermore, interviewees indicated that manually mapping lineage is a tedious task, while lineage is often required to back-trace the source of data issues, analyze the impact of IT changes or because regulators request to show it. Moreover, not having an overview of who is responsible for which data assets and systems can result in frustrating inefficiency when an employee has a question regarding a data asset or system. Furthermore, not having a glossary with standardized definitions for commonly used business objects can lead to unnecessary communication errors between employees and between departments. Additionally, interviewees indicated that employees dislike manual work for data integration related tasks.

Metadata management could contribute to general employee satisfaction by diminishing these issues by enabling catalogs, glossaries, lineage, data integration and interoperability and clear roles, rules and responsibilities.

The following statement by interviewee 3 illustrates how glossaries and lineage contribute to employee satisfaction

I am particularly talking about glossary for business people and lineage for the IT people in the organization as the two most important functions that contribute to employee satisfaction. If you no longer have to search endlessly for data, and for the meaning of data, then you can perform your daily activities much faster. If you have that same struggle every week, at some point it will drive you completely crazy.

Moreover, interviewee 11b commented that if data is made more available, accessible and understandable, it would contribute to the happiness of their employees:

So many people today want to use data, they want to be able to do data driven analysis, and make their decisions based on data. However, that data is not available. They don't know where it is, they have no place to find it, or they get it, but don't know what it means. If all this information is made available, it just helps everyone. Especially people working on the floor. People that actually need to work with data definitely will be most impacted. They would be so happy, if we have that.

5 DISCUSSION

This study advanced into exploring the business benefits of metadata management. The results indicate that metadata management indirectly leads to business benefits through the working of the functions it enables. This finding is in line with the notion that metadata leads to indirect value [13, 26, 27].

Notably, the discovered functions are interrelated. They are often combined in specific metadata-based tooling mentioned by the interviewees. For example, definitions that are included in a glossary tool could, at the same time, also be searched to reveal where the physical data resides. This search functionality, however, could be referred to as a data catalog. Moreover, information about roles, responsibilities and policies that apply to data assets and systems could also be presented, as well as relations to other data assets by means of lineage and logical relationships. The fact that various functions are combined in tools made it difficult for interviewees to distinguish between functions, and attribute specific functions to specific business benefits. However, it is promising that organizations are making various types of metadata available to their employees. In the current study, no single software solution that enabled all functions in one package was encountered. However, recent articles from Google™[21] and Lyft™[19] show that innovative technology companies are working on metadata-based software that enables a multitude of the functions found in this study.

Like the functions, the discovered business benefits are interrelated. For example, employee satisfaction is influenced by other benefits such as productivity improvement, IT complexity reduction and organizational alignment. This influence can be explained

through the notion that the Shang and Seddon framework [37] structures business benefits along organizational levels that influence each other. Nevertheless, the framework proved to be useful for structuring the business benefits of metadata management. Specifically, since it made it insightful that the business benefits of metadata management have an effect on various organizational levels.

The finding that the functions and the business benefits are both are interrelated explains the difficulty in attributing specific functions to specific business benefits. While the functions were used to explain the business benefits, the Delphi session did result in a consensus on coupling specific functions to specific business benefits. However, this does not take away that the functions and business benefits are all clearly the outcome of metadata management initiatives.

Organizations in different domains seem to value the business benefits differently. Notably, interviewees that were active in the finance domain strongly emphasized regulatory compliance. Considering the interviewees were active in various domains, no specific preferences for other industries could be distilled.

People seem to have the notion that metadata management is only an abstract and technical endeavour. Remarkably, the findings indicate that a significant part of metadata management revolves around simple and human-understandable actions. Defining what business objects exactly mean, assigning roles and responsibilities for data assets and systems and generally cataloguing what data assets are available in the organizations form an essential part of metadata management, that already contributes to various business benefits found in this study.

5.1 Theoretical and practical contribution

Literature that addresses the business value of metadata management was found to be scarce and an overview of the business benefits that can be expected to be achieved through metadata management initiatives was missing. This study posed a novel definition for metadata management and established a metadata management business benefit framework that explains how metadata management can indirectly enable business benefits through enabling functions. Furthermore, these functions and business benefits were illustrated and clarified with quotes from experts. Hereby, this study expands the theoretical knowledge on the business value of metadata management. Moreover, this study shows that the Shang and Seddon framework [37] is suitable to structure the business benefits of metadata management.

The results of this study have practical implications, considering organizations often struggle to implement data management initiatives due to the difficulty in establishing a strong business case [14, 28, 30]. This study sheds light on the possible positive outcomes of metadata management that can be expected on various organizational levels. The metadata management business benefit framework contributes to bringing business and IT closer together by lowering the level of abstraction and complexity in the metadata management domain. It could assist the IT department in establishing the business case for metadata management, and vice versa, business people in understanding the business value of metadata management. Hereby, the study contributes to lowering potential barriers to invest in metadata management initiatives.

5.2 Limitations

The cost of metadata management initiatives have not been considered in this study because the attainable business benefits, and thus the balance between costs and benefits, are expected to be unique for every organization.

Various interviewees indicated that metadata management initiatives are usually part of larger data management or data governance programs. Since their answers are based on personal experience, and this personal experience comes from metadata management initiatives that were embedded in larger programs, it is possible that not all business benefits that were identified can be fully attributed to isolated metadata management initiatives.

