MD3M Master Data Management Maturity Model - Developing an Assessment to Evaluate an Organization's MDM Maturity

Katharina Pietzka (3571572)

University of Utrecht

Author Note

Katharina Pietzka, Master Thesis in Business Informatics, Department of Information and Computer Science, University of Utrecht.

Contact: k.pietzka@students.uu.nl

Abstract

This research deals with assessing the master data maturity of an organization. It presents a thorough literature study on the main concepts and best practices in master data maturity assessment. A maturity matrix is developed and described in detail. Furthermore, an assessment questionnaire is presented which is used to assess the master data management maturity of the case company E.ON Energy Trading. Emphasis is laid on the academic validity of the model development process. The results of the organization's maturity are presented as well as conclusions on the research and the developed maturity model. The purpose of the paper is to show reasons and incentives for a decent master data management and provide a benchmarking tool with which different organizations can compare their levels. Concluding, further research areas are presented and the model is critically evaluated.

Keywords: Master Data Management, Maturity Assessment, Information Management

Glossary[1]

API#2	Pricing Index for steam coal regarding the delivery to the area of Amsterdam, Rotterdam and Antwerp (ARP); important price index for OTC financial swap trades.
API#4	Pricing Index for physical steam coal traded FOB in South Africa; it is commonly used for OTC financial swaps.
APX	European Energy Exchange in the Netherlands operates in electricity and gas markets for Netherlands, Belgium and the UK.
ARA	Abbreviation for the port and refining area Amsterdam, Rotterdam, Antwerp.
BI	Business Intelligence.
BPM	Business Process Management.
CER	Certified Emissions Reductions; output of the Kyoto Protocol, unit of greenhouse gas reduction, somes with the right to emit a certain amount of CO^2
Clearing	Offsetting and settlement of trading transactions; by clearing, the clearing house becomes the buyer or seller to each counterparty when dealing derivatives.
Clearing Member	Are members of exchanges and take the responsibility for trades cleared through them.
Crack Spread	Calculation of the value of a barrel of oil regarding the value of its final refined products (gasoline, heating oil,…).

[1] The explication of the trading terms in the glossary are based on (Sapient Global Markets, 2011) and (Energy Risk, 2010)

Dark Spread	Difference between the market price of one unit of electricity compared to the gas used to produce it.
EET	E.ON Energy Trading.
EFET	European Federation of Energy Traders; group of more than 90 energy trading companies to improve energy trading conditions.
Effective Date	Date on which obligations for a derivative transaction take effect.
Emissions Trading	Market for trading polluting emissions, evolved after set up of the Kyoto Protocol.
ERU	Emission Reduction Unit; measure of carbon dioxide emissions for participants of the Kyoto Protocol.
EUA	European Union Allowances – the governmental secured rights to emit a certain amount of CO^2
Exchange	Central marketplace for buyers and sellers to trade securities, futures and options.
FOB	Free-On-Board; seller provides oil product or gas in a way that the loading costs are taken care of, but buyer takes responsibility for risks.
Forward	Contract including the sale by one party to another one of a predefined amount of a commodity at a predefined price and a predefined date in the future.
Future	Exchange traded agreement to deliver an asset at a specific time in the future for a specific price agreed today; standardized forward contract.

Hedging	Trading strategy to reduce or mitigate risk; a second transaction is done to offset the first one's risk.
Index	Portfolio of assets calculated and published to a index sponsor to give indications of the market and measure the market's performance.
ISDA Master Agreement	Market standard agreement to facilitate OTC derivative trading; contains a negotiated details like credit, legal and operational terms.
Kyoto Protocol	International agreement to reduce greenhouse gas emissions under the United Nations Framework Conventions on Climate Change.
Load Shape	Combination of electricity contracts covering a period of time reflecting power requirements.
Maturity Date	Date when a debt security or OTC derivative expires.
MDM	Master Data Management.
Netting	Offsetting of positive transactions against negative transactions between counterparties to reduce settlement risks.
Option	The right, but not obligation to buy/sell (call/put) a financial instrument at an agreed price.
OTC	Over-The-Counter; trading between two counterparties without use of exchange as intermediate; trade terms are freely negotiated.
Proprietary Trading	Also Prop Trading; strategy of financial institutes to trade on their own account rather than with customer capital.

Settlement	Process of solving obligations of a trade through payment and/or delivery.
SME	Small and medium Sized Enterprise
Spark Spread	Difference between market price of a unit of electricity and market price of the fuel to generate it.
Spot (deal)	Standardized contract for bonds, foreign currencies or other things with same conditions which have to be fulfilled within two trading days after conclusion the business transaction.
SOA	Service Oriented Architecture.
Spread	Simultanuous purchase of one security and sale of a related security, which are called legs; the purpose is to achieve a net position whose value depends on the prices of the legs; ususally spreads are traded as one unit to minimize execution risk.
Swap	Two counterparties exchange streams of cashflows; these flows are called legs.
Swaption	An option to enter into a predefined swap transaction.
UML	Unified Modeling Language, a modeling language for e.g. databases, programme setups and use cases, often used in software engineering
Underlying	Asset, index, currency, rate, credit etc. which serves as the basis to calculate value and cash flow or deliverable of a transaction.
Vanilla	Derivative transaction with a simple structure; often traded in the relevant markets.

Table of Contents

List of Tables

Table of Figures

Declaration of Academic Honesty

I hereby declare that this master thesis has been written only by the undersigned and without any assistance from third parties. Furthermore, I confirm that no sources have been used in the preparation of this thesis other than those indicated in the thesis itself. This thesis, in same or similar form, has not been available to any audit authority yet.

Frankfurt, 19.08.2012 Signature

MD3M Master Data Management Maturity Model -

Developing an Assessment to Evaluate an Organization's MDM Maturity

Information and data became over time increasingly important and a crucial competitive factor. In literature, major societal shifts towards an "information society" or an "information age" were already predicted in the late 1970's and early 1980's. Beniger shows an overview of the societal shift predictions (1986). He also describes the development towards an Information Society starting in the 19th century. Some centuries ago, the world's economy was described in three sectors. The Primary Sector processes raw materials, the Secondary Sector manufactures things and the Tertiary Sector offers services. Societies will shift their main economic activities from the first to the third sector when increasing their stage of development. Nowadays, a new Quaternary Sector has been defined. The Tertiary Sector has grown over time and was used to cover too diverse activities. This additional sector contains information-based activities. (Kenessey, 1987) This shows the development towards an information-centered economy. Research in this area therefore has a significant economic and societal relevance.

Organizations' assets do not only consist of crude materials or finished products to sell. In many cases information and data that the company has collected and now provides is a valuable asset that even mergers have been conducted because of the data that one company owns and the other wants to obtain.

The first chapter introduces the motivation and research objectives of this thesis work. The targets and the research question will be presented as well as the research approach. The following chapter provides a thorough literature study of the fields of

interest. Chapter 3 covers the development of the master data management maturity model and the corresponding assessment tool. After this, the practical application at the case company is described. In the fifth chapter, the obtained results from the research undertaken are presented and recommended areas for improvement are shown. Consequently, the results are discussed and finally, conclusions are drawn in the final chapter.

1 Motivation and Research Objectives

Starting from a particular size, every organization has to deal with the question of how to integrate data from different units or areas. Furthermore, the organizational setup of the firm plays a big role. Does the firm operate in a data-intense business? In areas with strict regulations on the traceability of events and accountability (like pharmaceutical industries, finance, trading), with data as a main source for added value (like the finance industry), or with the urgent need for efficiency and agility (production, technical industries), a systematic integration of data is crucial for the business. (Wegener, 2008) The recent development shows that organizations have to cope with short innovation cycles and market launch times. Furthermore, the complexity is increasing due to globally harmonized business processes and global customer services. This results in shorter decision cycles basing on more information. (Otto & Hüner, 2009)

Problems in integrating data do not result from mechanical reasons. Databases and hardware are so advanced that they are able to deal with all clear formulated requirements in a business environment. Also performance or stability are problems that are generally a minor issue considering the state of the art. The biggest issue in organizing data is the logic behind it. (Wegener, 2008) Defining the logic is an ambitious challenge because the data landscape in a company is very complex, the data entities are interdependent, and there is a huge amount of data to be structured.

If organizations get acquired or consolidate, there is a huge effort to be taken in order to integrate and migrate the data from those companies. Usually, the data is not structured in a similar way based on the same data models. This can derive from the variety of business; in different branches and markets, there might be big differences

between the products, which lead to differences in the processes and data landscape. Another factor complicating the modeling of data is companies being located in different countries and therefore having to comply with different laws and regulations regarding processes and reporting. Even within one firm, different departments are often autonomous in terms of organizing their data. (Wegener, 2008)

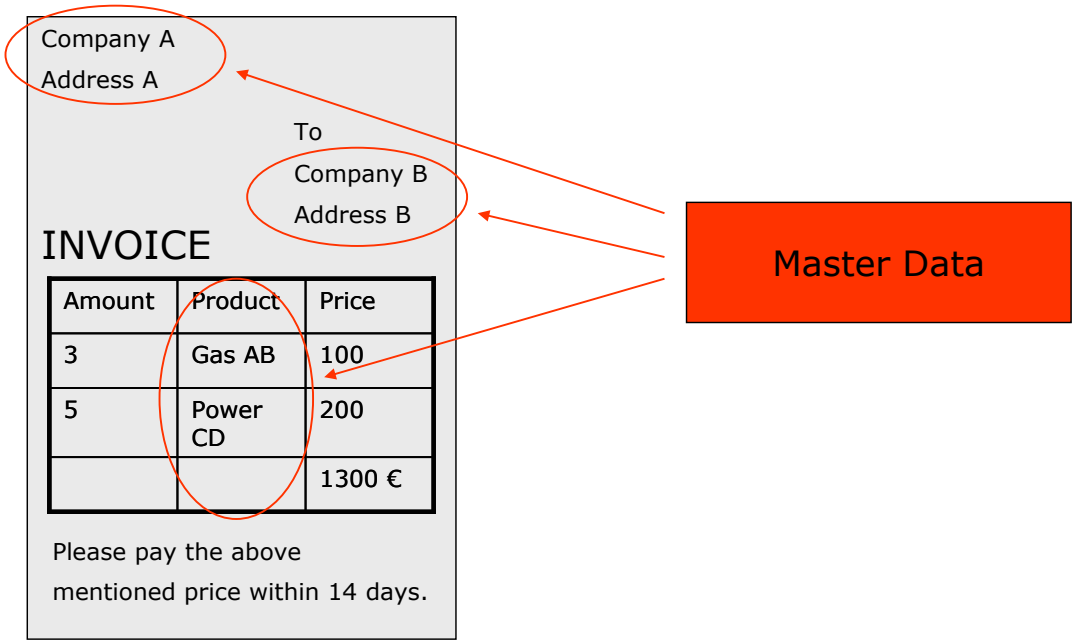
In an organizational context, synergies arise from the fact that the conducted activities of one organizational unit can be used for another unit's work. In addition, that organizational unit can help decrease cost or increase earnings by bundling competences and relating between the units. Synergies have to be seen from a broader perspective. It is possible that some units put a lot of effort in producing high quality data without benefitting from them. Other units however, benefit a lot without putting any effort in it. (Winter, Schmaltz, Dinter, & Bucher, 2008) These effects can create synergies and have to be considered from a global organization-wide perspective. Some departments fulfill supportive functions and therefore, from a business perspective, only cost money and do not generate any benefit, but from a transitive perspective, they do. So, an effective master data management can help reveal existing and possible synergies.

1.1 Problem Statement

The bigger the company, the bigger is the issue of managing data. This research will follow the definition of the European Commission, saying that big enterprises start from a number of employees of more than 250. The focus here lies on the people working within the company and the generated revenue plays no substantial role. Therefore this aspect will be ignored. (European Commission, 2003) Consequently, if it is referred to big companies, this number is taken as reference.

Master data is a subset of the data which is the necessary foundation of all processes in a company. This subset is the one this research is focusing on since it is the crucial base of information that all processes make use of. The following graphic displays fields of master data with the example of an invoice.

Figure 1: *Example for Master Data Elements*



If companies grow and/or merge, the data landscape gets more complicated and managing data becomes an issue that is hard to deal with. It grows in a heterogeneous and inconsistent manner, which can cause big problems.

It is general consensus and common sense that correct, available and timely data is of great importance and can be a competitive advantage (Borghoff & Pareschi, 1997; Kahn, Strong, & Wang, 2002; Otto & Hüner, 2009). However, many companies have insufficient data management strategies. Especially bigger companies struggle with the huge amount of data and have no sufficient strategy to exploit the data. (Davenport & Prusak, 2000; Otto & Hüner, 2009)

The objectives of this thesis project are of both a practical and an academic nature. From a corporate point of view, the objective is to give organizations the possibility to assess their own MDM maturity and benchmark against other organizations. This helps the firm figuring out improvement areas to become more efficient. From an academic point of view, the objective is to contribute to the body of knowledge on the field of master data management maturity assessment because little research was conducted on this particular topic so far yet it has a remarkable practical relevance.

1.2 Research Questions

Deriving from the previously described problem statement, the research question can be formulated as:

RQ *How can a company's current state in Master Data Management be measured to identify potential improvement areas?*

This research question aims to answer the questions of how master data management can be dealt with and how the current status and improvements can be measured. This question leads to the following sub-questions:

SQ1 *How can an organization's master data management maturity be assessed and provide a way for different organizations to be compared with each other regarding their master data management?*

SQ2 *What should such a maturity assessment look like?*

SQ2 *How could the organizational implementation of such a tool in an existing organization be undertaken?*

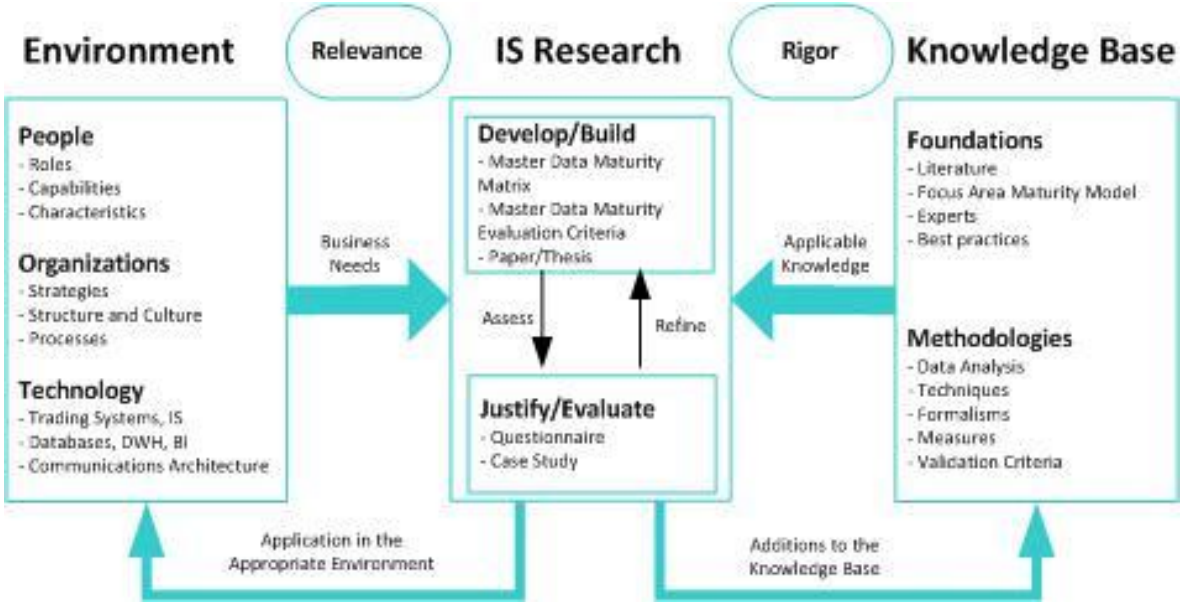
1.3 Research Approach

The "Design Science in Information Systems" approach that will be used in this research is based on the assumptions that information systems are designed to improve effectiveness and efficiency in the area where it is utilized. But the success of such a system does not only depend on the technical implementation. It is also highly dependent on the environment, which consists of people, organizational structures and technology within the company. The acquisition of knowledge in this context requires two different but mutually supportive paradigms; behavioral science and design science. Behavioral science "explains or predicts organizational and human phenomena surrounding analysis, design, implementation, management, and use of information systems". This research discipline is dedicated to analyze the interaction between people, technology and organizations that need to be aligned in order to make the implementation successful. Design science is meant to create innovations that define ideas, practices, technical capabilities and products which are used to achieve analysis, design, implementation, management, and use of information systems. Actually designing is important due to the fact that it stimulates critical thinking. (Hevner, March, & Park, 2004)

Figure 2 depicts the framework of the design science in information systems approach. It is based on the assumption that information systems and organizations are "complex, artificial and purposefully designed" (Hevner et al., 2004). They are composed of people, structures, technologies, and work systems. The information systems (IS) have the purpose of allocating resources to achieve a set goal of the organization. For this purpose, there are crucial alignments to be made between business and IT strategy and between organizational and IS infrastructure, so research in this discipline must also

address the interaction between the components. (Henderson & Venkatraman, 1993; Hevner et al., 2004)

Figure 2 *Information Systems Research Framework (Hevner et al., 2004)*



The environment scopes the problem space in which the area of interest resides in (Simon, 1996). This comprises the goals, tasks, problems and opportunities that define the business needs as they are perceived by the employees. This perception is influenced by the roles, capabilities and characteristics of the people. Business needs are captured and evaluated with regard to the company's strategies, culture and present processes. IS research is accomplished in two balanced phases. The behavioral science phase is about developing and justifying theories that explain and/or predict phenomena in the research scope. The design science phase is concerned with building and evaluating artifacts which are developed to address the business needs. In combination, they are used to deliver truth and utility. The knowledge base is the theoretical foundation on which the IS research is built on. The former research supplies theories, frameworks, etc which can be

used in the development phase. Rigor is accomplished by the adequate application of foundations and methodologies. (Hevner et al., 2004)

IT artifacts that fall in the scope of designing are constructs, models, methods and instantiations. Constructs contain vocabulary or symbols that are used to define a language to define problems and solutions. (Hevner et al., 2004) Models are abstractions and representations which use the constructs to depict an actual real world situation; thus the design problem and the solution space (Simon 1996). Methods contain algorithms and practices. They define processes and guidance on how to solve problems. (Hevner et al., 2004) Instantiations form the implemented prototype systems. They are meant to prove the feasibility of the artifact in a working environment. (Hevner et al., 2004)

The Design Science Research approach is suited very well in the context of the research. This is also stated in Becker, Knackstedt & Pöppelbuß (2009), who employed a procedure mode for the development of maturity models in IT management. They also base their research on Hevner et al. (2004). The research takes place in the field of Information Science and the area to be investigated is in the scope of information systems. There will be a literature study for thorough grounding on the topics of master data, master data management, data quality and related topics. Furthermore, there will be a developing of an artifact, i.e. a maturity assessment model for master data management. The literature base will form the foundation of the model development. In the case company, the model will be applied and evaluated according to academic standards.

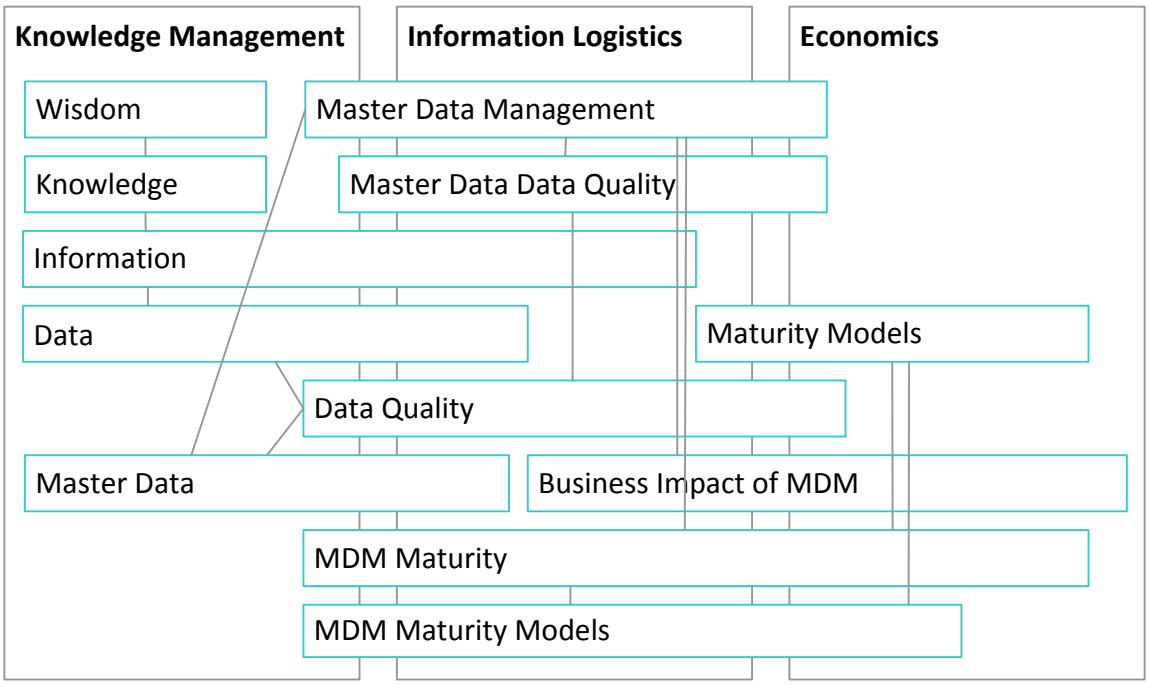
To ensure validity of the research, it will be evaluated based on (Yin, 2003). In his research, he proposes a tactic with tests to ensure validity. The appropriate ones for this research are presented in the following. The first is ‘Construct Validity’ which is about

setting up correct measures for the studied subject. ‘External Validity’ is about defining a scope in which this research’s results can be generalized. ‘Reliability’ ensures that the study can be repeated with the same results. (Yin, 2003)

2 Literature Study

In this thesis, the research areas are interdisciplinary. This research can roughly be classified into the areas of knowledge management, information logistics and economics. The topic of master data management with the economic focus has roots in many different academic fields of interest. The next diagram gives an overview about the broad research areas.

Figure 3 *Interdisciplinary Research Areas of the thesis*



The research focus is part of the concept of information logistics. According to Winter et al., (2008) the following topics are part of this concept; meta data management, (data) quality management, master data management, and data protection and data safety management. Information logistics is a concept that has an infrastructural, a strategic, and a procedural character. It is inter alia dedicated at providing and using data. Thus, it

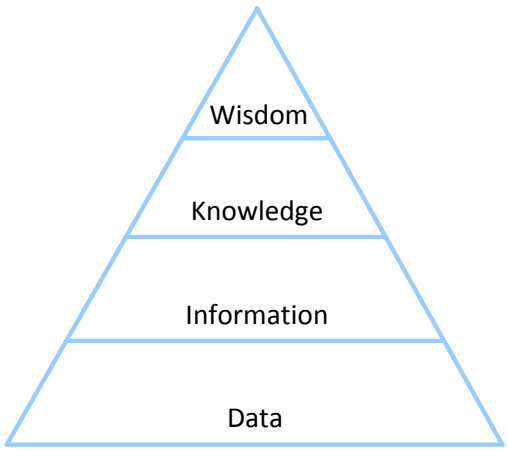
is advisable to view the supportive processes not isolated, but in line with an infrastructure. Infrastructure is the part that provides the necessary organizational and technical requirements for implementing an information system. It is not only about hardware and software but also about processes and all process knowledge which are needed for planning, regulation, control, maintenance and support of itself. There is also the need for standards and procedures which enable an efficient business operation. The strategic aspect derives from the fact that in this focus, long term decisions have to be made which affect large parts of the business. The process character is addressed because the focus is also on information systems that are integrated in processes. Analyzed data is used in the processes to help decision making in the operative business. (Winter et al., 2008)

The specialties in this area lead to some challenges for the organizations. Projects within that scope have to face some risks. The bigger and the more complex the projects get, the higher the risk will be. The more the area is defined by infrastructural topics, the more dependencies and possibly mutually exclusive requirements have to be considered and the more alignment with other units has to be done. Infrastructure is used by several applications, which on the one hand makes it possible to scale the costs, but on the other hand difficult to allocate the costs to a certain cost unit. Also the value of improved infrastructure is hard to quantify. Companies often conduct infrastructure projects on a local basis. Single units have their own strategies which cannot bring full benefit because they cannot be seized by the whole organization and lead to desired synergies. The goals are set for reaching local objectives, not organization wide improvements. This has to be encountered with overlapping resource management and clear communication. The

information logistic has to be embedded in the company wide IT architecture. The goal is to harmonize the organizational structure by seamlessly embedding the information logistics and the IT infrastructure. (Winter et al., 2008)

Since the topic of master data belongs to the research area of knowledge management, I will briefly introduce the concepts of data, information, knowledge and wisdom. There are definitions from various points of view; the sciences of philosophy, biology, psychology, pedagogies, social sciences and many more have their own definitions of those terms that are related to their field of study. However, the scope is a scientific one in a business environment, so definitions emphasizing that focus will be given. A famous structuring of the terms was first done by Ackoff (1989). The picture below depicts the pyramid of Ackoff. It shows how the concepts build upon each other and one could not exist without the founding one. The single terms are described in the following paragraphs.

Figure 4 *Knowledge Pyramid (Ackoff, 1989)*



2.1 Data

Data is the basic description of "things, events, activities and transactions that are recorded, classified, and stored" but not organized to derive a certain meaning. (Turban, Volonino, & Wetherbe, 2009) Davenport and Prusak (2000) define data as "a set of discrete, objective facts about events". From the business perspective, they describe it as "structured records of transactions". Data does not provide any information about reasons or give any help in predicting anything. In this state, data has no meaning or value since it has no context. It would need an informative description. (Davenport & Prusak, 2000; Jashapara, 2004) Data consists of characters; atomic elements that are assembled to form a piece of data like singular letters form words (Rehäuser & Krcmar, 1996).

2.2 Information

Information is data in context, ergo usable data (Van der Linden, 2009). It is the value of data plus the context of the value. As a practical example, 'AAA' is just a sequence of letters, but with the context it shows that it is a scale of a rating agency for creditworthiness of countries. Information is therefore data – which is by its nature of no or little relevance – enriched with relevance and purpose. So, "data becomes information when its creator adds meaning". (Davenport & Prusak, 2000)

To explain the concept of information, Davenport and Prusak use the metaphor of a "message". Information is a message in the form of a document or some kind of audible/visible communication. All kinds of messages need a sender and a receiver to be complete. So information is the entity that makes the difference. By receiving the message, the receiver will change his perception concerning a particular circumstance. (Davenport & Prusak, 2000)

This turning of data will be conducted through adding value in at least one of the following ways (Davenport & Prusak, 2000):

- *Contextualization*: it is known what purpose the data was collected for
- *Categorization*: Units of analysis or key components of the data are known
- *Calculation*: there might have been a mathematical or statistical evaluation of the data
- *Correction*: possible errors have been removed
- *Condensation*: there might have been a summarization of the data.

In this research area, the terms are often used interchangeably because a distinction is not crucial and people can usually distinguish between both terms intuitively (Otto & Hüner, 2009; Van der Linden, 2009). Therefore, the terms are used like synonyms in this research as well especially because there was a clear distinction made and the reader is now able to see the difference in case it might be necessary. In daily life of organizations, the terms are used interchangeably. (Pipino, Lee, & Wang, 2002; Rehäuser & Krcmar, 1996; Wang, 1998)

2.3 Knowledge and Wisdom

For the sake of completeness, the definitions of knowledge and wisdom will be presented as well.

> *"Knowledge is a fluid mix of framed experience, values, contextual information, and expert insight that provides a framework for evaluating and incorporating new experiences and information. It originates and is applied in the minds of knowers. In organizations, it often becomes embedded not only in documents or*

repositories but also in organizational routines, processes, practices, and norms.“ (Davenport & Prusak, 2000)

That is how Davenport and Prusak define knowledge. A shorter explanation comes from Van der Linden (2009), saying that knowledge is information in context combined with an understanding of significance of the information.

In general, it is clear for everyone that knowledge goes beyond information. It is deeper and richer than information. That can be derived from the fact that people can be considered as knowledgeable, but things as computers or books would not - even though they could have been the source of someone having much knowledge. Knowledge is rather complex. It is structured but also fluid and intuitive. Knowledge is something that exists in people, not in any artificial systems. It derives from information which derives from data. The transition from information to knowledge can happen through the following actions (Davenport & Prusak, 2000):

- *Comparison*: Comparing information about one situation with information about others
- *Consequences*: what decisions and actions to be taken come with the information
- *Connections*: in what way does that information relate to other pieces of information
- *Conversation*: what is other people’s opinion about this information

Knowledge is something that is created within and between humans. It gets delivered through media like books, documents and also through interpersonal communication. It is therefore closer to actions to be taken than information or data. (Davenport & Prusak, 2000)

Knowledge can be classified in different ways. A common distinction is tacit and explicit knowledge with explicit being expressible knowledge and tacit being more internal intelligence. (Van der Linden, 2009) Explicit knowledge is knowledge that can be formalized. In an organization, it is that knowledge stored and made accessible in different kind of documents. Tacit knowledge is embedded in the individuals coming from their experience and can be communicated via direct contact. The communication of tacit knowledge results in tacit knowledge of the receiver, whereas the acquisition of explicit knowledge is indirect and results in tacit knowledge as well. The first case is called socialization, the latter one internalization. A conversion from tacit to explicit knowledge is defined as externalization and the direction explicit to explicit is referred to as combination. The transfer of knowledge can be seen as a cycle starting with socialization and ending with internalization.(Borghoff & Pareschi, 1997; Nonaka & Takeuchi, 1995)

Knowledge – as the consequences from data and information – is very valuable to organizations. Companies rely on the knowledge that has been developed over time. The ability of deriving consequences from information goes hand in hand with experiences; therefore experienced workers are very precious to the company. However, it is intangible and very difficult to measure. (Davenport & Prusak, 2000) Knowledge is the factor enabling successful companies to exceed the sum of the visible assets in the balance sheet with their market value. (Borghoff & Pareschi, 1997) In order to keep the knowledge up to date and useful, it needs to be communicated. Only if knowledge is transferred among the people, can it generate new knowledge and not get outdated.

Wisdom is – in contrast to the philosophic view – the actual application of knowledge. Wisdom is utilized to decide on the right actions to be taken.

Knowledge and Wisdom fall out of the scope of this research since master data is clearly within the scope of data and information and only the application of the data belongs to the definition areas of knowledge and wisdom. However, this is not the focus of this thesis research.

2.4 Master Data

This chapter contains a comparison of different definitions of master data to finally present a derived definition on which this research is based on. There will be both definitions from academia and definitions from practice to ensure a definition that suits the needs of an academic work and of a deliverable to be used in business.

Master Data is the data describing the most relevant business entities, on which the activities of an organization are based on, e.g. counterparties, products or employees. In contrast to transactional data (invoices, orders, etc.) and inventory data, master data are oriented towards the attributes. They describe the main characteristics of objects in the real world. Single master data entities are rarely being changed, for instance the properties of some kind of material. Instances of master data classes are relatively constant, especially if they are compared with transactional data. Master data is the reference for transactional data. There would not be a single order or delivery without master data. (Otto & Hüner, 2009)

Microsoft describes Master Data the following way; they are the critical business data covering the four categories people, things, places and concepts. People is further subcategorized into customers, employees and salespersons. Under things fall products,

parts, stores and assets. Concerning concepts, there are contracts, warrantees and licenses to be found. Places comprise locations and geographic divisions. Furthermore, they describe master data as often being related to transactional data, whereas master data display the nouns and transactional data display the verbs; e.g. when a vendor sells a product, the vendor and the product belong to the group of master data and the act of selling is part of transactional data. Master data tends to be not very volatile. Their lifetime is considered longer as of other data types. Master data are very valuable to a company. Master data must be reused across many applications. (Wolter & Haselden, 2006)

BeyeNetwork defines Master Data as follows. Master Data is the aggregation of reference data, enterprise structure data and transaction structure data. From their perspective, reference data is that kind of data that is used to categorize other data or relating data in a database. Enterprise structure data is data that helps reporting business activity. Transaction structure data are those data that are precondition for transactions to actually happen. This contains categories like product and customer. It represents the participants in a transaction. (Chisholm, 2008)

The following table summarizes key attributes mentioned in different definitions.

Table 1 *Comparison of Master Data Definitions*

Criteria	(Otto & Hüner, 2009)	Microsoft	BeyeNetwork
Critical for business	√	√	√
Long lifetime	√		√
Reference for transactional data	√	√	√
Reuse across different systems	√	√	√

The definition that this research is based on includes the following criteria. Data is defined as master data if it is business-critical, serves as reference for transactional data, has a long lifetime and is used across several different systems.

Master data are important for the systems of an organization. It is the basic fabric the organization is working with. In case of any inconsistencies in the data, there are side effects for business and IT. It might affect operational efficiencies, customer relations and infrastructural optimizations. (Kumar, 2010)

Master data is important for the systems of an organization. It is the basic fabric the organization is working with. In case of any inconsistencies in the data, there are side effects for business and IT. It might affect operational efficiencies, customer relations and infrastructural optimizations. (Kumar, 2010)

There is often a lot of confusion about the terminology. Sometimes one hears the terms reference data or Meta data and the same meaning is being intended. Wegener

(2007) distinguishes the terms as follows. Master data is the overall term containing all other concepts. Meta data is a sub group of that which then contains reference data. Since all the concepts fall within the scope of master data, there will be a short definition of the terms.

Meta data are that kind of data which are used in a development or a business supporting process for adapting products. As the word meta indicates, they contain data about data. Reference data are reusable, easily structured meta data with the purpose of conducting changes in the technical context automatically. Reference data represents the context which needs to be considered. Examples for reference data are currencies, risk kinds, places, customers or products. Master data are data that are reused in different contexts to conduct changes automatically in a professional environment. This is the category also containing data without a natural lifecycle like interest rates. (Wegener, 2008)

To summarize, master data focus on automatic adoption of changes and reuse. They are not involved with the lifecycle of products etc and the usage is mainly in the operating business. Meta data is partly used for automatic adoption of changes, it is rarely reflected and it uses the lifecycle. It is mainly used in the development and also the operating business. Reference data is also used for automatic adoption and focuses on reuse and the lifecycle and it is mainly used in the operating business. However, they all aim at improving the realization of changes. (Wegener, 2008)

2.5 Data Quality

In this focus, it makes sense not to concentrate on the sole accuracy of data, but also to consider other criteria about data. (Van der Linden, 2009) It is consensus in

research that data quality is about information/data which is characterized by the fitting to the consumers' needs. In literature, this concept is mainly referred to as "fitness for use". (Huang, Lee, & Wang, 1998; Kahn et al., 2002; Pipino et al., 2002)

The German Society for Information and Data Quality published a definition of information criteria. (Malzahn, 2008) They ordered data quality in four different categories (Huang et al., 1998; R. Y. Wang, 1998): intrinsic data quality to encourage confidence in correctness and objectivity of data; contextual data quality to make sure the data is relevant at the time of use; representational data to support understandability of data and accessible data quality for assuring that the data consumer knows how data can be accessed. The single characteristics are self-explanatory and can be seen in the following table.

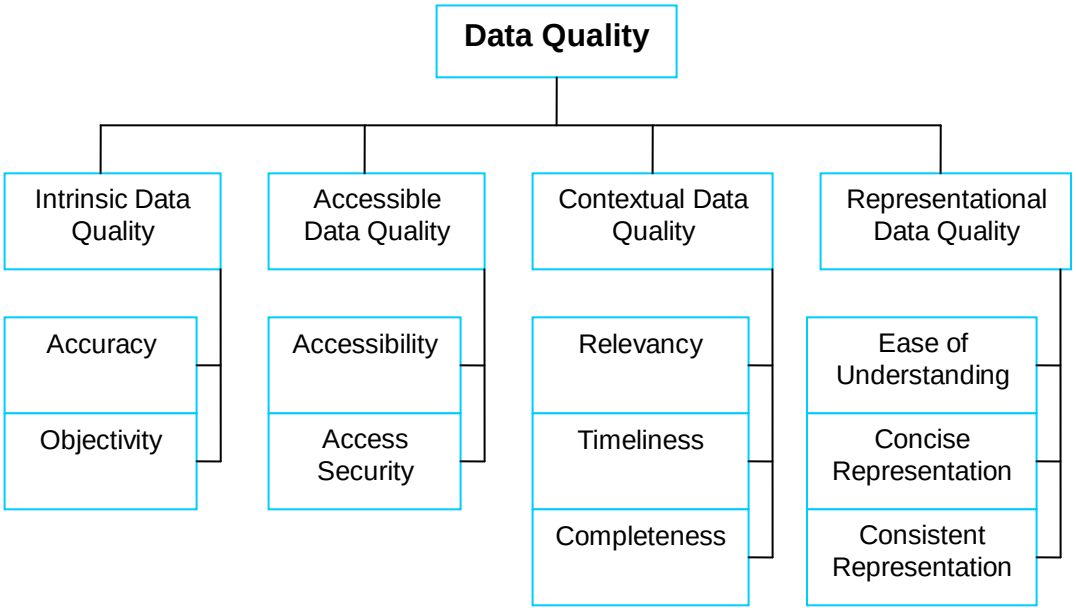
Table 2 *Data Quality Criteria*

Intrinsic DQ	Contextual DQ	Representational DQ	Accessibility DQ
Believability	Value-added	Interpretability	Accessibility
Accuracy	Relevancy	Ease of understanding	Access security
Objectivity	Timeliness	Representational consistency	
Reputation	Completeness	Concise representation	
	Appropriate amount of data		

These aspects represent an all-embracing overview of data quality in a broad sense. In the context of data quality in electronic systems, not all of the characteristics are necessary. The aspect believability is not relevant. Believability is clearly given since the data is one of the main assets an organization uses for daily business and gathers it from reliable sources. The same argument counts for the aspect reputation. The characteristics value-added and appropriate amount of data do not apply as well because the amount of data is clearly predefined and the added value is not subject to discussion because the data is substantial to the business. Furthermore, there is no room for interpretability in master data.

Some of the above mentioned criteria are not relevant for the scope of this project as depicted in the precedent paragraph. Therefore we use in this context the framework of Van der Linden (2009) to depict the dimensions of data quality that are relevant in this sense. This framework was developed to investigate the effects of data quality of organizations, so it was developed to fit the scope better. The criteria base on the above mentioned definition by (Huang et al., 1998; Kahn et al., 2002).