Involving metadata related software vendors and consultants raises ethical concerns, since they could have a personal interest in a publication that is positive about the benefits of metadata management. To prevent participation being seen as a marketing opportunity, it was clearly indicated upfront that all results would be anonymized. Furthermore, all statements were critically assessed during the coding process and the outcomes were tested during a Delphi session [23] to prevent inclusion of views that were not shared among participants.

The outcome of a qualitative study relies on the analysis and interpretation of the researcher, which can be influenced by personal traits and biases. Moreover, executing the study at a Deloitte could have led to a bias towards the practices of this firm. The fact that the study has been carried out by a single researcher does not contribute to mitigating these effects. Moreover, due to the relatively small sample of interviewees and the inexperience of the researcher, it is uncertain whether the interviews have led to full theoretical saturation [20]. Altogether, this makes it difficult to guarantee validity. By establishing a rigorous research method by adopting the Gioia methodology [17] and by doing a Delphi session [23] as a means of triangulation it was attempted to minimize these effects. Nonetheless, it is questionable if different research set-ups would lead to a valid general outcome, since the benefits that can be achieved through metadata management are expected to differ between organizations based on their unique traits.

5.3 Future research

To address the limited validity of the current study, future studies should address whether the functional capabilities and business benefits can be found in more organizations. Moreover, it could be investigated whether certain business benefits are more applicable to organizations based on specific traits (e.g. size, industry) to make the metadata management benefit framework more useful for specific organizations.

Business benefits of metadata management are indirect [27, 39] and the calculation of the cost of a metadata management initiative is a simpler affair than calculating its contribution to business performance after it has been put in place [27]. It remains a general difficulty to establish causality for business benefits [37]. The current study confirms these findings since only some quantifiable business benefits were encountered. However, qualitative research can offer

guidance for quantitative studies [17]. To quantify the business benefits, measurable constructs could be developed based on the findings of this study, after which detailed longitudinal case studies that consider the situation before and after the implementation of metadata management initiatives could be conducted. Nevertheless, it would remain a challenge to completely isolate metadata management from its context and prove that certain business benefits can be solely attributed to metadata management initiatives.

During the Delphi session, participants noted that reporting lineage is currently only possible within the boundaries of the individual organization. No lineage information is available for data that enters the organization and the lineage trace is lost as soon as data crosses the organization boundary. With ever more data being shared and increasingly strict regulations, a path for future research could lie in developing lineage mechanisms that enable organizations to report lineage across organizational chains.

6 CONCLUSION

The goal of this study was to increase the understanding of the business value of metadata management. Therefore, the following research question was posed: *What are the business benefits of metadata management and how are these business benefits enabled?*

To answer this question, an exploratory, qualitative and inductive study with experts that had multiple years of professional experience in metadata related projects was conducted. The study provides an answer to the research question by explicating that metadata management enables functions that contribute to business benefits. In this study, 7 of these benefit enabling functions and 11 related business benefits were identified. Together with the known concepts metadata management activities and metadata categories, the findings are visualized a metadata management benefit framework. Additionally, a novel definition for metadata management was posed.

Although some functions and business benefits might not have been identified in the current study and every organization might have unique priorities, the findings enhance the understanding about the business benefits of metadata management. Moreover, the metadata management business benefit framework provides guidance in explaining the business benefits of metadata management, hopefully contributing to bringing business and IT closer together.

ACKNOWLEDGMENTS

I would like to express deep gratitude towards my internal supervisor and first examiner Paul, my second examiner Benedikt and my external supervisors Niko and Kiean, for investing their valuable time in this thesis project. I am indebted to the experts that were involved in the interviews for their participation. I would like to thank Paul, Tessel, Erik en Yannick for all the fun discussions we had during the internship. Moreover, I would like to acknowledge the unconditional support provided by Noor, my roommates and my friends during this project. Finally, I wish to thank my parents for their support and encouragement throughout my studies. Without them, this accomplishment would not have been possible.

REFERENCES

[1] Ibrahim Alhassan, David Sammon, and Mary Daly. 2018. Data governance activities: a comparison between scientific and practice-oriented literature. *Journal of Enterprise Information Management* 31, 2 (2018), 300–316. https://doi.org/10.1108/JEIM-01-2017-0007

[2] Philip R. Bagley. 1968. Extension of Programming Language Concepts. *Philadelpia University City Science Center* (1968).

[3] Anitesh Barua, Charles H Kriebel, and Tridas Mukhopadhyay. 1995. Information technologies and business value: An analytic and empirical investigation. *Information systems research* 6, 1 (1995), 3–23.

[4] Besim Bilalli, Alberto Abelló, Tomàs Aluja-Banet, and Robert Wrembel. 2016. Towards Intelligent Data Analysis: The Metadata Challenge. *Proceedings ofthe International Conference on Internet ofThings and Big Data (IoTBD 2016)* January (2016), 331–338. https://doi.org/10.5220/0005876203310338

[5] Erik Brynjolfsson and Lorin M Hitt. 2000. Beyond computation: Information technology, organizational transformation and business performance. *Journal of Economic perspectives* 14, 4 (2000), 23–48.

[6] Milagros Castillo-Montoya. 2016. Preparing for interview research: The interview protocol refinement framework. *The Qualitative Report* 21, 5 (2016), 811–831.

[7] Saumya Chaki. 2015. *Enterprise Information Management in Practise. Chapter 7: Pillar no. 7: Metadata Management.* Number 7. 115–127 pages. https://doi.org/10.1007/978-1-4842-1218-9_10

[8] Guido De Simoni. 2019. Metadata Is the Fish Finder in Data Lakes. *Gartner Group* Januari (2019).

[9] Guido de Simoni, Alan Dayley, and Roxane Edjlali. 2018. Magic Quadrant for Metadata Management Solutions. *Gartner* August 2018 (2018), 1–31.

[10] Barbara DiCicco-Bloom and Benjamin F. Crabtree. 2006. The qualitative research interview. *Medical education* 40, 4 (2006), 314–321. https://doi.org/10.1108/11766091111162070 arXiv:arXiv:1011.1669v3

[11] Martin Eling, Hato Schmeiser, and Joan T Schmit. 2007. The Solvency II process: Overview and critical analysis. *Risk management and insurance review* 10, 1 (2007), 69–85.

[12] Päivi Eriksson and Anne Kovalainen. 2015. *Qualitative methods in business research: A practical guide to social research.* Sage.

[13] Adir Even, G. Shankaranarayanan, and Stephanie Watts. 2006. Enhancing decision making with process metadata: Theoretical framework, research tool, and exploratory examination. *Proceedings of the Annual Hawaii International Conference on System Sciences* 8, C (2006), 1–10. https://doi.org/10.1109/HICSS.2006.152

[14] Neil Foshay, Avinandan Mukherjee, and Andrew Taylor. 2007. Does data warehouse end-user metadata add value? *Commun. ACM* 50, 11 (2007), 70–77. https://doi.org/10.1145/1297797.1297800

[15] Neil Foshay, Andrew Taylor, and Avinandan Mukherjee. 2014. Winning the Hearts and Minds of Business Intelligence Users: The Role of Metadata. *Information Systems Management* 31, 2 (2014), 167–180. https://doi.org/10.1080/10580530.2014.890444

[16] Joel Gehman, Vern L Glaser, Kathleen M Eisenhardt, Denny Gioia, Ann Langley, and Kevin G Corley. 2018. Finding theory–method fit: A comparison of three qualitative approaches to theory building. *Journal of Management Inquiry* 27, 3 (2018), 284–300.

[17] Dennis A. Gioia, Kevin G. Corley, and Aimee L. Hamilton. 2013. Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods* 16, 1 (2013), 15–31. https://doi.org/10.1177/1094428112452151

[18] Jane Greenberg. 2003. Metadata and the World Wide Web. *Encyclopedia of Library and Information Sciences* 3 (2003), 1876–1888. https://doi.org/10.1081/E-ELIS

[19] Mark Grover. 2019. Amundsen - Lyft's data discovery & metadata engine. (2019). Retrieved June 11, 2019 from https://eng.lyft.com/amundsen-lyfts-data-discovery-metadata-engine-62d27254fbb9

[20] Greg Guest, Arwen Bunce, and Laura Johnson. 2006. How Many Interviews Are Enough? An Experiment with Data Saturation and Variability. *Field Methods* 18, 1 (2006), 59–82. https://doi.org/10.1177/1525822X05279903

[21] Alon Halevy, Flip Korn, Natalya F. Noy, Christopher Olston, Neoklis Polyzotis, Sudip Roy, and Steven Euijong Whang. 2016. Managing Google's data lake: an overview of the Goods system. *IEEE Data Engineering Bulletin* 39 (2016), 5–14.

[22] Deborah Henderson, Susan Earley, Laura Sebastian-Coleman, Elena Sykora, and Eva Smith. 2017. *Data Management Association - Data Management Body Of Knowledge (DAMA-DMBOK)* (2nd editio ed.). Technics Publications, LLC, New Jersey.

[23] Chia-chien Hsu and Brian A. Sandford. 2007. The Delphi Technique: Making Sense Of Consensus. *Practical Assessment Research and Evaluation* 12, 10 (2007). https://doi.org/10.1576/toag.7.2.120.27071 arXiv:ISSN 1531-7714

[24] Kai M. Hüner, Boris Otto, and Hubert Österle. 2011. Collaborative management of business metadata. *International Journal of Information Management* 31, 4 (2011), 366–373. https://doi.org/10.1016/j.ijinfomgt.2010.12.002

[25] Zahir Irani and Peter ED Love. 2002. Developing a frame of reference for ex-ante IT/IS investment evaluation. *European Journal of Information Systems* 11, 1 (2002), 74–82.

[26] Miltiadis D. Lytras and Miguel Angel Sicilia. 2008. Where is the value in metadata? *International Journal of Metadata, Semantics and Ontologies* 2, 4 (2008), 235–241. https://doi.org/10.1504/ijmso.2007.019442

[27] Miltiadis D Lytras, Miguel-Angel Sicilia, and Cristian Cechinel. 2013. The value and cost of metadata. *Handbook of Metadata, Semantics and Ontologies* (2013), 41–59.