Figure 5 *Information Quality Framework (Van der Linden, 2009)*



The framework divides quality of information or data into 4 categories; intrinsic, accessible, contextual and representational data quality. These categories are subcategorized into two or three characteristics. Accuracy is what is often referred to as quality, i.e. the extent to which the data is correct and error free. Objectivity describes the degree to which the data is unbiased and clear. These factors belong to the intrinsic category. The degree of availability and the possibility of easy and fast retrieval are addressed by accessibility. Security talks about access restrictions to ensure only authorized access. These are part of the accessible data quality. Relevancy is used to express the applicability and usefulness of data in a context. Timeliness means that the data is sufficiently prevailing. Completeness describes how sufficiently broad and deep data is to be used for a certain purpose and that no relevant parts are missing. The three mentioned aspects are contextual. Ease of understanding describes the degree to which the data is clearly understandable without ambiguity. The data is to the point and not

overwhelming is categorized under the aspect conciseness. Whether data is presented in the same format and in compatibility to previous data is collected in the term consistency. These three fall under the category of representational data quality

Good master data serves a purpose. This is why the quality is often defined as the fitness for use. Concrete quality criteria often build upon e.g. consistency and completeness. Consistency is defined as the degree to which the data values within different redundant data sources stay equivalent. Completeness measures the degree to which the data is gathered. (Otto & Hüner, 2009)

In literature, a single source of truth is often referred to as a quality enabler. If there is a high trust in the repository, the incentive of establishing a single source that feeds the others is given. Otherwise, one source could cause false data. Data quality can also indirectly be improved by persuading the employees that data is a valuable asset that improves the business. (Loshin, 2008)

2.6 Data Quality Benchmarks

It has been widely investigated how data or information quality can be classified. From a practical perspective, there is the need to assess the quality in a measurable manner. These benchmarks enable organizations to compare the information quality within the organization and to monitor quality improvements. The term fitness for use communicates the core meaning, but still quality is hard to measure. (Kahn et al., 2002)

For this purpose, a number of benchmarking methods exist. The one presented here was developed by Kahn et al. (2002). It makes use of the information quality characteristics that were also the foundation for the information quality framework

described in the chapter before. (Van der Linden, 2009) The information quality characteristics are widely used and agreed upon.

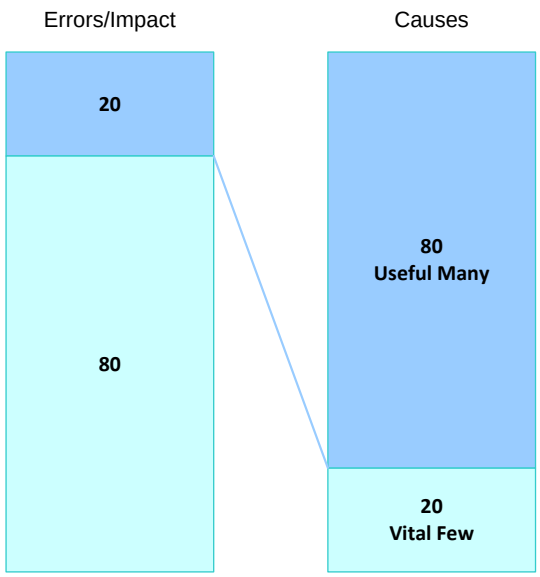
Table 3 *Information Quality Model mapped with IQ dimensions (Kahn et al., 2002)*

	Conforms to Specifications	Meets or Exceeds Consumer Expectation
Product Quality	**Sound Information:** Free-of Error Concise Representation Completeness Consistent Representation	**Useful Information:** *(Appropriate Amount)* Relevancy Understandability *(Interpretability)* Objectivity
Service Quality	**Dependable Information:** Timeliness Security	**Usable Information:** Believability Accessibility *(Ease of Manipulation)* *(Reputation)* *Value-Added)*

An approach to benchmark quality, widely used across different application areas is the statistical phenomenon which is called Pareto principle. It is also applied in quality measurements. It states that 20 percent of reasons cause 80 percent of the errors and vice versa. The 20 percent of reasons are called vital few, and the remaining 80 useful many.

(Schneider, Geiger, & Scheuring, 2008, p. 165) This shows that it will need relatively small effort to reach a decent quality of data, but it will need unequally more effort to reach perfection. This principle is applied in many fields where the relative proportion of cause and impact is relevant, e.g. in time management: 80 percent of the tasks can be done in 20 percent of the time. This kind of cause-impact view can help giving a good overview about the most important sources of errors. In 2002, Microsoft figured out that the Pareto principle applies to both features and bugs. Approximately 20 percent of the bugs caused 80 percent of the errors, and even more remarkable, one percent of the bugs were responsible for 50 percent of the errors. (Rooney, 2003)

Figure 6 *Pareto Principle (Rooney, 2003)*



For data quality, the empirical measurement is rather complicated. It can quite easily be measured if the data adheres to quality in the sense of correctness and availability and readability at the time of access. To measure an organization's data quality, the following binary approach should be enough: If the data is available and

correct at the time of access, it is considered as maximum quality. Any other state will be considered as minimum quality. This can be summed up for all data or for a defined control sample. From the sum, a quality percentage can be derived. Certainly, in some contexts, a closer look at different facets of quality must be taken, but in this context, it is enough to just give a percentage and look at the origins a bit closer.

3 Master Data Management

Some companies simply pile up masses of data with the expectation of gaining benefits from this. This expectation will not come true because the pure existence of data will not lead to any virtues. Even worse, it gets more and more complicated to find the particular relevant piece of data if it is somewhere among huge amounts of unmaintained data. The pure possession of data will not lead to anything if there is no logical structure that makes it possible to mine data according to criteria. Data only describes facts, there is still the lack of a judging, interpreting or action triggering dimension. (Davenport & Prusak, 2000)

Master data management is a complex task. This comes from the complexity of a master repository of this kind of key data which is integrated through a service layer with applications throughout all the company. (Loshin, 2008) Most master data is used everywhere in the organization. Companies have different databases which feed into various systems. The relations between possibly more than hundreds of systems are highly complex and it is not easy to see where the data comes from that is used in one application. It is very unlikely that people are so experienced to have an overview about all the data in the company.

MDM can neither be conducted only in the IT department nor only on the business side. The business' functional knowledge is needed but also the IT is responsible for displaying the entities in the information systems. (Otto & Hüner, 2009) The IT and the business side must understand each other so the business can communicate their needs and the IT can find solutions to provide. The task of MDM is to ensure timeliness, clarity and relevance of the provided data. It must be ensured that the data is available

when needed, it must be exactly what is needed and the recipient must be able to make use of it. (Davenport & Prusak, 2000) MDM needs strategic planning and the support of key stakeholders for success. It needs to be led by the business and be conducted jointly by business and IT. (Kumar, 2010) There must be responsible people forcing this topic to be addressed in order not to get forgotten.

Problems with MDM can occur due to its complexity. A responsible person is confronted with a highly complex topic on the one hand and on the other hand, he is also overwhelmed with the saturated offerings of software solutions regarding this topic. (Otto & Hüner, 2009) The organization has to be prepared for the technical, operational and managerial changes that will occur throughout the process of MDM. (Loshin, 2008)

Master data management is often not a sovereign management task, but just a sub area of business units. There is one person responsible for master data in his scope, but that person is not informed about data in other processes, even though the data might be the same and even changed in other processes. (Otto & Hüner, 2009)

3.1 Impact of Poor MDM

Master data of sufficient quality is the prerequisite for the performance potential of the company. If master data is not consistent, companies will not be able to react to upcoming bureaucratic obligations and they will not be able to set up an effective reporting process or to harmonize business processes in general. (Otto & Hüner, 2009) As can be seen, the impact of poor data management could be dramatic. These impacts, as business impacts in general, are usually expressed in monetary terms. So, the logical consequence of poor master data can be monetary loss to a variable extent.

In Van der Linden (2009) a retrospective case study is presented which mentions 11 different impacts of poor data management. It is assumed that these can mostly be applied to master data management as well. These impacts are mentioned in the following table:

Table 4 *Impacts of Poor Data Management (Van der Linden, 2009)*

Business Impact	Description
Lost Sales Opportunities	Possibilities of acquiring new customers or making existing customers buy new products are not seized because of poor data management.
Customer Service Costs	Costs that are spent for correcting wrong data; both actually changing the data set and finding the correct values.
Customer Dissatisfaction	Customers are dissatisfied because of errors in data quality concerning them. As a result they terminate relationships and possibly advise others to do the same.
Lost Revenue	Costs for wrong invoices or inability to bill customers due to errors in the data.
Operational Inefficiencies	Inefficiencies in processes because of poor data can have many impacts including insufficient resource planning or the inability to react to changes in time, increased workload etc.
Delays in System/Project Deployment	Projects or new system introductions get delayed or cancelled because of bad data quality.

Regulatory Compliance	This implies costs that result from a company's inability of adhering to regulatory compliances. It can result in liability risks or even the whole company's life.
Poor Decision Making	Making wrong decisions because of wrong or outdated data serving as a source for business intelligence and decision support systems.
Lost Business Opportunities	The missed chance of buying resources at a cheaper price because of an insufficient overview about the market and the prices.
Employee Morale	If bad data quality leads to frustrated employees either because they cannot work as productively as they would like to or because they have to correct faulty data instead of working on their primary task.
System Credibility	Low trust in general data will bring departments to setting up own data sources, which leads to multiple data sources with different values.

The impacts will be further elaborated on in the following paragraphs.

1) Lost Sales Opportunities

Due to poor data management, opportunities for sales (cross-selling of products, missing ability to identify trends or capability to analyze customer needs) might get lost, e.g. wrong customer data mining leads to incorrect target group analysis and thus to eventually failing marketing approaches. If activities fail due to problems with the data

management, the costs of failed campaigns and the resulting missed customer value can be allocated to poor data management to give an overview about the costs that this produces. (Van der Linden, 2009)

2) Customer Service Costs

Employees have to correct data on a regular basis. Even though the costs for this kind of tasks may vary, it is still time consuming, so costs are created. Even though, this is not their main task, but employees are forced to correct the data to keep the business running. Customer service costs can easily be measured by simply allocating the time of an employee spent on this purpose as a percentaged value of the FTE multiplied with the salary. This calculation has to include both the employees changing the data and the ones finding the correct values. If special software is used to track wrong data, the costs for that have to be added to the calculation as well. (Van der Linden, 2009)

3) Customer Dissatisfaction

There are different reasons for dissatisfaction of customers which is therefore hard to measure. Dissatisfaction could originate in errors in names, addresses, billings or product information which can lead to various impacts. The indirect effects of spreading the bad experience with others who might not start a business with the company even if they would have wanted to, is impossible to estimate. At least the direct effect of a customer terminating business with the company is approximately computable by the means of the customer lifetime value, which is defined as the net profit or loss that one customer brings to the company throughout the whole life span of relation between both parties. The specific customer lifetime value highly depends on the business area, the kind of customer and relationship. More information on this topic can be found in Jain &

Singh (2002). Once the segmental customer lifetime value is calculated, it can be used to calculate the costs of customer dissatisfaction resulting in poor data quality. For this, the company has to find sources of data about customer complaints (e.g. the help desk). Data quality related complaints have to be filtered. This approach is a good opportunity to monetarily visualize the impact of poor data quality. (Van der Linden, 2009)

4) Lost Revenue

Lost revenue consists of costs that were produced because the low quality data led to errors for instance in the invoices or the inability to bill the customer at all. Van der Linden states a case study which indicates that more than 2/3 of all people asked said that they have problems with invoices on a regular basis. Most errors are corrected within a week, but for 15 % the correction time adds up to more than one month. The delay and the extra time spent on those problems can be measured easily. (Van der Linden, 2009)

5) Operational Inefficiencies

Inefficiencies in the operational work flow cause consequences like poor resource planning, the lack of ability to react in time to external developments, increasing workloads in systems and rising costs because of wrong or duplicate mails to customers. Costs for duplicate mailings are not only the direct costs for resending the correspondence, but also the hidden costs of finding the origin of this. If an employee has to stop working for the purpose of resolving errors, this is an incident related to data quality. The work process is being stopped while the data is corrected. If it is possible to reduce delays due to these circumstances, more transactions will be processed in greater volumes and at lower transaction costs. The impacts can be very different, depending on

the business area and also the costs strongly differ, therefore there is no standard way for all companies to calculate the costs. (Van der Linden, 2009)

6) Delays in System/Project Deployment

This field contains delays or cancellation of new products or system introductions because of bad data quality. Improving information quality can ensure that projects will not get delayed or cancelled because of bad data. If still projects from the area of business intelligence (BI) or data warehousing (DWH) fail, then the reason would not be data quality. The risks and costs are not predictable. Projects succeed or fail, but it can be assumed that low quality data definitely is a burden for success. By improving the data, the atmosphere for new projects is more proactive. (Van der Linden, 2009)

7) Regulatory Compliance

Regulatory Compliance impacts come from an organization's inability to adhere to regulatory compliances. In some industries, like banking or (energy) trading, the impacts can be enormous and very high costs can result. This sub-item includes the ability to fulfill service level agreements (SLA). Failures in this area can have unpredictable results with a wide range. This depends on the industry and the kind of problem. It can result in liability issue or even risking the whole organization's existence. The impact will result in a penalty that depends on the magnitude of failure. (Van der Linden, 2009)

8) Poor Decision Making

If the business intelligence applications had better data, the company would have more appropriate information about the customers and therefore could have increased sales, better tailored services and valued customer retention just because of a better

understanding of the own customers and target groups. If the data is wrong or outdated, it leads to incorrect forecasts and decisions. In organizations it is common sense that there is a correlation between data quality and the quality of the analysis and decision support. It is hard to estimate the costs of poor decision making, but it is clear that it can substantially harm a business in the worst case. (Van der Linden, 2009)

9) Lost Business Opportunities

Lost business opportunities are partially related to customer opportunities but here, the focus is on the upstream relationships. Costs are produced as a consequence of missing business to business opportunities because of incorrectly analyzed internal and external data. If the company has an insufficient view on the company landscape, it misses opportunities of buying resources at cheaper prices. This risk comes with incomplete market data. The value of a better identification and monitoring of saving opportunities is clearly less expenditures which can be achieved through more reasonable procurement. The precise value changes for every organization. (Van der Linden, 2009)

10) Employee Morale

This term is used if costs are created because of sinking motivation and rising dissatisfaction of the employee due to bad data quality. The employees have to correct data and cannot work on their primary tasks or they are less productive because of the poor quality of the data. Direct costs of this impact are the decrease in productivity as a percentage of the wage. Indirect costs derive from the possible staff attrition resulting from the situation. Then the costs for hiring and training new employees are partially related to the bad data quality. (Van der Linden, 2009)

11) System Credibility

As a consequence of low trust of the employees in company data and data of other departments, the different departments build their own databases and spreadsheets which they keep to themselves and fill in by themselves. This can have the impact of multiple systems containing the same data sets but with different values. Having different values for the same data can result in many of the earlier mentioned impacts. The costs for this impact are those for running the different systems and the inefficiency of multiple sources for the same data. They can be calculated by adding up the costs for running the systems and the costs for time and money spent by an employee to collect data and maintain it in the system. (Van der Linden, 2009)

The following table compares the impact on general data to master data along the definition of master data that was derived beforehand.

Table 5 *Comparison of Business Impact for DM and MDM*

	DM	MDM
Lost Sales Opportunities	Possibilities of acquiring new customers or making existing customers buy new products are not seized because of poor data management.	Since this includes the use of product and customer data, which clearly is part of master data, this impact applies to MDM as well.
Customer Service Costs	Costs that are spent for correcting wrong data; both actually changing the data set	This includes correcting customer data, which is a core part of master data.

	and finding the correct values.	
Customer Dissatisfaction	Customers are dissatisfied because of errors in data quality concerning them. As a result they terminate relationships and possibly advise others to do the same.	This also includes customer data with the impact of customers being unsatisfied.
Lost Revenue	Costs for wrong invoices or inability to bill customers due to errors in the data.	Wrong invoices or billing issues may result on wrong customer or account data, which belong to master data.
Operational Inefficiencies	Inefficiencies in processes because of poor data can have many impacts including insufficient resource planning or the inability to react to changes in time, increased workload etc.	Operational inefficiencies result from inter alia redundancies or duplicates in data sources, including master data, e.g. no updates of customer data in a redundant system. Therefore, within the scope.
Delays in System/Project Deployment	Projects or new system introductions get delayed or cancelled because of bad data quality.	The successful introduction of system relies on data; usually the systems are fed with master data.
Regulatory	This implies costs that result	Regulatory compliance issues

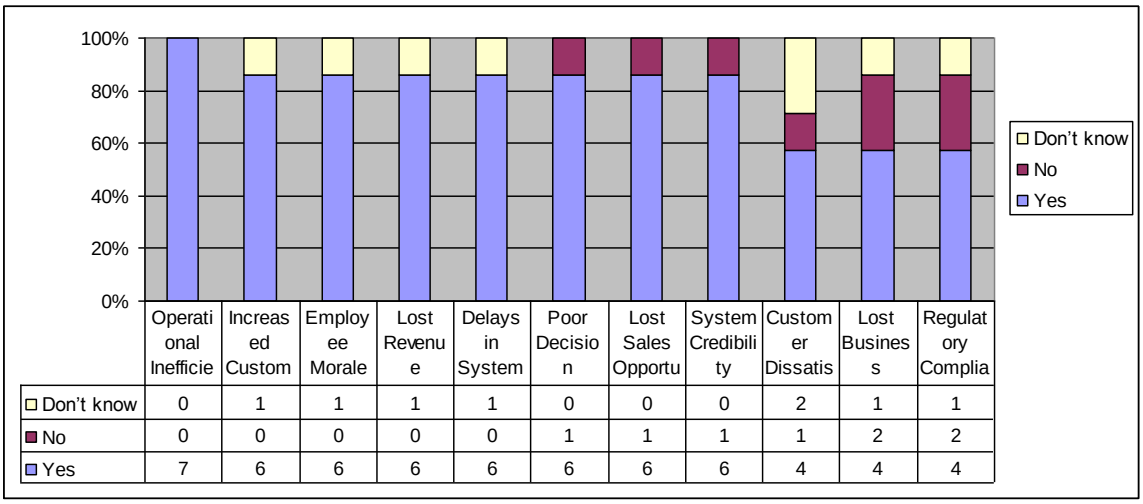
Compliance	from a company's inability of adhering to regulatory compliances. It can result in liability risks or even the whole company's life.	transitively result from poor master data. If regulations/SLAs are not adhered to, it might be because of incapability due to wrong master data as a source.
Poor Decision Making	Making wrong decisions because of wrong or outdated data serving as a source for business intelligence and decision support systems.	Data sourcing BI solutions are at least partly master data.
Lost Business Opportunities	The missed chance of buying resources at a cheaper price because of an insufficient overview about the market and the prices.	Master data about customers, vendors and markets feed the dynamic systems to find cheap prices and trading opportunities.
Employee Morale	If bad data quality leads to frustrated employees either because they cannot work as productively as they would like to or because they have to correct faulty data instead of working on their primary task.	If employees work with faulty data, this applies to master data as well. Constantly having to correct data impacts morale, regardless if it is master data or not.
System	Low trust in general data will	If there is low trust in data

Credibility	bring departments to setting up own data sources, which leads to multiple data sources with different values.	sources, regardless of their content, business units will set up own sources which are very likely to contain master data.

3.1.1 Survey about Business Impact

To ensure the applicability for master data management, a survey was conducted at E.ON Energy Trading among employees who actually work with master data in their daily life. The interviewees are all experts on the field of master data. Their daily tasks contain maintenance of data from a business-oriented view. They have a thorough understanding of the characteristics and impact of data issues. A group of 11 people were asked to answer the questionnaire. The questionnaire was rather simple. It contained the 11 impact factors and three choices to indicate whether the factor impacts the business or not or if the person does not know. The questionnaire can be found in the Appendix A. It was constructed as an online form to fill in. The results were automatically inserted in a spread sheet. From the 11 requests, 7 people actually replied. The results are depicted in the following table sorted by relevance.

Table 6 *Survey results about impacts of poor MDM*



	Operational Inefficie	Increased Custom	Employee Morale	Lost Revenue	Delays in System	Poor Decision	Lost Sales Opportu	System Credibility	Customer Dissatis	Lost Business	Regulatory Complia
Don't know	0	1	1	1	1	0	0	0	2	1	1
No	0	0	0	0	0	1	1	1	1	2	2
Yes	7	6	6	6	6	6	6	6	4	4	4

As can be seen from the results of the survey, the business impacts seem to be relevant from the perspective of expert users. The most significant impact is caused by operational inefficiencies. 100 percent of the interviewed people believe that operational inefficiencies are caused by insufficient MDM. 85 percent think that bad MDM causes increased customer service costs, impacts the employees' morale, causes loss in revenue, and delays in the systems or project development. 15 percent are not sure about the impact on these factors. For poor decision making, lost sales opportunities and system credibility, also 85 percent see MDM as the influencing factor. 15 percent do not think that these factors rely on poor MDM. 57 percent of the interviewees are convinced that lost business opportunities, customer dissatisfaction, and regulatory compliance result from unsatisfying MDM. 15 percent think that customer dissatisfaction cannot be derived from MDM issues and 28 percent are not sure about this. 28 percent are convinced that there is no relation between MDM and lost business opportunities and regulatory compliance and 15 percent are uncertain.

As the survey shows, for each characteristic, at least more than half of the questioned people are convinced that the characteristics are influenced by MDM.

3.2 Maturity Models

IT support for business processes is indispensable.(Becker et al., 2009; Müller, von Thienen, & Schröder, 2004) For almost every company, it is virtually impossible to run the business without the basis of IT applications. This becomes more and more crucial the bigger the companies are and the more sites they have in different locations. Since IT solutions are not optional anymore, an innovative one provides the organization with a competitive advantage. In order not to lose the competitive factor, a company always has to have a realistic view of how they are positioned in terms of IT capabilities and the quality of goods and services taking into account the organization's "goals, external requirements (customer demands, laws or guidelines) or benchmarks". (Becker et al., 2009)

An objective self-assessment is per se quite difficult; therefore it is handy to have some guidelines to work along. For this purpose maturity models are well-suited. They offer scaling for estimating the current position of the organization, providing criteria and characteristics to be achieved in order to achieve a certain level of maturity. A maturity model is a snapshot of an organization along some predefined criteria. They consist of a class of maturity levels for a number of entities. Those maturity levels represent a planned, wished or typical path of development for the entities ordered along stages. (Becker et al., 2009)

A lot of different maturity models can be found in literature. Each of them is suited well for one purpose, but less good for another one. In general, maturity models

describe the development of an entity over time. (Klimko, 2001) They contain sequentially ordered levels; on the bottom they depict the initial stage and on the highest state the level of full maturity in that particular focus area. The moving from the initial to the final stage displays the progress of regarding an organization's capabilities or process performance. (Becker et al., 2009; Klimko, 2001; Sacu & Spruit, 2011) The levels are characterized by certain prerequisites that need to be fulfilled for reaching a level. (Klimko, 2001) A maturity model can be used as an assessment of the current situation within the evolutionary path the organization is in, because it offers criteria and characteristics that have to be implemented to get to a certain level of maturity. (Sacu & Spruit, 2011)

There are different kinds of maturity models. It can at least be distinguished between the staged models, continuous models and focus area oriented models. The staged level models are characterized by the fact that there is a number of maturity levels defined. Within these levels, there are some focus areas determined that need to be fulfilled. The continuous level models have a number of maturity levels and a number of focus areas. Within each focus area, there is a fixed number of levels to be achieved. Focus area oriented models do not have a set number of maturity levels, but each focus area has a specific amount of maturity levels. (Steenbergen, Berg, & Brinkkemper, 2007)

Maturity models have some advantages that come in handy for the purpose of this research. The first one to mention is the simplicity. Due to their structure, they are easy to understand and explained to others. Additionally, they are suited for comparison. If the descriptions of the levels include some required characteristics, they can be used to rank

and compare entities. (Klimko, 2001) The application can be facilitated by the use of additional tools, for instance questionnaires. (Becker et al., 2009)

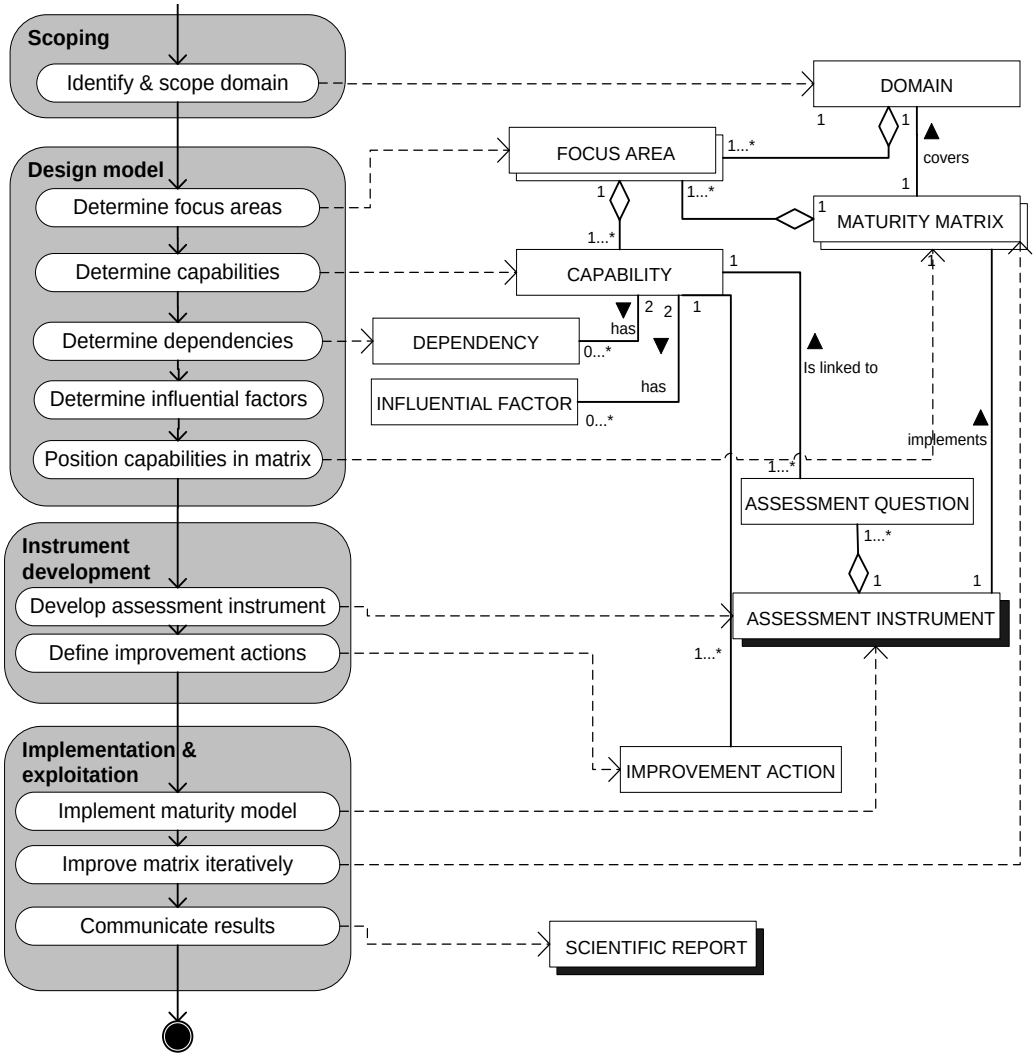
3.2.1 The focus area maturity model.

To assess the maturity of a company's master data management, the focus area maturity model (FAMM) approach is well suited. This functional domain is complex because of the huge amount of data and the amount of stakeholders in different departments. This implies that it is not possible to develop a fully mature process from the beginning. The whole MDM process needs to be improved incrementally. The FAMM approach supports the incremental progress of the functional domain. (Steenbergen, Bos, Brinkkemper, & Van De Weerd, n.d.)

The FAMM method builds upon the concept that for a functional domain there is a certain number of mutually disjoint focus areas. These focus areas have to be emerged to achieve maturity. They can be defined dynamically depending on the functional domain. If all focus areas reach maturity, one can speak of overall maturity. They make it possible to distinguish different levels of maturity within the functional domain. Within the areas, there will be a number of steps defined. These steps are natured as capabilities that rise in maturity gradually. (Steenbergen et al., n.d.)

Steenbergen et al. have developed a generic process to implement focus area maturity matrices (n.d.). Figure 7 depicts the process and the resulting deliverables (PDD). The notation that is being used for displaying this is called process-deliverable-diagram and is based on UML. This notation is suited for the purpose since it has the possibility of illustrating both the processes & activities and the resulting deliverables & results. (Van De Weerd & Brinkkemper, 2008)

Figure 7 *Focus area maturity model development process (Steenbergen et al., n.d.)*



On the left side, the boxes with the rounded corners illustrate the activities. An atomic activity can be grouped with others. This is depicted with the grey boxes. The black dot indicates the starting point of the process. The circled black dot indicates the end of the process. The dotted lines lead from the processes to the corresponding deliverables. The boxes with the sharp corners are the deliverables that will be produced during the process.

3.2.2 COBIT.

For the definition of the maturity levels, there are plenty of possibilities. Now, the levels according to the COBIT framework are presented. COBIT stands for “Control Objectives for Information and Related Technologies”. COBIT is a tool for controlling IT from a business perspective and is also used for compliance purposes(IT Governance Institute, n.d.)

They distinguish 6 levels of maturity, which are depicted in the following table.

Table 7 *Maturity Levels (IT Governance Institute, n.d.)*

Maturity level	Description
0	Non-existent
1	Initial
2	Repeatable
3	Defined process
4	Managed and measurable
5	Optimized

The levels have the following meanings. Level 0 indicates that there is total lack of any remarkable process related to the topic. The company does not realize there is an issue to be worked on. Level 1 is the one, where the company finds out that there is indeed an issue to be addressed. Still, there are no defined procedures and instead, there are ad hoc initiatives that are applied either on an individual or a case by case base. The whole approach is unorganized. On the second level, the processes are grown to a stage

with similar procedures that are carried out by different people who conduct the same task. However, there are no formal guidelines for standard procedures and no responsible roles. Mistakes are very likely because there is a high level of dependence on individual persons. On level 3, there are already non-sophisticated procedures that are standardized and documented and it is also communicated that those guidelines need to be obeyed. But there would not be a discovery of non-adherence. The fourth level is reached if the management is able to monitor and measure the adherence of the procedures and actually takes necessary steps in case it is necessary. The processes are advanced regularly. There is also a limited use of automation tool support. On the last maturity level, the processes are improved to a level of good practice, being grounded on the results of the last steps. There is a decent integrated use of IT to support and automate the workflow and tools are provided to improve the quality. (IT Governance Institute, n.d.)

3.3 MDM Maturity Models.

This chapter presents some master data management maturity models and then compares the models.

3.3.1 Oracle.

Oracle presents a MDM Maturity Model. Oracle has a lot of expertise in the field of data logic and structuring. Their model is centered on five focus areas: profiling data sources, definition of a data strategy, definition of a data consolidation plan, maintaining of data, and utilization of data. Profiling the data sources contains inventorying all sources of data in the IT landscape. From there, the decision can be made whether to include the data in the MDM scope or not. Defining the data strategy means

understanding the use of the data. The strategy consists of controlling the data, security rules, data structure and usage policies. Defining the consolidation strategy contains the definition of the operational data model. It needs to be assured how cross referencing is conducted and how synchronization will happen. Data maintenance entails the desired degree of standardization to be defined, cleansing rules and the quality monitoring requirements. Utilizing data addresses the issues of what particular data gets published, who consumes the data, how the right data will be put to the right place in the right format, and validating the data and insuring the adherence to security rules. (Butler, 2011)

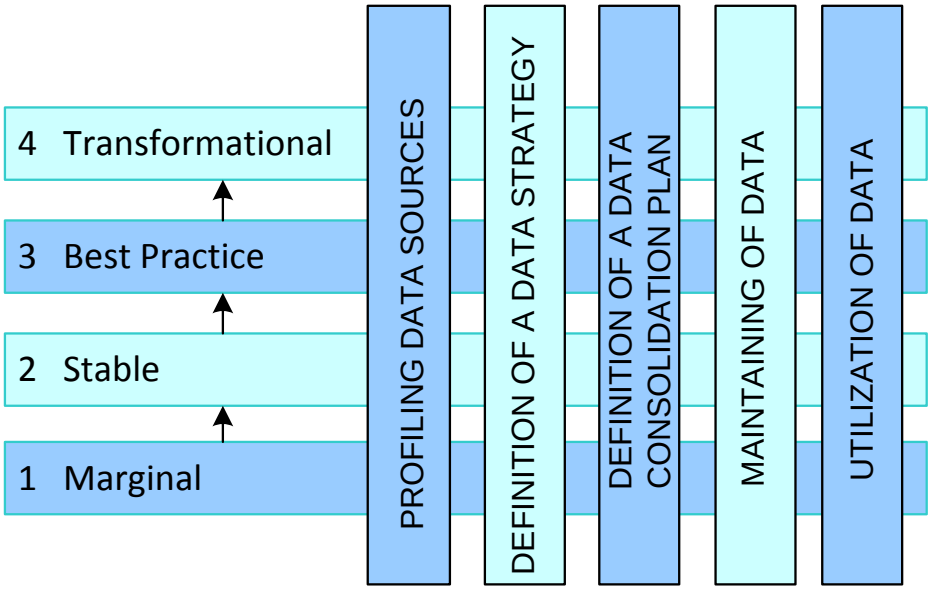
Oracle mentions four maturity levels. The initial one is called “marginal”. It means that the employees are maintaining their trusted sources individually. There are no consistent structures and limited integration and no significant automation.

The next level is named “stable”. Limited tactical implementations have been conducted which are targeted at special divisions. It also contains some limited data stewardship.

The third one is a “best practice” containing process automation improvements. It has an enterprise wide scope and provides a single version of the truth. It is organized from an enterprise architecture team with representatives from business and IT.

The final maturity level is called “transformational” and on this level, the MDM is managed quantitatively. It is integrated in the business’ BI, SOA and BPM processes. (Butler, 2011)

Figure 8 *Oracle MDM Model*

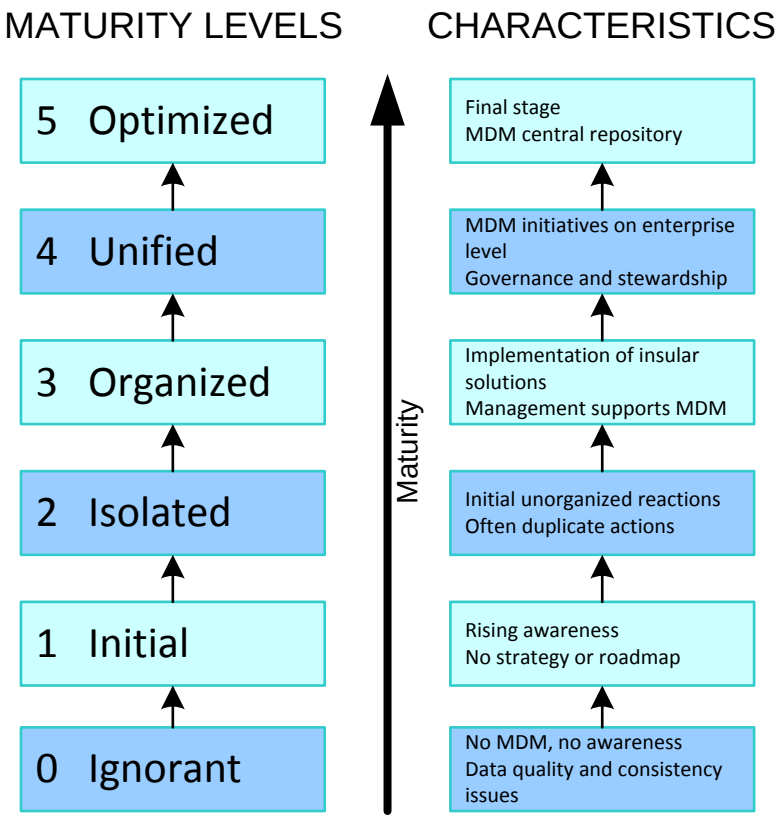


The Oracle model has been chosen because Oracle is a company with a lot of experience and knowledge in the field of data management in enterprise environments. The model addresses the major focus areas of MDM. However; the structure does not perfectly fit the maturity model. It provides only broad areas of interest. The target of this research is to go more in-depth. Nevertheless, all topics from the Oracle model will be part of the MDM maturity model presented in this research. The maturity levels are thought through but they do not cover the total ignorance of an issue. Still, the model gives a good overview about the usual progress from inconsistent isolated ideas to department wide solutions to the final global management which is integrated with other IT solutions.

3.3.2 Information management newsletter.

The newsletter on Information Management published an MDM Maturity Model. They propose that enterprises must follow a sequential approach to improve maturity.

Figure 9 *MDM Maturity Model (Information Management Newsletter)*



Six levels of maturity are distinguished.

The initial level 0: Ignorant indicates that there is no MDM in place and also no awareness of the issue. On this level, different divisions suffer differently from data quality and consistency issues. A serious amount of processes runs without precise and complete data. Analytical processes rely on outdated data. Data integration processes are conducted manually. The paper proposes to initiate an MDM awareness program. The needs for and benefits of MDM, data quality and data governance should be communicated. (Kumar, 2010)

Level 1 is the initial stage with a rising awareness for MDM. The organization recognizes that the existing master data problems affect the business. So, the need is

recognized, yet there is no strategy or roadmap developed. There are no data ownership concepts or governance. If there are initiatives for data quality, they are conducted in an isolated manner. There is no centralized repository for data and processes. According to Kumar, the organizations should continue their initiatives and try to make it popular in the company. MDM should get closely related to the business.

The following level called isolated consists of the step when the organization realizes that problems derive from redundant and duplicate master data. Resulting from this, the enterprise starts reacting. Unorganized groups within the company start taking isolated actions. Since no coordination taking place, duplication of activities is likely. Often the IT is taking a lead in this project but they are struggling to get the business side on board. It is advised to create business cases that can be solved by MDM and to connect the different isolated groups.

Level 3 is described as organized; there are “planned siloed MDM initiatives”. A serious approach has been taken for special master data domains or situations. Actions are taken to implement insular solutions for a broader audience. The upper management is now supporting the MDM because they can see the potential to increase effectiveness and accuracy of processes. A business driven MDM strategy is in place. It is recommended to finish ‘quick win’ projects to have some achievements to present.

The next level is named unified, meaning that there are MDM initiatives on an enterprise level. The strategy is organization wide and on a priority agenda. MDM has reached the level of being recognized as the key enabler for information management. There are policies developed for the integration, exchange, synchronization, and quality management of data. The technology is shared throughout the company and company-

wide governance and stewardship for MDM is installed. For improvement, the author recommends to operationalize the strategy of MDM. Major MD areas should be identified and implemented on an organization-wide level. Any siloed solutions should be integrated into and aligned to the strategy.