[28] Nick Martijn, Joris Hulstijn, Mark de Bruijne, and Yao Hua Tan. 2015. Determining the effects of data governance on the performance and compliance of enterprises in the logistics and retail sector. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9373 (2015), 454–466. https://doi.org/10.1007/978-3-319-25013-7_37

[29] Nigel Melville, Kenneth Kraemer, and Vijay Gurbaxani. 2004. Information technology and organizational performance: An integrative model of IT business value. *MIS quarterly* 28, 2 (2004), 283–322.

[30] E Nass and M Scheibmayer. 2011. Defining a research framework for the business impact of data management. *2011 17th International Conference on Concurrent Enterprising* Ice (2011), 1–8.

[31] Harshvardhan J Pandit, Declan O'Sullivan, and Dave Lewis. 2018. Queryable Provenance Metadata For GDPR Compliance. *Procedia Computer Science* 137 (2018), 262–268.

[32] Florentina Peels, Roger W H Bons, and Marijn G A Plomp. 2016. The Business Value of Enterprise Data Models. *AMCIS 2016 Proceedings* (2016), 1–10. http://aisel.aisnet.org/amcis2016/EntSys/Presentations/4

[33] Lukas Prorokowski and Hubert Prorokowski. 2015. Solutions for risk data compliance under BCBS 239. *Journal of Investment Compliance* 16, 4 (2015), 66–77.

[34] Ron Roszkiewicz. 2010. Enterprise metadata management: How consolidation simplifies control. *Journal of Digital Asset Management* 6, 5 (2010), 291–297. https://doi.org/10.1057/dam.2010.32

[35] Shazia Sadiq (Ed.). 2013. *Handbook of Data Quality: Research and Practise.* 41–73 pages. https://doi.org/10.1007/978-3-642-36257-6

[36] E.D. Scheirer. 2002. About this business of metadata. *Proc. International Symposium on Music Information Retrieval* (2002), 252–254. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.6443

[37] S. Shang and P. B. Seddon. 2002. Assessing and managing the benefits of enterprise systems: The business manager's perspective. *Information Systems Journal* 12, 4 (2002), 271–299. https://doi.org/10.1046/j.1365-2575.2002.00132.x

[38] G. Shankaranarayanan and Adir Even. 2004. Managing Metadata in Data Warehouses: Pitfalls and Possibilities. *Communications of the Association for Information Systems* 14, September (2004), 247–274. https://doi.org/10.17705/1cais.01413

[39] Ganesan Shankaranarayanan and Adir Even. 2006. The Metadata Enigma. *Commun. ACM* 49, 2 (2006).

[40] D Stock and R Winter. 2011. *The Value of Business Metadata: Structuring the Benefits in a Business Intelligence Context.* Physica-Verlag HD, Heidelberg. 133–141 pages. https://doi.org/10.1007/978-3-7908-2632-6

[41] Daniel Stock, Felix Wortmann, and Jörg H Mayer. 2010. Use Cases for Business Metadata - A Viewpoint-Based Approach to Structuring and Prioritizing Business Needs. *International Conference on Business Information Systems* (2010), 226–237. https://doi.org/10.1007/978-3-642-41687-3

[42] Besiki Stvilia, Les Gasser, Michael B. Twidale, and Linda C. Smith. 2007. A Framework for Information Quality Assessment. *Journal of the American Society for Information Science* 58, 12 (2007), 1720–1733. https://doi.org/10.1002/asi

[43] Paul P Tallon, Kenneth L Kraemer, and Vijay Gurbaxani. 2000. Executives' perceptions of the business value of information technology: a process-oriented approach. *Journal of Management Information Systems* 16, 4 (2000), 145–173.

[44] Toomas Tamm, Peter B Seddon, Graeme Shanks, and Peter Reynolds. 2011. How does enterprise architecture add value to organisations? *Communications of the association for information systems* 28, 1 (2011), 10.

[45] Adrienne Tannenbaum. 2001. *Metadata Solutions: Using Metamodels, Repositories, Xml, and Enterprise Portals to Generate Information on Demand.* Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA.

[46] Thiru Thangarathinam, Gregg Wyant, Jacque Gibson, and John Simpson. 2004. Metadata Management: the Foundation for Enterprise Information Integration. *Intel Technology Journal* 8, 4 (2004), 337–345.

[47] Marcel van Oosterhout, Eric Waarts, Eric van Heck, and Jos van Hillegersberg. 2006. Business agility: need, readiness and alignment with IT-strategies. In *Agile information systems: Conceptualization, construction and management.* Butterworth Heinemann, 52–69.

[48] Patrick Verweij. 2017. Understanding the value of data governance. *Masters thesis* December (2017).

[49] Lubos Vnuk, Andy Koronios, and Jing Gao. 2012. Enterprise Metadata Management: Identifying Success Factors For Implementing Managed Metadata Environments. *PACIS 2012 Proceedings.* (2012).