On the optimized level 5, the final step is reached. Here, the MDM is the central repository from which data is shared with other applications correctly, consistently and completely. Managing master data is a core activity of the organization. Now it is important to maintain the maturity on the level and even look for further improvement. (Kumar, 2010)

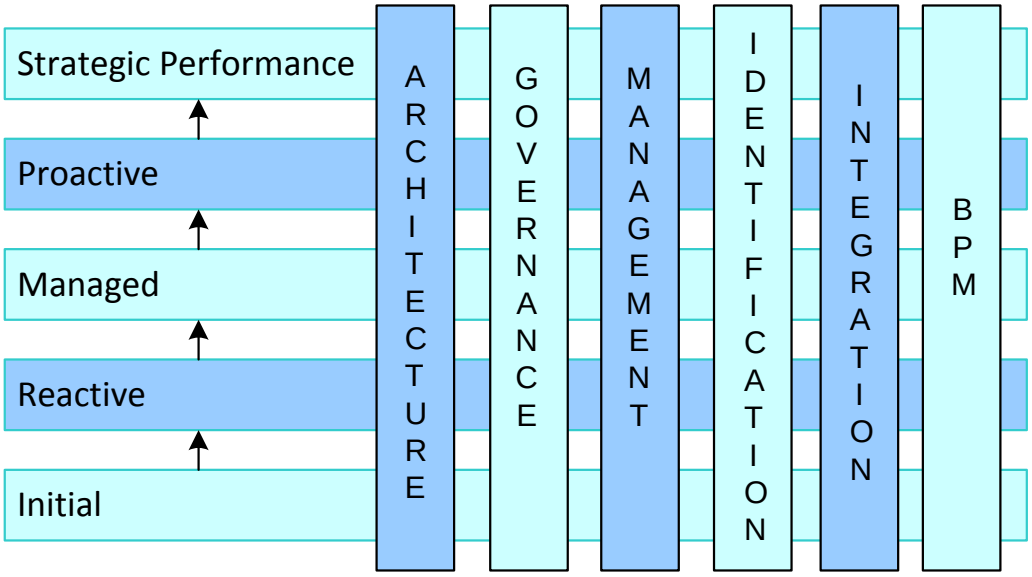
The maturity model of Kumar (2010) provides a good broad overview. It shows maturity levels that are well developed. Nevertheless, the topic MDM maturity is only looked at from a very high level perspective. This is too abstract for this approach. Still, it gives a good insight and proves of some experience in the topic. It can serve as a basis, especially regarding the maturity levels. The lack of focus areas has to be filled from different sources.

3.3.3 DataFlux

The DataFlux whitepaper by Loshin (n.d.) describes MDM as a topic that comes along with technical, operational, and management challenges. At the start of the MDM maturity process, he regards data object consolidation, followed by increasing levels of integration, service and synchronization and in the final state, the MDM environment as an enterprise resource that is integrated within the application architecture with provided services. The purpose of this maturity model is to describe how already existing services

and components can be used for the sake of a master data repository. Additionally, it describes additive capabilities to improve maturity. (Loshin, 2008)

Figure 10 *Data Flux MDM Model*



Loshin distinguishes six conceptual areas in which maturity can be achieved; architecture, governance, management, identification, integration, and business process management (BPM). This is displayed in the picture above. The architecture level is seen as the foundation of MDM. It consists of three pillars. Firstly, the master data model is a consolidated representative master data model which is the core source for all representations. Secondly, it must support all possible formats and structures which could be used. Finally, the MDM systems architecture relies on a framework that takes into account the data life cycle of the master data objects. For master data objects, this means creating, accessing, updating and retiring. The MDM service layer architecture includes the mandatory technical components for covering the data life cycle. (Loshin, 2008)

On the governance layer, one can find the policies of MDM. They consist of data standards, metadata management, data quality, and data stewardship. Artificial systems lack the ability to resolve ambiguities. This is the origin of problems when integrating data sources if data elements do not have the same meaning but the same name or vice versa. Here, a process must be implemented to find global definitions for the same contents. Metadata management derives from the before mentioned step. Through the investigation of data elements, meta data information gets discovered. This meta data should contain the business use of the data elements, these concepts are referred to by the same element definitions, applications that refer to the concept, how the different data elements are handled over the life cycle in the different applications, a determination of data quality properties and monitoring locations within the business process, and how the uses are interconnected. Data quality is also an important part of governance. Since the repository is the single source of truth, a high trust in it is necessary. An understanding must be developed that high quality data impacts business performance. On this layer, there is also the implementation of a data governance and stewardship program. Ownership concepts and oversight mechanisms ensure the adherence to the rules and the active information quality management. (Loshin, 2008)

The management area covers the logic behind the whole MDM. It contains identity management, which is involved with ensuring that for an individual entity not more than one record exists or one can be created which can uniquely be distinguished from all others. It enables and manages the determining of all required attributes for unique identification of attributes. Hierarchy management deals with two aspects. The first is defining a primary source of data if there are multiple ones during the

consolidation process. The second one is about the connections between master data objects, which must be addressed in linkage hierarchies. Concerning migration management, the objective is to substitute existing data interactions, so that the new ones facilitate different versions operating at the same time. The ability to cope with application migration using the same data repository is required. The administration part is concerned with the need to configure and administer applications because several applications use the master repository. (Loshin, 2008)

The identification and consolidation part contains three capabilities. Identity search and resolution deals with determining that more than one representation can be resolved by representation in a unique object. This step includes discovery – the profiling of data and manual review to develop probabilistic models – and then scoring and matching for linking of records. This step feeds into the record linkage. The developed models can be used for a big amount of records to identify records that lead to the same entity. Then, the data entities will be merged and consolidated according to algorithms qualifying the values. (Loshin, 2008)

The following part is about integration. On this layer, the applications get integrated with master data components. Application integration means that the data must easily be entered into the repository but also be easy to access for the applications. For this purpose the infrastructure must provide a standardized path. The MDM component service layer is composed of the ability to consolidate application functionality by the use of service layers. (Loshin, 2008)

Business process management is the part that empowers the MDM initiatives to improve daily business. It is the highest level of abstraction in the model. It must be

considered when designing the technical side. All business process requirements should be reflected by a business application. This includes linking individual activities together. A revision of the business process model empowers to figure key triggers for business rules. This process is also concerned with displaying required conditions that need to be fulfilled. The MDM Business Component Layer contains underlying requirements and definitions with a rules-based system. (Loshin, 2008)

The model distinguishes five maturity levels. The initial level indicates a lack of capabilities; the possibilities for master data exploitation are limited. However, there is already consciousness about the fact that some replicated data sets are relevant to several applications. On this level, the preparation for exploring possibilities to consolidate data sets gets initiated. On the different layers, this level has the following meaning.

Architecture: For every business application, architectures are designed. There is no representative model consolidation. There are no master data models and data overviews are stored in various forms and places.

Governance: There is only restricted cleansing of data by either business units or applications. There are no data stewardships or ownerships defined. There is an awareness of the necessity to achieve overview.

Management: If it is necessary, there is identity management in place. Some application configuration is established, but it is not steered through a management function.

Identification: Identity management is used in a limited volume. Certain units manually take care of data consolidation if it gets required by applications or changing legal regulations.

Integration: There are replicated copies of reference data. The reuse of data is rather limited and there is no reuse of applications.

Business process management: There is very little or no involvement of the business for requirements definitions except for high priority ones.

The second level is called reactive. On this level, there is awareness of the fact that duplicate data can cause problems and also first attempts to solve those problems. New tools are starting to be used but learned lessons and added value is not shared throughout the enterprise. Some primary attempts of consolidating master data are made.

Architecture: There are already attempts to gather data dictionaries in one repository. There are first efforts to investigate unsophisticated application services. Different possibilities for information sharing are being evaluated.

Governance: External applications are utilized for management of metadata. There is also a start in data quality management concerning parsing, standardizing and consolidation.

Management: Personal resources are explicitly allocated to managing the use of applications. There are trainings for the staff about rolling out new tools and technology. The administration of metadata and master indexes gets centralized.

Identification: Identity search and match is used to find duplicates. A basic solution is used for record linkage.

Integration: Initial investigations of data consolidation for new analytical tools are conducted. A data warehouse is utilized as the main source for master data. Still, there is no integration into the connected applications.

Business Process Management: There are already descriptions of conceptual business process models. There is an initial effort in using business rules within applications.

The third level is managed. On this level, the organization gets able to integrate new applications and existing ones. The focus is on moving from single data areas to a repeatable process.

Architecture: There is a clearly defined core data model. There is already a basic architecture for shared master repositories. Furthermore, there is a framework in place for basic master data lifecycle activities.

Governance: Tools for data quality, and policies and procedures for data quality management have been established. Data quality issues are tracked and processes for data standards are in place.

Management: Identity management is centralized and utilized across different applications. Discovered relationships and hierarchies between data objects are used by analysis tools. A migration concept is available for some applications.

Identification: Identity search and match applications are available for the applications. Record linkage is connected to the MDM service layer. There are rules for merging and consolidation and processes are established.

Integration: For the application integration, there are component services available and services synchronize applications with the repository.

Business Process Management: Business rules are integrated with master data operations. There is a strong connection between business applications and data objects. Business process analysis is included in master data engineering.

The following level is named proactive. At this level, the organization succeeded in integrating applications with the master repository via the service layer.

Architecture: Master models have been established. The organization obtained the competences to move from an index based MDM framework to a transaction based one. For the application architecture, SOA is established and for the business meta data, there is a centralized management.

Governance: There is an enterprise government program and enterprise data standards and metadata management implemented. The proactive monitoring of data quality control feeds into governance programs.

Management: The identity management is integrated in the enterprise processes and all master object instances can uniquely be identified. Analytic and operational activities are supported by a hierarchy management. This hierarchy management makes the correction of false positive consolidation errors possible.

Identification: Identity management includes identity search, match and resolution. All operations about the data life cycle are organized around merging and consolidation services. The consolidation takes place in the background.

Integration: The data synchronization is fully integrated in the life cycle services. On a master object level, application integration is supported by a component layer. Business application integration is driven by SOA.

Business Process Management: There is business logic in place which is being reused. The rules are integrated in a rules engine and available on a business process level. Business analysis is an important part of application development. The relationships with customers are personalized and business processes are automated.

The last level for this model is described as strategic performance. On this level, operational and analytical business requirements are supported by an MDM combined with SOA.

Architecture: For internal applications, the complete transaction integration is available. There are public interfaces enabling a quick processing of master data.

Governance: Information sharing on a high quality level is guaranteed by cross organizational data governance.

Management: In internal and external representation, the identity management of all data objects is synchronized and the migration of legacy application has been completed.

Identification: Identity resolution services are exposed externally to the organization. Business performance directly is tied to master dimensions.

Integration: Business process models and master object models drive the application development.

Business Process Management: Application design and development is fully driven by the business. The applications integrate business rule engines. Data instances are managed within the master repository. Through MDM, embedded predictive analytics are made possible.

This model provides a very detailed description of maturity levels and focus areas. However, there is a layer missing in which there could have been a more granulized distinction within the focus areas.

3.3.4 Comparison of the maturity models.

The presented models have maturity levels ranging from four to six either in- or excluding a maturity level of complete ignorance of the existing issues. This range seems to be good because it is precise enough to enable a distinction between first attempts in MDM from more elaborate solutions, but not too detailed to get lost in fine differences between maturity levels that probably do not fall into account. All the models share the progress view from complete ignorance of the importance of MDM leading towards awareness and initial steps followed by first decent attempts to a fully mature MDM implementation. For all the models, it can be seen that a shift takes place from an operational to a strategic approach towards MDM.

The Oracle model has five focus areas. They follow a possible timely structure of a project. Apart from that, they are not very detailed. This model is fine to assess a broad overview about the state of maturity. The model by Kumar does not provide focus areas at all (2010). This is also only suitable for a first insight into the organization's MDM maturity. But for the assessment with a focus area maturity model, this is not suitable. However, they indirectly describe focus areas, which should be considered from their point of view. The last model is the most elaborate one. It distinguishes 6 focus areas and gives examples for the different maturity levels. Here, the focus areas are grouped after broader business tasks than after processes. This gives good insight into the maturity already, but unfortunately there is no clear structuring within the areas.

The purpose of this comparison is not to choose the best model to use it, but to evaluate the different approaches and take this as input for the development of MDM Maturity Model. All the models have in common that they are not detailed enough for the

direct assessment of an organization's maturity. The models in total give a good overview of the areas that should not be forgotten. Therefore, all the mentioned focus areas will serve as input for the MDM Maturity Model, even though the granularity will be finer.

The following table summarizes the models and which topics they comprise. There are topics identified that are in total covered by at least one model.

Table 8 *Comparison Maturity Models*

	Oracle	IMN	DataFlux
Definition of Master Data	√	√	√
Master Data Model	√	√	√
Data Landscape	√		√
Assessment of Data Quality	√	√	√
Impact on Business		√	√
Awareness of Quality Gaps	√	√	√
Improvement	√	√	√
Data Usage	√		√
Data Ownership	√		√
Data Access	√		√
Data Protection			√
Storage	√	√	√
Data Lifecycle	√	√	√

4 Development of the MD3M

In this chapter, the actual practical development of the MD3M (**M**aster **D**ata **M**anagement **M**aturity **M**odel) will be presented. To assess the maturity of the master data management of an enterprise, the MDM maturity model is suggested. The MDM maturity model is a means of assessing the whole process of master data management including the data point of view and also focusing on the whole operational process. This chapter will describe that the assessment bases on the focus area maturity model.

In order to keep this research consistent according to academic requirements, the development will be based on guidelines and frameworks from academia. As already mentioned in earlier sections, this research is based on the design science approach described by Hevner el al. (2004). It is also in accordance with the research of Becker et al. (2009) who published a paper on the development of maturity models for IT management and proposed a procedure model for the development. Also they base their research on the design science approach to achieve a reasonable catalogue of requirements for the design of maturity models.

According to Becker et al. (2009), a maturity model is an artifact that aims at be solving the problem of defining an organization's current status regarding their capabilities and deriving means for improvements. They developed a procedure consisting of eight steps for developing scientifically valid maturity models in accordance with Hevner et al. (2004)

The first step is the comparison with existing maturity models. Researchers should have an overview about what already exists, so that their development can be based on already existing models or improve an already existing one. This instruction

takes into account Hevner's 'Guideline 6: Design as an artifact' and 'Guideline 4: Research Contributions'. Hevner's 6th guideline: 'Design as a Search Process', which implies the need of an iterative solution development, refining, evaluating and possible enhancement, is considered in the second and third step. The second advises an iterative procedure saying that models must be developed step by step and the third gives advice about the evaluation. All principles and preconditions for developing a maturity model as well as usefulness, quality and effectiveness must be evaluated iteratively. This one is also in accordance with Hevner's third guideline about evaluation, stating that results must be evaluated with appropriate scientific grounding. The next instruction deals with the topic of multi-methodological procedure. The development of maturity models employs a variety of research methods, where the application needs to be well-founded and finely tuned. Here, the guideline of 'Research Rigor' stating that selected methods have to be rigorously attuned is applied. Hevner's second guideline about 'Problem relevance' recommends that the problem solving artifacts should be innovative but also relevant to researchers and/or practitioners. This requires thus a precise definition of the problem and therefore is in line with the fifth and sixth procedural step; identification of the problem relevance and definition of the problem. These imply that the solution's relevance must be demonstrated and that the future application domain, the conditions and the benefits must be defined before designing the model. Hevner's seventh guideline (Communication of Research) finds place in the last two principles. The results should be targeted at specific user groups. The rule 'Targeted Presentation of Results' indicates that the presentation of the model must be targeted at an audience regarding the conditions of the applications. The last one (Scientific Documentation) recommends the detailed

documentation of the design process, taking into consideration every relevant step in the design process. (Becker et al., 2009; Hevner et al., 2004)

The development of the Master Data Management Maturity Model will follow the guidelines presented above to ensure the validity for an academic design science research.

4.1 The MD3M Development Process

The MDM Maturity Model aims at analyzing the maturity of a company's master data management maturity throughout the whole process. It does not only consider the data perspective itself, but the whole cycle from analyzing the data to the implementation. The MDM maturity model is an instance of the focus area maturity model. The different steps from Figure 7 will be explained in the following.

4.1.1 Scoping.

This block is all about defining the scope of the project and having documents that define it. Since this type of projects normally originates from the perception that more insight into the domain is needed, this block is a rather short one because it benefits from preliminary knowledge.

In this field, the main step is to *identify and scope the functional domain*. Here, it should be decided on what and what not to include. The deliverable of this step is the domain in which the maturity assessment will be conducted. In this research, the scope comprises master data management. The research comprises the whole implementation path of master data management from the initial data analysis to the implementation path. Within this scope, the aim is to figure out a company's maturity. With a valid assessment

in this area, companies can use the results as benchmarks and compare themselves with others to position their performance and identify potential for improvement.

4.1.2 Design model.

This block is the biggest part of the whole process. It is all about developing the matrix for assessing the current state of maturity in a theoretical manner.

The first step in this area is the determination of focus areas. These will be grouped into functional domains. The focus areas will be mutually disjoined, but will all represent a part of the whole assessment topic. The focus areas are a result of the comparison of the presented maturity models. Grouping the focus areas will help making the model accessible and activities that belong together can be grouped together. (Steenbergen et al., n.d.) The deliverable of this process step it thus the list with focus areas grouped into key topics. This can happen with a bottom-up or a top-down approach. Either the functions can be defined and then the focus areas are gathered and then grouped into the functional domains or the focus areas are defined first and then functional domains are found to integrate the focus areas in.

The following step is the determination of capabilities. Each focus area consists of a number of different capabilities which represent a progress in the level of maturation. The definition of the capabilities relies on the rational for improving focus areas steps by step. The capabilities and their order are defined separately for each focus area. The capabilities are developed and improved based on thorough literature review and expert opinions. (Steenbergen et al., n.d.)

In this case, the process is the definition of the capabilities by investigating the processes. The deliverables are therefore the capabilities documented thoroughly.

The third move is to determine the dependencies between the capabilities. Now, it has to be determined whether the capabilities are interrelated. Since the capabilities represent a progress, it is generally advisable to adhere to the given order. Since this is only advised, there can be a valid reason not to follow that order, but in some cases, the order is inevitable and must be followed because some capabilities precede other ones. (Steenbergen et al., n.d.)

Hence, the actual process here to identify the dependencies by investigating the capabilities and the processes. As a result, the documentation from the step before will be extended by the dependencies. Additionally, there will be influential factors identified. In order to keep the model applicable for all kinds of organizations, some capabilities might not apply for all. This will be covered by the influential factors. If some conditions apply, certain capabilities will either be enabled or disabled.

Finally for this area, one has to position the capabilities in the matrix. It is now possible to arrange the capabilities in the matrix. This arrangement is partly based on the dependencies that were defined in the preceding step but also bases on practical reasons. (Steenbergen et al., n.d.)

In this step, the capabilities and the dependencies that were defined in the steps before will be arranged in the matrix in a logical reasonable order. With finishing this, the MDM maturity matrix is completed.

4.1.3 Develop instrument.

After the theoretical and generic development of the matrix, measures and actions have to be defined to actually assess the state of maturity. Within this functional domain, there are the following two measures to be implemented.

The assessment instrument needs to be developed. Measures for each capability have to be defined. This is likely to be done by formulating control questions to be used in an interview or a questionnaire. The formulation of these is derived from the description of the capabilities and also on experience and practice. (Steenbergen et al., n.d.)

The deliverable of this rather big process is therefore the questionnaire that is to be filled in by responsible experts in the company to assess the level of maturity.

Afterwards, improvement actions will be defined. Originating in the capabilities and the incremental path, actions can be determined that will improve the maturity in a particular area. This can help the practitioners scoring higher in the focus area. (Steenbergen et al., n.d.) Of course, from a generic view, this can only result in some general advice, but for the actual assessment at the case company, this will lead to actual feasible suggestions.

4.1.4 Implement and exploit.

This phase consists of the practical realization. The model will be implemented at the company so the maturity of the master data management can be assessed. Following, the matrix can be improved iteratively to achieve a higher level of maturity.

The maturity model must be implemented. This can be conducted via filling in the before mentioned questionnaire. Then the results can be evaluated and the maturity for the different focus areas is visible. (Steenbergen et al., n.d.) For this purpose, a questionnaire was developed. The questionnaire is documented in the Appendix F. It is half-structured and uses non-suggestive questions in order to avoid manipulation of the

interviewees. Furthermore, it contains open questions to encourage the respondents to provide individual input.

After the preceding step, the matrix can be improved iteratively. With the results from the evaluation, the matrix can be improved. Now it is assessed which areas are of main concern and are least mature. The latter chapters will in detail describe the actions that are suggested to be taken as a consequence of the analysis of the matrix. The deliverables of the improvement process highly depend on the findings and the consequences that are drawn from those.