[50] Jane Webster and Richard T Watson. 2002. Analyzing the Past To Prepare for the Future: Writing a Review. *MIS Quartely* 26, 2 (2002), 12. www.jstor.org/stable/4132319

[51] Joost F. Wolfswinkel, Elfi Furtmueller, and Celeste P.M. Wilderom. 2013. Using grounded theory as a method for rigorously reviewing literature. *European Journal of Information Systems* 22, 1 (2013), 45–55. https://doi.org/10.1057/ejis.2011.51

[52] Anneke Zuiderwijk, Marijn Janssen, and Iryna Susha. 2016. Improving the speed and ease of open data use through metadata, interaction mechanisms, and quality indicators. *Journal of Organizational Computing and Electronic Commerce* 26, 1-2 (2016), 116–146. https://doi.org/10.1080/10919392.2015.1125180

[53] Anneke Zuiderwijk, Keith Jeffery, and Marijn Janssen. 2012. The Potential of Metadata for Linked Open Data and its Value for Users and Publishers. *JeDEM - eJournal of eDemocracy and Open Government* 4, 2 (2012), 222–244. https://doi.org/10.29379/jedem.v4i2.138

APPENDIX A LITERATURE REVIEW

The adopted literature review method involved iterative cycles of the stages *define, search* and *select*, whereafter the findings were analyzed. The *define* stage marked out the scope of the study. Inclusion and exclusion criteria were established (see Table 2), the focal fields of study "business" and "information technology" were determined, Google Scholar was selected as literature search engine, and search terms were established (See Table 3). The *search* stage is the actual execution of the search that was prescribed in the define stage. While searching through databases, surprises such as unknown synonyms and new suitable search terms arose, after which the define stage had to be revisited. All iterations were documented (See Table 3). The *Select* stage concerns selecting papers that fit the study. Papers were selected on title and keywords first. Hereafter the selection was refined on abstract and conclusion. Next, the selection was refined on the full text and general quality (See Table 4). Backward and forward citations were checked for every suitable paper to exhaust the available literature. To organize the literature and to prevent including duplicates, reference manager Mendeley™ was used.

The selected papers were read. Relevant quotes were copied into MS Excel™ and assigned a code that indicated the subject of the quote. Furthermore, *any* positive outcome enabled by metadata or metadata management was added to a concept matrix. The concept matrix contained 134 concepts found in 25 papers, of which only 8 occurred more than twice and 117 occurred only once. The result was not helpful towards understanding the actual business benefits of metadata management. However, it did help to understand the indirectness of the benefits and the distinction between functions and business benefits. Furthermore, it inspired to use an existing business benefit framework to structure the business benefits.

Include	Exclude
- Papers that adress sources of value of metadata (management)	- Papers not written in English
- Papers that adress the practicalities of metadata (management)	- Filetypes other than .pdf
- Papers <2000 in cross-reference search	- Papers <2000 in database search
- Practical literature	- Papers that only discuss algorithms and software
- Papers from the first 50 search results	- Papers that approach metadata management from a purely technical perspective

Table 2: Inclusion and exclusion criteria

Query or search method	Round	Number of results
- "Metadata management"	1	5
- "Metadata management" + Business Value	1	11
- "Metadata Management" + definition OR define	1	1
- "Metadata management" + Value	1	2
- "Metadata management" + business benefits	2	4
- Metadata + "Business benefits"	3	3
- Metadata + "Business value"	3	2
- Metadata + value	3	6
- Cross-reference search (backward)	4	11
- Cross-reference search (cited by)	4	13
- Cross-reference search (backward)	5	1

Table 3: Definition of search terms and documentation of literature search

Selection	Number of papers
- Initial papers selected	59
- Excluded on abstract and conclusion	9
- Excluded on full-text	19
- Excluded on quality	5
- Final papers included and fully read	26