Finally, the results should be communicated to the scientific community. The results are of interest for practitioners and the scientific community. First of all, the results are interesting for the company the research is conducted at. Secondly, it is interesting for all companies dealing with master data management and which are willing to improve their processes and wanting to have some benchmarks. Nonetheless, the topic is of relevance for science. It shows improvement potential for master data management and provides possibilities to academically validly compare the efficiency of the organizations.

4.2 The MDM Maturity Levels

The presented maturity models are not far away from this research's approach. The Oracle model in chapter 3.3.1 presents four maturity levels with the first one indicating initial awareness. The second and third model in chapter 3.3.2 and 3.3.3 define 5 levels, also with initial awareness on the first level.

The following table gives an overview of the amount of maturity levels and their meaning.

Table 9 *Comparison of Maturity Levels*

Level	COBIT	Oracle	IMN	DataFlux
0	non-existent	-	-	-
1	initial	Marginal	Initial	Initial
2	repeatable		Isolated	reactive
3	defined process	Stable	Organized	managed
4	managed and measurable	Best practice	Unified	Practice
5	optimized	transformational	optimized	Strategic performance

From this information, the decision was taken to exclude a level with a non-existing maturity. The ignorance of existing issues within the organization concerning master data management will be considered as no maturity at all and therefore not be considered in the matrix as a unique level. Hence, the first level will be – like it is consensus among the different models – an initial one. On this level, the attention has just been raised for the topic and initial plans are developed to investigate and tackle the problem. Consequentially, there are isolated measures initiated by single units or persons in the organization without relation to others. They are meant to solve internal problems of the particular unit that are due to the insufficient MDM in place. On the third level, the

initiatives start to be aligned among each other and awareness is in place for the initiatives and problems of other units. This level is called organized. Different units start to collaborate for certain projects. On the following level, the initiatives tend to be best practices. The organization adopted common frameworks and implemented them. Processes have been defined and are adhered to. The fifth level is called optimized. On this level, all processes concerning MDM are optimized for the purposes of the organization beyond the best practices. The maximum of benefits can be drawn from the MDM initiatives and it is constantly reviewed whether the circumstances change.

The following table depicts the chosen maturity levels for the MDM maturity model and a short description. They are mainly based on the COBIT model earlier described, except that here the level zero is left out as an explicit maturity level. This framework is based on the same perspective like this research. It is not too detailed for this top-down view and since this is an initial prototype, the overall view seems to be the best approach. Achieving a maturity level means all capabilities of the stage are fulfilled.

Table 10 *Description of the MDM Maturity Levels*

Level	Description
1: initial	A first awareness for issues regarding the topic of MDM has been raised on an operational level. Initial steps are initialized.
2: repeatable	Measures from individuals are conducted to solve individual problems. No connection to other units or projects. Still operational
3:defined process	First collaborations take place on a tactical level. Awareness was created for the existence of other initiatives.
4: managed and measurable	Best practices are in place for handling of MDM. There are defined processes on a tactical level.
5: optimized	Optimized handling of MDM. The organization's efficiency has been improved. Tactical approach on the topic.

This decision implies that this maturity model can be classified as a staged model.

4.3 The MDM Maturity Attributes

In this chapter, the structure of the maturity matrix will be developed. The construction cycle is based on Steenbergen et al. (n.d.) which was described before. The development of the maturity matrix is the following. It is structured in a loop. First, basing on the literature review, a maturity model is developed. Then, two expert interviews are conducted. Two interviews were conducted in order to improve the matrix. The interview setup can be found in the Appendix C. It was a semi-structured approach which started with an introduction of the concept of maturity models with all relevant

terms and afterwards a presentation of the MD3M. The interviewees were two experts on the topic; each with a different focus on MDM. One has a more technical background and view on it, while the other one is more focused on the business perspective. This is an adequate way to cover both sides of the coin, since the topic has as well an IT and business focus. The main results of the improvement loop were renaming some capabilities to make the meaning clearer and to incorporate two key topics into one because they were overlapping. The influential factors were developed in cooperation with the experts.

The following table summarizes the experts' knowledge on the topic and justifies their status as experts.

Table 11 Interviewees' expertise overview

	Expert A	Expert B
Position	Manager	Manager
Academic Background	Engineering	Business Informatics
Working Experience	>15	>10
Focus areas in work	Data management, data strategy	Data management, implementation strategy

During the deployment of the matrix, several interviews are conducted at EET. There is a strict disjunction between the generic matrix setup and the specific situation at the case company. In this step, the focus is on the development of the matrix for a general use. As interviewees, several leading persons in the company were chosen who are experienced in this topic.

The interview follows a semi-structured interview. First, the topic and the general idea of the matrix are presented to the interviewee. The interviewee is being informed about the goal of the interview in advance before the meeting, so he has the possibility to think about it in advance.

In the second part, the interviewee is asked to present relevant parts of master data management from his point of view. He is asked what factors have to be measured and what is important in general. If the interviewee is lost in the situation, the interviewer will give some hints regarding the subject or the scope.

The third part is the one, in which the developed model is being presented. The setup of the matrix will be shown to the interviewee as well as the logic behind it. It will be explained why the matrix is built up in that way.

Based on that, there will be a discussion about both approaches in the last part of the interview. The interviewee will be asked some predefined questions about the presented matrix. Then, an open discussion will be conducted about the differences of both models and how the matrix can be improved.

Based on the interviews, the matrix will be successively improved until consensus is reached with the interviewees. The setup of the interview can be found in the Appendix C of this thesis.

The capabilities in the matrix are described in a document. The structure of this document is based on Bekkers & Van De Weerd (2010). In this document, each capability in the matrix is described in terms of the factors goals, actions, prerequisites, and notes. Goals are what must be achieved to get to the according level of maturity. Actions are what need to be done in order to reach the goal. Prerequisites contain

interdependencies that have to be considered, like other capabilities which need to be achieved before. The whole document can be found in the Appendix D.

The next paragraph shows the general structure.

KEY TOPICS

Description of the key topics

Focus Areas

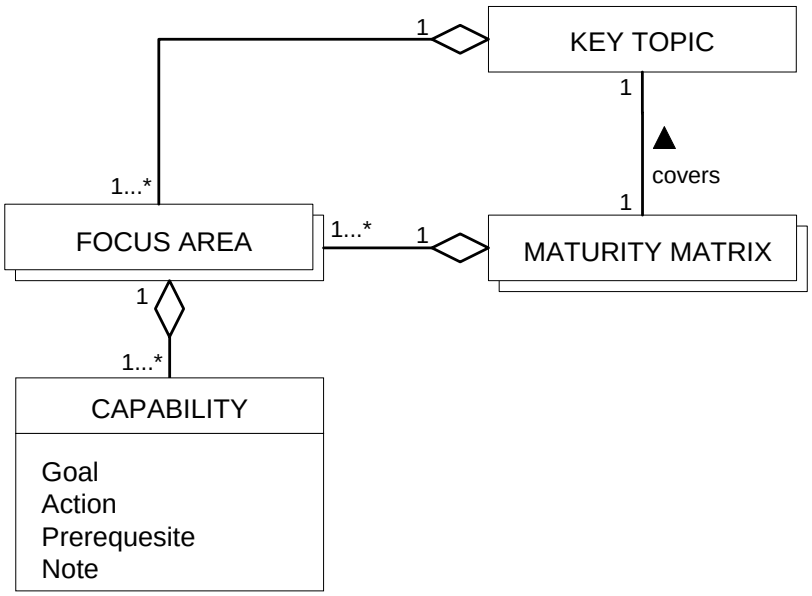
Description

A ***Capabilities***

Goal:	Description of goal
Action:	Description of action
Prerequisites:	Optional
Note(s):	Optional

The following figure graphically displays the structure of the attributes.

Figure 11 *Structure of the maturity matrix*



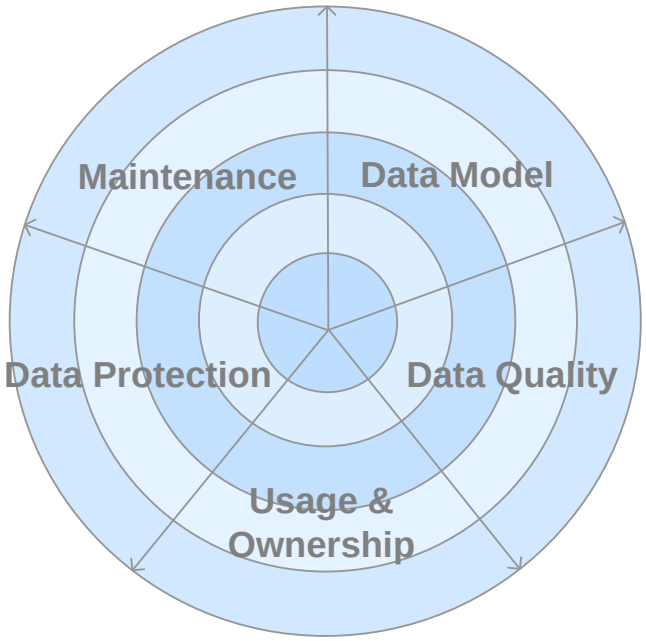
The development of key topics, focus areas and capabilities will be described in the following.

4.3.1 The key topics.

The key topics and the focus areas were developed with a bottom up approach. They were deployed after a thorough analysis of the literature study of existing MDM models and other literature and studies on the topic. They were chosen to cover all aspects of Master Data Management that are relevant for an organization. An appropriate granularity was developed. All factors from the presented models are somehow covered in the new MDM maturity model, even though the order is different. This is due to the finer granularity and the structuring with the key topics and the focus areas within. Furthermore, there was practical input from the company regarding topic-unrelated practices. So the only input for this model was general practice that a company follows,

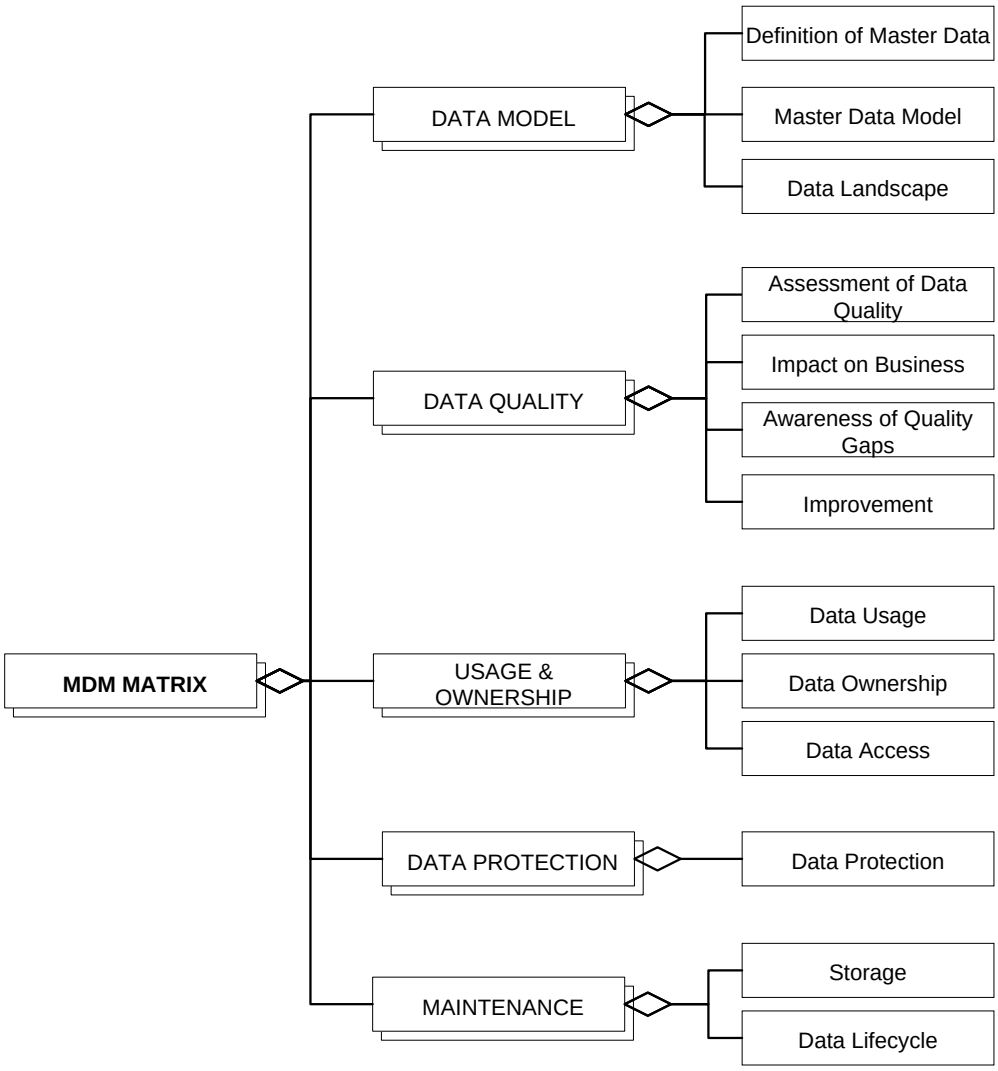
but no company-specific factors. The model is designed to fit in general to all – especially bigger – companies dealing with master data. For too small organization, the effort to install an elaborate MDM would be exaggerated. The model was deployed and validated in a loop approach. The initial draft mainly based on literature was presented to experts within the company and evaluated with them. Iteratively, the final matrix was developed, ensuring that it covers all important aspects of the topic.

Figure 12 *Key Topics of the MD3M*



This figure depicts the key topics with the according focus areas. They are further elaborated on in the following paragraphs.

Figure 13 *Key Topics and Focus Areas*



Concerning the key topics, the goal is to find mutually exclusive and logical chunks in which the focus area can be grouped logically. Resulting from the previous research and the focus on MDM, the choice is to have a topic-oriented approach. This approach is chosen because processes might look different across different companies. If the MD3M was too much focused on processes, it would not be generic anymore and could not be applied by all companies. The other option would be to keep the process descriptions very general that it would still work for every company. Therefore, the

model is organized around topics that can be included in any process depending on the organizational setup.

There are five key topics identified, with 13 capabilities in total:

DATA MODEL. This key topic deals with the data and the infrastructural and organizational view on it. It contains topics like what data is considered as master data, how the data is structured, which systems use what data, and where the data is stored.

Davenport and Prusak mention the aspects of 'Contextualization', meaning that it must be clear for which purpose data is collected, 'Categorization' meaning that key components are known and 'Condensation', meaning a summary of the data is essential for overview purposes (2000). It is crucial to have a shared understanding of the data in the organization. Shared definitions help align the understanding of different stakeholders with e.g. technical and non-technical perspectives. (IBM, 2007; EDUCAUSE…, 2009) It is important to have one single view on the data and on the systems that use the data in order to consolidate different sources into one for the whole enterprise. (IBM, 2007; Loshin, n.d) Experience has proven that the understanding of elements and terms differ throughout the business units if they are not explicitly agreed upon and communicated.

The following are the three focus areas belonging under this category:

- Definition of Master Data
- Master Data Model
- Data Landscape

DATA QUALITY. The key topic **DATA QUALITY** is dedicated at data quality in all regards. It includes finding ways to assess the data quality and assess it, finding ways to improve data quality and investigating the reasons for and impact of quality issues.

Furthermore, the organization can assess what the most frequent and critical sources for problems are. Data quality is very important for an organization. Therefore, the organization must find out which quality the data has, how this can impact the business, where the sources for poor quality lie and how to improve the data. Data is an organizational asset, so the quality must be good in order to be a competitive advantage and in order not to cause problems. Therefore, data quality must be known and improved. (Redman, 1995; Pipino et al., 2002) In order to do this, objective criteria must be defined that match the needs of the organization (Kahn et al. 2002). Awareness must be raised that poor data quality impacts the business and therefore necessary measures can be taken in an appropriate seriousness (Van der Linden, 2009).

The focus areas are the following:

- Assessment of Data Quality
- Impact on Business
- Awareness of Quality Gaps
- Improvement.

USAGE & OWNERSHIP. This key topic is dedicated at defining who uses the data in which systems. Which employee has read/write access and is it clear why people are granted or denied access to certain data? The organization can find out if there are data ownership concepts implemented and see whether the historical grown way still displays the needs. It is generally consensus throughout academia and practitioners that divided responsibilities shared with different people are not effective. Therefore, ownership concepts have proven to be an adequate concept. (EDUCAUSE Center For Applied Research, 2009; IBM, 2007; Loshin, 2008) Furthermore, due to data privacy and data

protection reasons, data has to be distributed to appropriate users and not be made available for users without access rights. Of course, data availability must be ensured at all time. (Anderson & Moore, 2007; Casassa Mont & Beato, 2007; EDUCAUSE Center For Applied Research, 2009; IBM, 2007; Loshin, 2008)

Under the umbrella of usage and ownership fall the following focus areas:

- Data Usage
- Data Ownership
- Data Access

DATA PROTECTION. This section is about the technical security of data; whether and how it is secured against possible incidents. . Incidents can be of different kinds; either failure of components, software bugs or steered by people on purpose, like sabotage, hacking, fraud or theft. To ensure confidentiality and the running business, data must be protected. (IBM, 2007; Loshin, 2008; Shaw, Chen, Harris, & Huang, 2009) This protection must be conducted via physical measures and software precautions (Bernard, 2007; Borgman, 2000).

This key topic contains one focus area:

- Data protection

MAINTENANCE. Here, the focus is on physical storage and the data lifecycle. The main points to investigate are how the data is stored and how is the data treated during the lifecycle. Organizations rely on software solutions. These solutions regularly become faster and more performant quickly. In order to make use of the technical innovations, systems should be up to date and be maintained properly. Furthermore, the inserted data

should be kept clean as well, so outdated data is to be removed according to the data lifecycle. (Bowker et al., 2010; Youssef, Butrico, & Da Silva, 2008)

This area contains the following sub-topics:

- Storage
- Data Lifecycle

4.3.2 The focus areas and capabilities.

This chapter will elaborate on the focus areas and capabilities. These were developed resulting from literature research and comparison of best practices. The before mentioned key topics result from grouping the focus areas into mutually exclusive areas. For each of the thirteen focus area, there are 5 capabilities that need to be fulfilled to reach full maturity. The capabilities and the focus areas will be as a whole shown in the Appendix D. They form a subsequent order of activities to be conducted in order to reach a higher maturity. In the following paragraphs, the focus areas and related capabilities will be described and reasoned upon their selection. To improve readability, abbreviations in brackets will be used to indicate relationships between the arguments and the capabilities. As an example, DMD stands for ‘Definition of Master Data’ and the following letter stands for the maturity level; ergo DMD-A indicates first maturity level in the area ‘Definition of Master Data’. The full list of abbreviations can be found in the appendix G.

4.3.2.1 Definition of Master Data [DMD].

In order to achieve a successful master data management, it is necessary to have one shared understanding of the term. Creating common definitions helps aligning

technical and non-technical understanding [DMD-A] [DMD-B] [DMD-C] [DMD-D] [DMD-E] (IBM, 2007). According to the IBM paper, “organizational structures and awareness” [DMD-A] and “classification and metadata” [DMD-A] [DMD-B] [DMD-C] [DMD-D] [DMD-E] are key elements of effective data governance (2007). This indicates the responsibility of both business and IT. Misunderstandings can be solved by people easily. But in order to resolve possible misunderstandings, communication must take place to uncover these misunderstandings [DMD-A] [DMD-B] [DMD-C] [DMD-D] [DMD-E]. This global understanding is crucial when it comes to designing technical solutions. (Loshin, 2008; Otto & Hüner, 2009) The organization needs a shared terminology, possibly in the form of a glossary to have defined unbiased words to talk about. [DMD-D] [DMD-E] (Otto & Hüner, 2009) Shared definitions of master data are a key enabler for successful data governance. (EDUCAUSE Center For Applied Research, 2009)

4.3.2.2 Master Data Model [MDM].

Loshin calls the master data model a “single view of the key data entities across the organization” (n.d.). He furthermore describes the data model as “complex, but integral” for MDM. The model provides the possibility to enable data availability and sharing with other users. (IBM, 2007; Otto & Hüner, 2009) [MDM-D] [MDM-E] Owning this model enables the organization to create a better understanding between the technical and the human view of data. The graphical representation facilitates human understanding and updating.(Otto & Hüner, 2009) There should be one model which is consolidated across the whole organization which contains an overview of all master data across the different systems [MDM-D] [MDM-E]. The information for this model can be

gathered from the source systems (Loshin, 2008). Inclusion criteria derive from the definition of master data. [MDM-A] [MDM-B] [MDM-C] [MDM-D] [MDM-E]

Also one of the described models defines the primary level as having no master data models in place. Furthermore, there are unstructured collections of data overviews spread across the organizations and there is a limited consolidation of models throughout the organization. [DMD-A] However, the necessity to have an overview is recognized [DMD-A]. (Loshin, 2008) On the following maturity level, attempts are mentioned to collect data from different sources into one single repository [DMD-A] [MDM-C] The second level is characterized by showing some trials to collect all different data overviews into a single one. Possibilities of sharing information are investigated and master data overviews are now initially centrally administered [DMD-B]. (Loshin, 2008)

Basing on the steps in this focus area, a better overview of the data model is achieved and the experiences from this field can be used when defining the master data model [DMD]. (Loshin, 2008) The author's third, i.e. managed level relates to the MD3M's. According to him, core data models should be defined in order to ensure persistence [MDM-C]. A basic architecture for a shared repository should be provided [MDM-D]. On his fourth level, the author suggests having master data models established and the management of master data should be centralized [MDM-D] [MDM-E]. (Loshin, 2008) The organization is advised to gather experienced personnel from different business units to ensure thorough understanding and adequate implementation. [MDM-C] [MDM-D] [MDM-E] (Otto & Hüner, 2009)

4.3.2.3 Data Landscape [DLS] [DLS-A] [DLS-B] [DLS-C] [DLS-D] [DLS-E]

The focus area Data Landscape covers the data and the systems which are using that data. It is about providing the data to users with the appropriate data systems and applications. The need to get an overview results from the fact that data is spread across different sources. These sources are independent from each other and likely to contain redundant information. The goal is to consolidate different instances into a unique one for the whole enterprise. (IBM, 2007; Loshin, 2008)

The data's usefulness for the business needs to be defined within the organization [DLS-A] [DLS-B] [DLS-D] [DLS-E]. Redundancies must be solved into a single representation of the data element [DLS-C] [DLS-D] [DLS-E]. (Loshin, 2008) The structuring of the data must be organized in a sensible way. [DLS-E] (EDUCAUSE Center For Applied Research, 2009)

4.3.2.4 Assessment of Data Quality [ADQ]

It is important to be informed about the actual master data quality since the data serves as a crucial organizational asset and competitive factor. (Redman, 1995) In order to improve data quality in the long run, the data quality needs to be assessed. . [ADQ-A] [ADQ-B] [ADQ-C] [ADQ-D] [ADQ-E] (Pipino et al., 2002) Data quality has not only objective measurable aspects, but also subjective perceptions of users. (Pipino et al., 2002) The users must develop a feeling about data quality and not take the status as given and unchangeable. [ADQ-A] The information about data quality needs to be objective and according to standards even though assessing the quality is a rather imprecise science. (Kahn et al., 2002) Different use cases might require different quality. (Pipino et al., 2002) Sufficient data for one team can still be insufficient for another one

(Redman, 1998; Winter & Meyer, 2001). [ADQ-B] [ADQ-C] Strong and weak aspects of the master data management must be investigated. (IBM, 2007) The quality must be assessed because this has a strong impact on business performance and operational productivity. For this, quality tools need to be used. (Loshin, 2008) Data quality policies need to be in place. Data needs to fulfill a defined purpose, also referred to as "fitness for use". [ADQ-C] (Loshin, 2008; Otto & Hüner, 2009) The quality needs to be defined and assessed objectively. This helps providing a standard for assessing the quality objectively throughout the organization. [ADQ-C] [ADQ-D] (Kahn et al., 2002; Otto & Hüner, 2009) Elaborate quality systems are needed for this purpose. [ADQ-D] [ADQ-E] (Redman, 1995)

4.3.2.5 Impact on Business [IOB]

Having a decent master data management is a prerequisite for a working business. (Otto & Hüner, 2009) Poor data quality results in different impacts for the business. [IOB-A] [IOB-B] [IOB-C] [IOB-D] [IOB-E] There are different kinds of impact that need to be investigated. [IOB-A] [IOB-B] [IOB-C] These impacts result in monetary and reputational effects. [IOB-C] [IOB-D] [IOB-E] (Van der Linden, 2009) Problems with the data due to various reasons result in unsatisfied counterparties, ergo in reputational problems. (Redman, 1998) [IOB-A] The organization must know to which extent these problems impact the business. [IOB-D] Furthermore, the organization suffers from a monetary perspective. [IOB-B] This includes a precise calculation of how the organization gets harmed financially. [IOB-D] (Redman, 1998) Finally, the firm must be able to accumulate those factors to one calculation. [IOB-E]

4.3.2.6 Awareness of Quality Gaps [AQG].

Among other things, data quality management contains the measurement and certification of data quality [AQG-A] [AQG-B] [AQG-C] [AQG-D] [AQG-E]. (IBM, 2007) The company needs to become aware of the fact that business performance and operational productivity are strongly dependent on the quality of data. By assessing the quality, the organization can identify reasons for low quality. [AQG-A] [AQG-B] [AQG-C] [AQG-D] [AQG-E] (Pipino et al., 2002) It is important not to assume a decent quality level that does not reflect reality. [AQG-A] (Redman, 1995) Different external triggers like changes in regulations can make the organization unable to act because the master data is not timely, correct or complete. [AQG-E]. (Loshin, 2008; Otto & Hüner, 2009) The organization must investigate which impacts apply for the company. [AQG-A] [AQG-B] (Van der Linden, 2009) These can be identified with different analyses on the data. [AQG-D] [AQG-E] (Otto & Hüner, 2009) An awareness of the extent of quality gaps must be established. (Redman, 1995) [AQG-D] [AQG-E]

4.3.2.7 Improvement [IMP].

IBM emphasizes the importance of ways “measure, improve and certify the quality and integrity of production, test and archival data” [IMP-A] [IMP-B] [IMP-C] [IMP-D] [IMP-E] (2007). Improving the data quality is relevant because the business highly relies on the master data and requires a good quality to have few errors and the lowest possible manual effort. (Loshin, 2008; Pipino et al., 2002; Redman, 1995; R. Y. Wang, 1998)

For this, the organization needs to assess areas in which the data quality is not satisfying [IMP-A]. Building upon that, awareness should be raised for the relation

between quality of data and efficiency and effectiveness. The value of the data needs to be promoted [IMP-B] (IBM, 2007). Then, the quality needs to be assessed objectively with the help of automated tools [IMP-C]. As a following step, proactive measures need to be installed in order to improve the data quality [IMP-D]. (Loshin, 2008) These improvement measures can be identification of double entries, correction of spelling mistakes and drop-down lists instead of free entry fields. [IMP-D] (Otto & Hüner, 2009; Pipino et al., 2002) Incorrect data need to be corrected. [IMP-D] [IMP-E] (Otto & Hüner, 2009; Pipino et al., 2002)

4.3.2.8 Data Usage [DTU].

This focus area is concerned about how to "enable data availability and distribution to appropriate users". (IBM, 2007) The organization must have a user administration taking care of the users of the data within the different systems. [DTU-A] [DTU-B] [DTU-C] (EDUCAUSE Center For Applied Research, 2009; Otto & Hüner, 2009)

It needs to be determined for which purpose the data is needed [DTU-A]. (Loshin, 2008) Basing on this, the organization can decide for whom they need to grant access [DTU-B] [DTU-C]. Data systems tend to get outdated easily, so it is important to trying to keep it up to date. [DTU-D] (Bowker, Baker, Millerand, & Ribes, 2010) Since users are on different levels of knowledge concerning the use of the technology, it is important to make them able to do so in order to have the employees seize all possibilities. [DTU-E] (Bowker et al., 2010)

4.3.2.9 Data Ownership [DTO] [DTO-A] [DTO-B] [DTO-C] [DTO-D] [DTO-E]

Ownership of data is an important factor to ensure responsibilities to be taken. It can help making sure that quality is maintained proactively. (EDUCAUSE Center For Applied Research, 2009; IBM, 2007; Loshin, 2008)

On the initial level, there is no existence of ownership models. However, the need is recognized and responsibility for data is given randomly [DTO-A]. (Loshin, 2008) It is often not clear whether data is owned by data processing or managing units. However, this is important so responsibilities are clear and the data's purpose can be defined. The data owner is accountable for content and quality. [DTO-B] (Winter & Meyer, 2001) Clear responsibilities for data must be assigned. [DTO-C] (Winter & Meyer, 2001) There should be one ownership model in place for the whole organization. [DTO-D] [DTO-E] (Hart, 2007)

4.3.2.10 Data Access [DTA] [DTA-A] [DTA-B] [DTA-C] [DTA-D] [DTA-E]

Access control must be established for the systems [DTA-B] [DTA-C] [DTA-D] (Anderson & Moore, 2007; Casassa Mont & Beato, 2007; EDUCAUSE Center For Applied Research, 2009; Loshin, 2008). The data must be distributed to users with the according rights. The access rights are established by responsible employees. [DTA-B] [DTA-C] [DTA-D] [DTA-E] (Bernard, 2007; Casassa Mont & Beato, 2007; EDUCAUSE Center For Applied Research, 2009; IBM, 2007) There must be roles and access rights for different user groups. Users get access to only what they need; there is no access to unnecessary data. [DTA-A] [DTA-B] [DTA-D] (Bernard, 2007; Casassa Mont & Beato, 2007; Otto & Hüner, 2009) There should be a policy in place defining

appropriate access rules for the employees. (EDUCAUSE Center For Applied Research, 2009)Since users are always at different levels of skills, interest and technological awareness, this needs to be pushed in order to motivate the employees to use the systems. [DTA-D] [DTA-E] (Bowker et al., 2010)

4.3.2.11 Data Protection [DTP].

Data protection is a key factor of MDM and therefore must be addressed as focus area. This is general consensus in models and literature. The targets are defining policies guidelines and controls for minimizing risks and protect the data. (IBM, 2007; Loshin, 2008; Shaw, Chen, Harris, & Huang, 2009) Security measures must ensure "confidentiality, integrity, and availability" of the sources. Physical precautions for data protection must be taken. [DTP-A] (Bernard, 2007; Borgman, 2000) Organizations more and more prioritize this topic, especially when considering the change of the working culture with people working from home using virtual solutions to access internal data. (Anderson & Moore, 2007; Gordon & Loeb, 2002) There must be security measures in place for access control and user right management. (Anderson & Moore, 2007) [DTP-B] [DTP-C] Furthermore, information security must be ensured by establishing security with safe passwords which the employees are obliged to use. This identity control ensures that only entitled users get access to the data, ergo it is secured from unauthorized access and the personal users' accounts are secured. So the data that a users works with is secured and the privacy of the users is secured. (Bernard, 2007)[DTP-D] People tend to lack awareness of data protection and privacy issues, so this must be actively raised among the employees. Standards must be enforced. (Anderson & Moore, 2007; Hwang, Yeh, Chen, Jiang, & Klein, 2011; Shaw et al., 2009) [DTP-E]

4.3.2.12 Storage [STO]

This focus area is about maintaining data objects through the life cycle and supporting the needs of the applications from a rather technical perspective [STO-A] [STO-B]. (Loshin, 2008) The storage solution is meant to be available, reliable, performant and consistent. [STO-A] (Bowker et al., 2010; Youssef, Butrico, & Da Silva, 2008) Tools should be used in order to reduce duplications [STO-C]. (Loshin, 2008) The storage setup regularly needs to be checked to ensure up-to-datedness. [STO-D] (Hwang et al., 2011) The organization should be storing the data with innovative technology to benefit from BI solutions delivering forecasts and analytics. [STO-E] (Loshin, 2008) Common data warehousing solutions help with analyzing data stored in traditional databases. Loshin also mentions this as one of his final capabilities. (n.d.)

4.3.2.13 Data Lifecycle [DLC]

Data elements undergo a data life cycle (Bernard, 2007; Loshin, 2008; Otto & Hüner, 2009). This is a fact the organization needs to be aware of [DLC-A]. (Casassa Mont & Beato, 2007) Data needs to be appreciated as an important asset because it aligns business targets and IT initiatives [LDC-B] (Bernard, 2007; IBM, 2007; Redman, 1995). Managing the lifecycle is an important task for an organization containing a guideline based way of using data over the lifecycle (IBM, 2007). The data characteristics need to be maintained in the way the applications require it depending on the status in the life cycle. The synchronization must be embedded in the life cycle. (Bernard, 2007; Loshin, 2008)[DLC-C] During its lifecycle, every logical data item needs a single source of truth which can establish high quality (Loshin, 2008). Therefore, its value needs be openly communicated. Redundant data items must be consolidated into a single source of truth

ensuring high quality. [DLC-D] The organization must be able to merge concepts that should be unique into a single one. (IBM, 2007; Loshin, 2008) Systems maintaining master data should automatically log changes as much as possible. Additionally, processes must be in place to monitor and document any changes on master data. The organization needs to define how its data entities are used over the life cycle, i.e. how they are created, read, changed or deleted by different systems. [DLC-E] (Loshin, 2008) It also needs to be clearly defined how the data is treated throughout the different steps of the lifecycle. [DLC-C] [DLC-E] (Otto & Hüner, 2009)

4.3.3 The dependencies.

Dependencies are restrictions that hinder an organization from executing an activity before another one was executed before. These dependencies are fairly logical and are integrated in the full capabilities description in the Appendix D. For instance, responsibilities for particular fields of data can only be taken if the data fields are defined beforehand. It is handy for the user to know these in advance to have in mind that some activities can only be conducted if others are done before.

The dependencies derive from logical conclusions. They were analyzed and defined with the experts. When the experts were consulted regarding the maturity matrix, they were also asked to consider possible dependencies. The dependencies were developed in an open interview. They were asked to think about dependencies for each capability. Resulting are the dependencies that are presented in the following. [MDM-A] displays initial attempts to design a model. This requires cooperation between different business units before [DMD-B]. Only when cooperation takes place, the model will be able to display a holistic view on the topic. [DLS-A] indicates an existing overview of

systems using master data. To achieve this, it is necessary to having at least partly a definition of what the organization understands under the term master data [DMD-C]. This is self-explanatory since one can only define systems using the data if it is known what the data comprises of. Furthermore, there should have been some preliminary attempts towards designing a model of master data. Also this capability is directly related to having an overview about master data [MDM-A]. [IOB-A] indicates that there is awareness in the organization about the impact of poor data in certain areas of master data. This requires an initial feeling about the actual tendency about the quality [ADQ-A]. This is because the employees can more easily estimate possible impacts if they know about weak spots already. To figure out whether the data quality does not suffice standards [IMP-A], the firm must have objective means defined to measure data quality [ADQ-D]. In order for the firm to know who is using which kind of data [DTU-A] it is clearly necessary that it can be made use of a shared definition of the term master data [DMD-C] to have criteria which data belongs to the group of interest. Furthermore, there must be data ownership in place [DTO-A] which is a precondition for gaining overview about who has defined rights to access. Additionally, the employees must all have access to the needed data [DTA-C], else data would be shared next to the official ways which would undesirable because of access rights restrictions. Data must be activated on request [DTP-B], so that everyone gets access if he needs it and no unauthorized people get it. Technical prerequisites for data protection must be in place [DTP-A], which requires a persistent way of storing the data [STO-A]. The capability requiring access to be activated on request [DTP-B], indirectly requires that access is denied to people that are unauthorized [DTA-B]. Guidelines must be established for treating data over their

lifecycle [DLC-C]. This has as prerequisite that there are responsible persons who communicate these guidelines [DTO-C].

The next table gives an overview about the dependencies. MDM-A

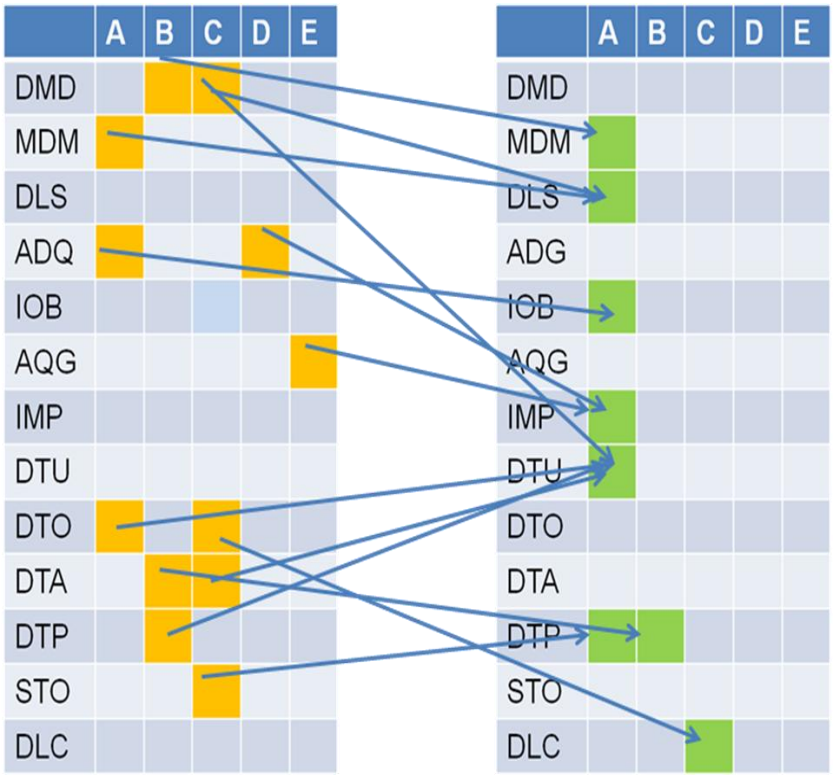
Table 12 *Dependencies between Capabilities*

Capability	Meaning	Requires First	Meaning
MDM-A	Initial attempts to design a model	DMD-B	Cooperative data definitions
DLS-A	Overview about systems using master data	DMD-C	Definition based on shared understanding
		MDM-A	Initial attempts to design a model
IOB-A	Knowledge about quality issues	ADQ-A	Feeling about data quality
IMP-A	Insufficient data quality is known	ADQ-D	Objective quality measurement
DTU-A	Known who uses which data	DMD-C	Definition based on shared understanding
		DTO-A	Data ownership
		DTA-C	Necessary data access
		DTP-B	Access on request
DTP-A	Technical data	STO-A	Persistent storage

	protection		
DTP-B	Access on request	DTA-B	Necessary data access
DLC-C	Guidelines for treating data	DTO-C	Communicated responsibilities

The following graphic displays the interdependencies. On the left side are the capabilities whose implementation enables the capabilities on the right to be implemented afterwards. So, if one capability on the right is to be implemented, it required the adequate ones from the left to be implemented already.