Table 4: Paper selection

APPENDIX B INITIAL CONCEPTUAL FRAMEWORK

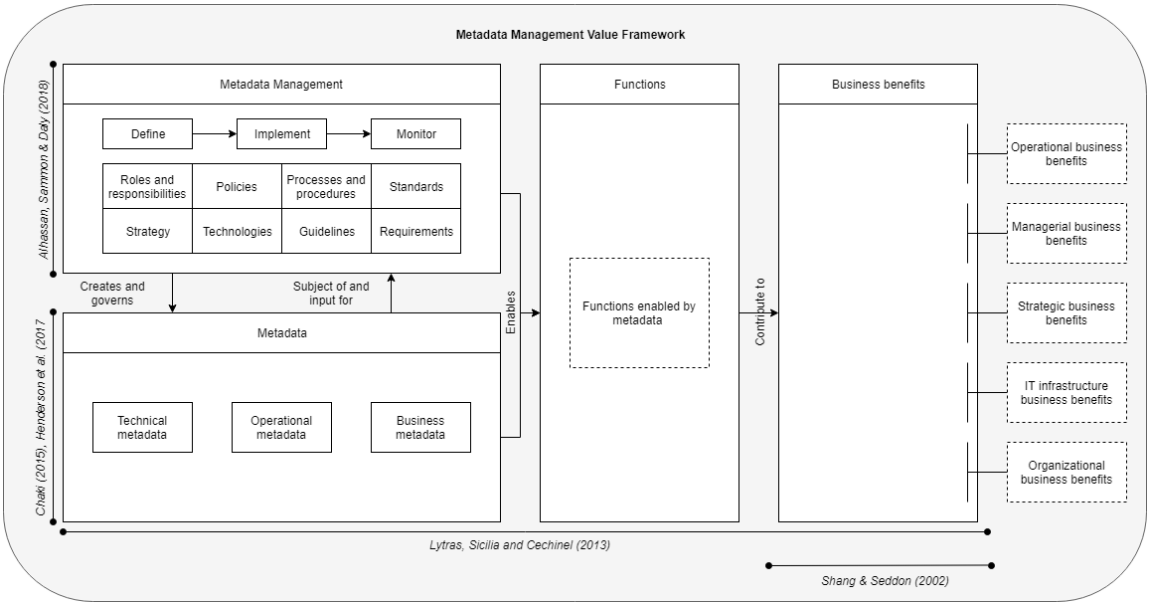


Figure 3: Conceptual Framework

APPENDIX C INTERVIEWEE INFORMATION

#	Sector	Role	Duration
01	Consulting	Manager in Enterprise Data Management	00:42:47
02	Consulting	Manager in Enterprise Data Management	00:41:01
03	Consulting	Consultant in Enterprise Data Management	00:50:47
04	Software	Senior Sales Consultant & Regional Leader Data Governance	00:55:07
05	Consulting	Founder and Vice President	00:38:43
06	Software	Founder and CEO	00:46:07
07	Consulting	Founder and Managing Partner	00:48:32
08	Consulting	Manager in Data and Technology	00:30:33
09	Consulting	Senior Manager in Strategy & Operations	00:52:26
10	Banking	Enterprise Data Advisor	00:59:03
11a	Insurance	Solution Expert in Data Management	01:07:00
11b	Insurance	Metadata Analyst	01:07:00

Table 5: Background of the interviewees and duration of the interviews

APPENDIX D INTERVIEW PROTOCOL

- **Introduction**
  - Introduce myself.
  - Express gratitude for participation.
  - Explain the current study.
  - Explain the overall structure of the interview.
  - Ask permission to record the interview.
    - Recordings will be transcribed and anonymized.
    - Informed consent form if required.

- **Open questions about personal experience with Metadata Management**
  - Who is your current employer and what is your current role?
  - Could you tell me about your professional experience with metadata related projects?
  - When do you typically have to deal with metadata within a business context?
    - In what situation is not having well managed metadata problematic?
    - In what situation is having well managed metadata beneficial?
  - How would you define Metadata?
    - Let the interviewee define Metadata first.
    - In this study, Metadata is defined as:
      - *[Metadata is] data about an object that supports functions associated with the designated object - (Greenberg, 2003)*
    - Do you have any objections to this definition?
  - How would you define Metadata Management?
    - Let the interviewee define the term Metadata Management first
    - In this study, Metadata is defined as:
      - Metadata Management is an organization-wide approach concerned with defining, implementing and monitoring all metadata related decisions, with the goal of achieving accessible high-quality metadata that leads to business benefits.
    - Do you have any objections to this definition?
  - What are typical positive outcomes of metadata management activities for organizations?
    - Could you describe how this leads to business value?
    - Do you have practical examples?
  - What would you include if you were to establish a business case for metadata management?

- **Briefly explain conceptual framework to guide the next part of the interview**
  - Metadata management
  - Metadata
  - Functions
    - Business benefits are enabled through functions that can be clear seen as metadata management outcomes
    - Examples of functions are lineage, impact analysis or glossaries (only provide basic examples on request!)
  - Business benefits

- **Questions about metadata management and business benefits**
  - Do and ask the following for all five business benefit categories (A-E):
    - Explain the business benefit category
    - Do you think Metadata Management enables organizations to achieve [category] business benefits?
    - What functions or functional capabilities enabled by metadata contribute to these benefits?
    - Could you give any practical examples?

  A Operational business benefits
    - *"Operational activities process day-to-day activities that involve acquiring and consuming resources"* [37, p.278]. Typical operational business benefits are increased productivity, cycle time and cost reductions and improvements in general quality and customer service. Exemplary, Information Systems can contribute to operational benefits by enabling automation, substituting manual labour and speeding up processes.

  B Managerial business benefits
    - Managerial business activities are concerned with *"allocating and controlling the firm's resources, monitoring of operations and supporting of business strategic decisions"* [37, p.278]. Typical managerial business benefits are better management of resources, improved planning and decision making and organization wide performance improvements. Information Systems can, per example, contribute to managerial business benefits by providing summarized (real-time) enterprise information.

  C Strategic business benefits

    - The strategic benefit category is concerned with strategic future planning and the overall direction of the enterprise. Strategic activities involve *"long-range planning regarding high-level decisions, such as business merging and acquisition, marketing competition, product planning, customer retention and capital sourcing"* [37, p.279]. Supporting innovation and alliance, building innovations, cost leadership and external linkages and enabling product differentiation are typical strategic business benefits enabled by Information Systems.

  D IT infrastructure business benefits
    - IT infrastructure concerns shareable and reusable IT resources that form that basis for business applications. IT infrastructure benefits can be anything that contributes to these IT resources or that supports business applications. Exemplary IT infrastructure business benefits are support for business flexibility, IT cost reduction and increased IT infrastructure capabilities [37, p.279].

  E Organizational business benefits
    - Organizational benefits are enabled when Information Systems *"benefit an organization in terms of focus, cohesion, learning and execution of its chosen strategies"* [37, p.279]. This benefit dimension has a strong focus on employees, how they perform their work, and how they function within the organizational context. Exemplary organizational business benefits enabled by information systems are empowerment of employees, improvements in work patterns and organizational culture, enabling organizational learning and helping to create a common vision across an organization.

- **Closing**
  - Are there any questions from your side? Are there questions that you expected me to ask, but I didn't? Are there any other insights you want to share?
  - Do you want to participate in a Delphi Session?
  - Again express gratitude for participating and the valuable insights given by the interviewee
  - Thesis will be shared when finished.

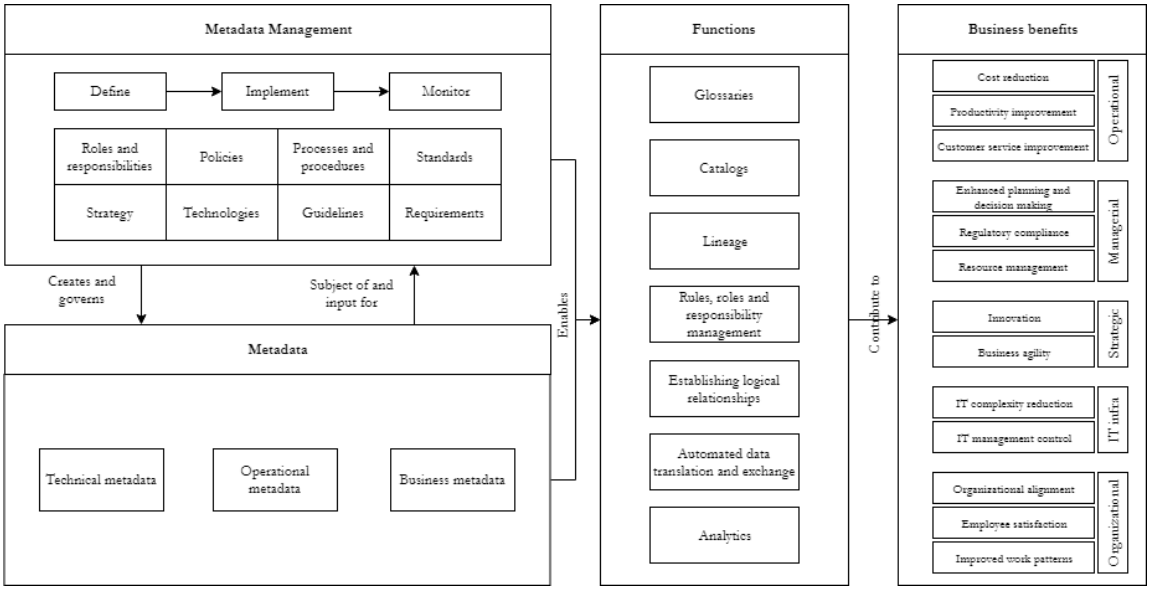
APPENDIX E DELPHI SESSION

E.1 Participants

Sector	Role	Interviewee
Consulting	Manager in Enterprise Data Management	02
Software	Senior Sales Consultant & Regional Leader Data Governance	04
Consulting	Senior Manager in Strategy & Operations	09
Banking	Enterprise Data Advisor	10
Consulting	Consultant in Enterprise Data Management	-
Consulting	Senior Consultant in Enterprise Architecture	-

Table 6: Participants of the Delphi session

E.2 Metadata management value framework at the time of the Delphi session



E.3 Outcome for functions

- Glossary - rephrased to specification and standardization (Glossaries)
- Catalogs - rephrased to metadata aggregation (Catalog)
- Roles, roles and responsibility management - rephrased to enforce rules, responsibilities and policies
- Automated data translation and exchange - rephrased to data integration and interoperability
- Lineage - Only consensus if all types of lineage including horizontal lineage are clearly explained.

E.4 Outcome for business benefits

- Resource management - deleted because it can be divided between planning and decision making and productivity improvement
- Changing work patterns - deleted because it was seen as a weak category that could be divided among productivity improvement, organizational alignment and employee satisfaction
- IT management control - rephrased to IT and data management control to emphasize that this type of control also concerns data.
- Innovation - consensus, but only if innovation also refers to improving existing products

E.5 Relationship between functions and business benefits

- No consensus on coupling specific functions to specific business benefits
- Participants noted that this connection was hard to establish since functional capabilities are strongly related.

APPENDIX F METHODOLOGY AND FINAL CODING CATEGORIZATION

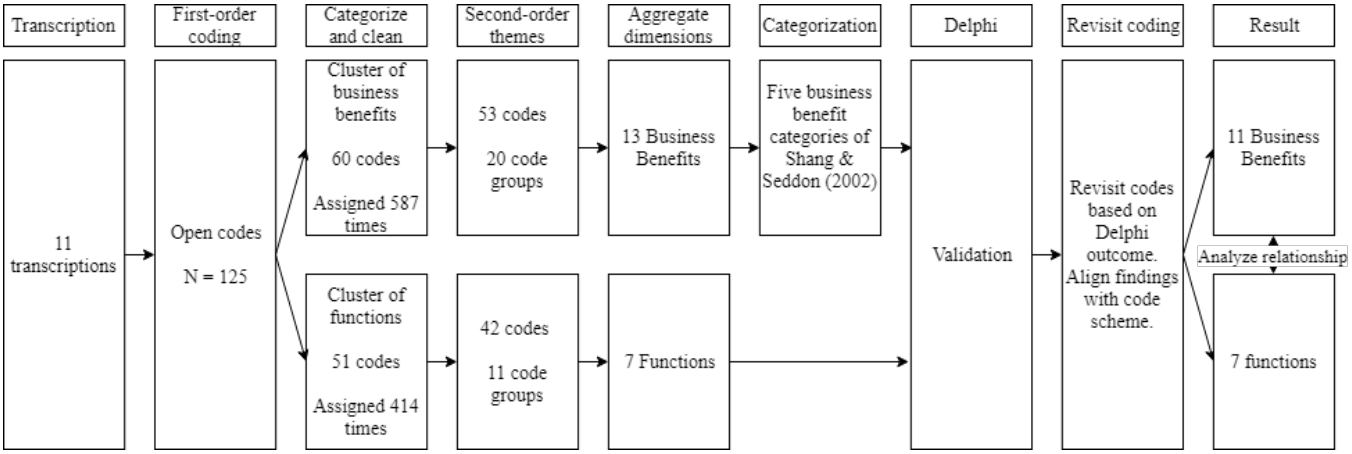


Figure 4: Overview of intermediate steps in methodology

Metadata aggregation (Catalogs)	**Metadata specification and standardization (Glossaries)**	**Logical models**	**Lineage**	**Enforcement of roles, responsibilities and policies**	**Data integration and interoperability**	**Metadata analytics**
- Data catalog (22) - Overview of available data (17) - System overview (15) - Cross platform centralization (11) - Program or project overview (6) - Metadata repository (3) - Record data quality issues (2) - Map IT landscape (2)	- Create definitions (35) - Glossary (30) - Data dictionary (11) - Record implicit knowledge (5) - Data specification (3)	- Common interchangeable format (4) - Taxononomies (3) - Semantical understanding (3) - Machine understanding (1) - Ontologies (1)	- Lineage (44) - Impact analysis (17) - Data flow insights (11) - Record data usage (4) - Traceability (3) - Provenance (1) - Data tracking (1) - Data flow chart (1) - Workflow tooling (1)	- Record ownership (38) - Record accountability (16) - Enforce rules (8) - Sensitive data classification (4) - Data governance (4) - Record roles and responsibilities (2)	- Automation (20) - Data integration (17) - Data exchange between systems (11) - Data exchange between parties (5) - Integrate external data (4)	- Analytics (26) - Data quality insights (18) - Measures (2) - Data clustering (1)
I=10, A=78	I=11, A=84	I=5, A=12	I=9, A=83	I=10, A=72	I=11, A=57	I=11, A=47

Table 7: Final coding schema for functions. Column headers indicate functions. Text in columns indicate open codes. The frequency of assignment of a code is indicated in parentheses after each code. I indicates the number of interviews in which codes related to a function occurred. A indicates the total number of times codes from a group were assigned. The bottom row shows the grand totals.

Operational	Managerial	Strategic	IT infrastructure	Organizational
Operational cost reduction - Cost reduction (30) - Manual work reduction (12) - IT cost reduction (6) - Optimize external data acquisition (5) - Employee reduction (4) I=11, A=57 **Productivity improvement** - Quicker data discovery (27) - Efficiency increase (27) - More efficient BI reporting (22) - Search time reduction (13) - Error reduction (10) - Operational process optimization (6) - Productivity improvement (4) - Overwork reduction (1) I=11, A=110 **Customer service improvement** - Customer service improvement (10) - Customer trust (2) - Customer satisfaction (1) - Customer retention (1) I=5, A=14	**Planning and decision making** - Decision making (47) - Trustability of data (19) - Portfolio steering (6) - Resource management (4) - Business process management (3) - Inventory planning (2) - Supply chain management (1) I=11, A=82 **Regulatory compliance** - Regulatory compliance (45) - Risk reduction (9) I=9, A=54	**Innovation** - Innovation (9) - Reduced time to market (4) I=5, A=13 **Business agility** - Support during mergers and acquisitions (9) - Business agility (8) - Flexibility for change (2) I=7, A=19	**IT complexity reduction** - Redundancy reduction (13) - Complexity reduction (6) - Phase out old IT systems (5) - Consolidate systems (2) I=8, A=26 **IT and data management control** - Data control (23) - Shadow IT reduction (3) - Monitoring system operations (2) - Assetize data (1) - Interference reduction (1) - Technology and data landscape control (1) I=8, A=31	**Organizational understanding and alignment** - Transparency (15) - Enhanced human understanding of data (15) - Business understanding (14) - Collaboration (13) - Communication (12) - Communication of context (10) - Awareness (5) - Information sharing (3) - Common data vision (1) I=11, A=88 **Employee satisfaction** - Improved work patterns (11) - Employee satisfaction (7) - Empowerment (5) I=8, A=23
I=11, A=181	I=11, A=136	I=9, A=32	I=10, A=57	I=11, A=111

Table 8: Final coding schema for business benefits. Column headers business benefit categories. Bold headings within columns indicate business benefits. Normal text in columns indicate open codes. The frequency of assignment of a code is indicated in parentheses after each code. I indicates the number of interviews in which codes from a business benefit occurred. A indicates the total number of times codes from a business benefit were assigned. The bottom row shows the grand totals of I and A for each business benefit category.