Figure 14 *The interdependencies between capabilities graphically displayed*



4.3.4 Influential factors.

This assessment method is meant to be used by companies of different size and business area. Therefore there might be influential factors for the maturity which cannot be covered by the capabilities. These factors were developed in cooperation with the experts and cross-validated. The influential factors are described in the following. If the organization is rather small, some actions tend to be too ambitious. Smaller companies have less data to process, less complicated structures and a smaller amount of complex information systems and software. Therefore, e.g. an elaborate data warehousing solution would be too much effort to deliver the desired effects. Furthermore, they might not have the human resources to allocate to those kinds of tasks. Resulting, it can be stated that smaller companies do not need to implement highly sophisticated solutions for master

data management to reach full maturity regarding their situational circumstances. The influential factors mainly relate to the size of the company, but still are independently regarded in the following since they cannot be generalized only to the size.

Form of organization. This describes the legal setup of the company; whether the company is a non-governmental organization, a public or a military one. Is it profit-oriented or not? If yes, is it noted at the stock exchange? These are factors influencing the data landscape. If the organization is not forced to generate profits and operate efficiently, there is less pressure on the organization.

Legal framework. If the company is set up in various locations across different countries, it complicates the data landscape. If it needs to operate in different countries, it needs to comply with different regulations, like tax and accounting standards. (Wegener, 2008) These need to be displayed in the data landscape. Complying with different standards complicates the setup. It might also be necessary to exchange data across the different companies in different countries. (Wegener, 2008) Another factor is the business; is the organization settled in a highly regulated business (e.g. banks) and has to deal with sensitive data.

Anonymity. Resulting from a particular size and/or the geographical setup in different locations, a certain degree of anonymity occurs. The reasons can also lie in the company culture. If there is a rather open, sharing culture, the people are more likely to know each other. A low degree of anonymity is good because people more easily know who their contact person is and who to ask in case of ambiguity. If the employees know each other in person, things can get clarified faster and less bureaucratically. The size

which is referred to as “big” is defined as at least 250. This follows the recommendation of the European Commission (2003).

Heterogeneity of processes and systems. If the organization needs to work with many different complicated IT systems, it is prone to have a complicated data landscape. Different systems rely on different data logic, which needs to be depicted in the data setup. This means, the data cannot easily be reused for other systems. Also, the use of different systems is likely to cause a variety of processes around the systems. The more interfaces there are, the more danger there is to lose information or to have more work in finding the same information twice.

The above described influencing factors differ from organization to organization. However, they have to be taken into account to guarantee the most possible objectivity. Therefore, they need to be included in the assessment of the maturity. The following table gives a comprehensive overview of how the influential factors impact the maturity model.

Table 13 *Impact of influential factors on MD3M*

Capability	Consequences of situational factors
DMD-E (company-wide definiton)	This capability is only applicable for companies that are part of a group and have to exchange data with other parties on a regular basis. For all others, this capability is disabled. (Legal Framework)
IOB-D (impact estimation) / IOB-E (impact valuation)	This does not apply for organizations not aiming for profit. So for non-profit, military or governmental organizations these capabilities do not apply. (Form of Organization)
ADG-C (patterns about bad data)	This capability does not apply for SMEs. In smaller enterprises the communication ways are so short that people know the requirements of other departments. (Anonymity)
DLS-E (inventory of data sources)	This capability does not apply for organizations with only few systems and processes around it. (Heterogeneity of processes/systems)

4.4 The MD3M Maturity Assessment Questionnaire

The MD3M Assessment Questionnaire forms an assessment measure for evaluating an organization's maturity level. The questionnaire contains questions for every capability in the maturity model and the influential factors. After answering the questions in the questionnaire, the spreadsheet tool will automatically calculate both the

individual maturity levels per key topic and the overall level for the organization. Every question is formulated neutrally and not suggestive in order to assess whether the content of the capability is fulfilled objectively.

The questionnaire is the result of the before developed maturity model. The questionnaire can be found in the Appendix F. For the situational factors, the questionnaire dynamically changes the questions to be answered. If a capability gets disabled, the respondent will not need to answer the question and the related maturity level will be disabled.

The questionnaire contains two lists of questions. The first set of questions tests the influential factors and the second asks about the capabilities themselves.

4.5 Final Version of the Matrix

The matrix provides an overview about the maturity of an organization. The matrix shows the capabilities for each key topic and focus areas sorted by maturity level. It can be seen whether the capability is implemented for one maturity level in a certain focus area or if it is still missing.

The left column shows the key topics and the focus areas. The grey shaded lines depict the key topics whereas the blue ones show the focus areas. The following five columns indicate whether a capability is implemented or missing for each focus area. This is displayed with either a red cell with the word “Missing” or a green one with “Implemented”. The next table depicts the matrix in an unfilled state, thus showing “missing” for every capability.

Table 14 *Structure of the MD3M Unfilled*

	Initial	Repeatable	Defined Process	Managed and Measurable	Optimized
Data Model					
Definition of Master Data	Missing	Missing	Missing	Missing	Missing
Master Data Model	Missing	Missing	Missing	Missing	Missing
Data Landscape	Missing	Missing	Missing	Missing	Missing
Data Quality					
Assessment of Data Quality	Missing	Missing	Missing	Missing	Missing
Impact on Business	Missing	Missing	Missing	Missing	Missing
Reasons/Sources for poor Quality	Missing	Missing	Missing	Missing	Missing
Improvement	Missing	Missing	Missing	Missing	Missing
Usage & Ownership					
Data Usage	Missing	Missing	Missing	Missing	Missing
Data Ownership	Missing	Missing	Missing	Missing	Missing
Data Access	Missing	Missing	Missing	Missing	Missing
Data Protection					
Data Protection	Missing	Missing	Missing	Missing	Missing
Maintenance					
Storage	Missing	Missing	Missing	Missing	Missing
Data Lifecycle	Missing	Missing	Missing	Missing	Missing

The next table shows the maturity levels overview divided by key topics. The single different key topics are depicted individually to have clearer insight into weak spots in the MDM of an organization. The overall maturity equals the lowest maturity

level from the key topics and the key topics' maturity levels are the lowest level from their capabilities.

Table 15 *Maturity Levels Overview*

Maturity Levels	
Data Model	0
Data Quality	0
Usage & Ownership	0
Data Protection	0
Maintenance	0
Overall	0

5 Application of the MD3M

To validate the model and provide practical information, the whole path is executed at a case company. This chapter describes the practical implementation of the before mentioned MDM Maturity Model. The case company will be introduced.

5.1 The Case Company

In this chapter, the case company, where the research took place, will be described. Since it is a global group, the whole structure and business will be explained before talking about the particular part of the company.

5.1.1 E.ON AG.

The E.ON AG is a stock corporation under German law. Headquarter of the whole group is situated in Düsseldorf, Germany. From there, the whole groups' activities are overseen and steered. There are five global units: Generation, Renewables, New Build & Technology, Gas, and Trading. In Europe, there are twelve regional units responsible for sales operations, regional energy networks and distributed-generation businesses. Furthermore, they form the contact for the global units. They provide the global units with market knowledge about e.g. political or regulatory issues. They also interact with the national stakeholders from politics or media. The twelve units are Germany, UK, Sweden, Italy, Spain, France, Netherlands, Hungary, Czech, Slovakia, Romania, and Bulgaria. Additionally, there is Russia as a focus region outside of the global unit structure. Supportive functions are organized functionally.

The regional unit in Germany is dedicated to power and gas distribution, energy sales and distributed generation in Germany. Energy sales are conducted and steered for

all customer segments, private and industrial. The network companies maintain power lines and gas pipelines through which energy is delivered. The high-voltage network is the interface between trans-European power transport and the regional energy networks. E.ON UK is a leading energy supplier and delivers power and gas to the whole variety of customers from private to corporate clients. They also provide maintenance and consultancy services concerning energy efficiency and partners a number of cities in terms of reducing their carbon footprint. Furthermore, the company is highly active in the field of combined heat and power and will soon operate the world's biggest plant in the UK. The Swedish unit provides and markets power, gas and heat and a variety of energy services. This unit is also responsible for the business in Finland, Denmark and Norway. In Italy, E.ON operates as an energy provider and is active in the fields of renewable energy and energy efficiency and conservation. The Spanish unit is mainly active in power distribution and retail. The French branch trades power and gas to industrial customers, is active in supporting the Generation and Renewables units and develops conventional, wind, solar and hydro capacities. The Dutch unit's core tasks are the delivery of power and gas to private and industrial customers in the Netherlands and Belgium. The Hungarian unit operates on the local power and gas market. The main fields are power trading and delivery of gas and power plus the depending services. The Czech unit is active on the gas and power market. In Slovakia, E.ON is operating on the power market with the tasks of delivering, trading and sales and services for both private and corporate clients. In Romania, the business focus is on gas and power whereas in Bulgaria, the portfolio covers power for the northeast. In the special focus area in Russia, everything is about power, from the generation to the delivery for industrial and

wholesale customers. There is the third biggest amount of power plants after Germany and the UK. E.ON attempts to make the energy supply safer, more efficient and more environment-friendly.

The whole E.ON group currently employs about 85.000 people and generates revenues of about 93 billion €. The biggest business parts are the generation of power from both renewable and conventional sources, trading of energy, the international gas business and the distribution of innovative energy solutions for clients. Currently there is a company shift from a European energy provider towards a global, specialized provider of energy solutions. This means that the company wants to become more competitive and achieve higher growth rates outside of Europe which leads to less capital need and more stakeholder value. Furthermore, they want to ameliorate the standards of energy provision, ergo cleaner energy. To achieve this, E.ON is researching and developing in the fields of highly efficient coal power plants, carbon capture and storage (CCS), new generation nuclear energy, offshore wind parks, modern and flexible gas power plants, biological natural gas, concentrated solar power (CSP), intelligent networks, micro plants with combined heat and power, gas heat pumps, electro mobility and energy storage.

Due to a complicated market environment, amplified by maximized regulations, the nuclear power phase out in Germany after the recent incident in Japan, sinking margins in the gas market and other reasons, E.ON needs to increase performance in order to stay competitive. This is why E.ON brought to live a program which is called E.ON 2.0. In a nutshell, it means that costs have to be reduced, structures facilitated to fasten decisions, and administration needs to be reduced to strengthen the operative

business. The program will lead to consequences in every country from 2012 until end of 2014.

5.1.2 E.ON Energy Trading SE

To be able to react to the dynamic wholesale energy market, E.ON connected its trading businesses under one roof. This was the birth of E.ON Energy Trading in 2008. EET oversees the generation of all north-west Europe. It determines which power plants are running for how long.

E.ON Energy Trading SE (EET) is a trading company in the energy sector based in Düsseldorf, Germany. The company is one of the leading energy trading businesses in Europe. EET serves as the link between the whole E.ON group and the world's wholesale energy market. It forms the commercial core of E.ON. EET sells electricity, natural gas, oil, coal, biomass, freight and emission certificates.

EET is operating at energy exchanges in Europe and the US and also in OTC markets. On those platforms, the company trades a broad portfolio of standard and exotic products. (E-ON Energy Trading, 2011a) At EET, there work over 1000 people from 40 countries. They are active on more than 20 exchanges and in over 40 countries. In 2010, EET conducted roughly 600.000 trades.(E-ON Energy Trading, 2011b)

E.ON Energy Trading acts as the interface between the whole E.ON group and the international energy market. The following table gives an overview about products EET is operating with.

Table 16 *Products of EET (E-ON Energy Trading, 2011a)*

	Products
Power	Spot, forwards, Vanilla Swaps and options, profiles, Dark/Spark spreads, location spreads, index products
Gas	Spot, forwards, Vanilla Swaps and options, profiles, location spreads, index products, formula swap, oil indexation
Emissions	EUA& CER forwards, EUA/CER swaps, time spreads, vanilla options on EUA, ERU, CER
Oil & Oil Products	Vanilla swaps and options, on EU/US/Asia crude, distillates and residuals, crack spreads and differentials, physical oil
Coal & Freight	Vanilla swaps on API2/4/6 and freight (FFAs), physical coal and/or freight at all major locations/routes
Weather	Vanilla swaps and options on temperature, precipitation or wind, cross-commodity (E.g. temp/gas) indices
Storage & Transport	Physical capacity swaptions

In line with the E.ON 2.0 project, EET will face changing circumstances in the near future. Within the next few years, EET will be merged with E.ON Ruhrgas, the biggest gas provider in Germany. This will lead to significant changes in the daily routine. The different company sites need to be integrated; and there will be personal changes for the employees.

5.1.3 Background and motivation.

Here, an overview about the motivation for this research within the case company is given. The organization has different motivational factors for investigating this topic. First, the extrinsic motivation deriving from the business field is described, then the E.ON-specific reasons are presented and finally, some explanatory details are given.

E.ON Energy Trading has several reasons for investigating the internal master data landscape and the improvement potential. The business, the company is operating in, is by nature very data-intense. The settlement and administration of trades in different exchanges and over the counter requires a lot of information. This whole business branch is strongly dependent on high quality data which needs to be highly available. If the data is incorrect, E.ON faces a high risk; the company might not be able to operate normally and lose profits. This topic was in-depth discussed in former chapters. Furthermore, the company faces a high organizational risk; side-effects on third parties might lead to legal problems. Therefore, not only from sensible reasons and a profit-thriving attitude, the organization needs to have the master data landscape under control from a legal perspective.

The organization itself also gives motivation to investigate the master data landscape. E.ON Energy Trading is a rather big company and belongs to huge group. Bigger companies have a more complicated data landscape. For those, it is especially important to have a good overview about the current situation. This company was founded as a merger of different organizations within the group which all had a focus on trading. Merging of organizations always leads to unequally more complicated master data setups. In this case, the different merged organization all had their own important

trading systems, which were all brought into the new organization. It was not possible to just leave out some because they all cover partially different markets and products, which are only displayable with huge effort in other systems. And the financial risk of unforeseen consequences is very high. Therefore, the systems were initially all included. However, this includes a very high maintenance effort. So, sequentially, now it is tried to make the setup more efficient and less maintenance intense.

Next to assessing the MDM maturity, the company is particularly interested in having a well-documented overview about the master data structure. More into detail, this means having an overview about the master data, having a graphical representation of it and the definition of major classes of master data. Additionally, EET wants to assess the risks; what are the origins of wrong data, what is the impact of wrong data on the trading business, what are the risks of inconsistent data and an illustration in use cases or examples. Furthermore, the case company is interested in having knowledge about the usage of the data in relevant systems, which means that they would like to get to know more about the source systems for each class of master data and about the maintenance of cross systems references.

Apart from that, a data quality assessment should take place. Herein, the questions of how to measure data quality, the impact of it and the assessment of quality should be answered. Building up on that, a master data maintenance concept will be developed, consisting of existing ways within the company, which will be evaluated and compared with best practices; a cost-benefit-analysis and a gap analysis. Because of the closer investigation of the company in different units while assessing the maturity, it will be possible to answer EET’s questions. For other employees, it would be quite difficult to

do so, because they would have expertise in one or maybe two areas, but not all of them. Additionally, they are busy with their daily tasks, so a thorough investigation would be difficult. Therefore, one resource who is only dedicated at this interdisciplinary project appears to be the best choice.

5.2 Practical Application at the Case Company

The practical application of the MD3M at the case company is described in this sub-chapter. This includes filling in the questionnaire and assessing the maturity level of the organization.

5.2.1 Filling in the questionnaire.

For the case company, this means that responsible and knowledgeable persons have to fill in the questionnaire. After this, the questionnaires will be evaluated. To ensure a higher data quality, the questionnaire filled out by the same experts within the organization which participated in validating it. The persons, who were chosen to fill in the questionnaire, are experts on the topic because they work with the master data in their daily work life. The experts were asked to fill in the questionnaire once. These answers were used to calculate EET's master data maturity. The results are in-depth presented in the next chapter. Furthermore, the experts were chosen to define dependencies between capabilities and influential factors. Table 9 presents an overview about the expertise areas of the interviewees.

5.2.2 Data modeling.

The development of models in order to visualize and make data structures understandable seems to be a well-suited approach. Models can be used to give an easier

description of a complicated, real world situation. (Wegener, 2008) Modeling the relevant data provides the user with a realistic overview about the scope of the data. It is important to regularly take care that the model stays up to date. (Wegener, 2008)

Due to the business field of EET, the data structure is rather complex. It is therefore modeled on a higher level to indicate the interdependencies between the different fields of master data. This perspective provides the reader with more clarity and an overview about important parts of the master data landscape. The current approach is rather a logical one than granularity-focused one.

After conducting several interviews with EET employees from different departments with a different view on the data and a different focus on the importance of data, a data model was developed. Since the data landscape in a global energy trading organization is very complex, the model is not developed with a data entity level granularity. It is meant to communicate the logical structure on a detailed level without losing focus. Furthermore, a more detailed granularity would increase the model's complexity while decreasing understandability. To also improve understandability, the data model was grouped. The data groups are depicted in appendix I. In appendix J, the whole model is presented. For the sake of comprehensibility, the different groups are differently colored in both the broad overview and the detailed model. The model was revised in a loop by employees of the case company who have a good overview about the data landscape since it is part of their daily work.

5.2.3 Special areas of interest of EET

During the placement at E.ON Energy Trading, I worked at some topics that are closely related to the master data maturity. The case company was particularly interested

in some factors related to the maturity, so emphasis was laid upon these next to the development of the MD3M. The organization is was interested in finding a shared definition of the term master data which would be usable in a trading house. This topic includes analyzing what data is used in the IT systems and who is using the data. Additionally, a grouping of the data into a model was desired.

Another area of interest is the assessment of risks. EET wanted to know where wrong data originates from. This also includes counteractions against those factors. On top of this, the impact from wrong data belongs to this area.

Furthermore, the organization has a stake in the development of a maintenance concept for the master data landscape. The goal is to find a competitive way of maintaining data with least effort and most flexibility possible.

The company’s interest overlaps in many regards, therefore the development of the MD3M and the research of the company’s interests are well to align in one thesis work without having interest conflicts.

6 Results

6.1 Results of the Implemented MD3M at the Case Company

In this chapter, the results of the questionnaire's application and the final matrix with EET's results are presented. The questionnaire was answered by the same experts that helped validating the questionnaire. Those experts qualify themselves because both their background experiences and their daily work life lead to a thorough understanding of the data situation within EET.

6.1.1 Influential factors.

The following table shows EET's answers to the questions concerning the situational factors.

Table 17 *Influential Factors at EET*

Influential Factors	
Does your company belong to a group and your company needs to interact regularly with other internal members of the group and exchange data?	Yes
Is your company a non-profit organization, and/or a governmental or military organization?	No
Does your company exceed a number of employees of approximately 250?	Yes
Do the employees need to work with many different systems for executing their daily work and have to follow different processes when doing this?	Yes

The table shows that EET does belong to a group which triggers capability 'Definition of Master Data'-E to be active. Since the organization is not a non-profit or a

governmental one, the capabilities ‘Impact On Business’-D and ‘Impact on Business’-E are enabled as well. E.ON Energy Trading exceeds the status of a SME, so ‘Assessment of Data Quality’-C will not be disabled. For the last influential factor, the existence of many different information systems enables capability ‘Data Landscape’-E. The following table gives an overview about the influential factors at EET.

Table 18 *Influential Factors at EET*

Influential Factors Question	Enabled (+) /disabled (-)	EET specific reason
1	Definition of Master Data-E (+)	Enabled because the EET belongs to a group and interacts with the different internal counterparties
2	Impact on Business-D (+) Impact on Business-E (+)	The capabilities stay enabled because EET is a private profit-oriented organization
3	Assessment of Data Quality-C (+)	This capability is enabled because the organization does have more than 250 employees and cannot be considered SME anymore.
4	Data Landscape-E (+)	The employees are confronted with many different systems in their daily work, so this capability is applicable for EET.

6.1.2 The maturity level of E.ON Energy Trading.

This chapter contains the assessed maturity level of E.ON Energy Trading. The table shows the maturity levels for each key topic and the overall one.

Table 19 *Maturity Levels of E.ON Energy Trading*

Maturity Levels	
Data Model	1
Data Quality	0
Usage & Ownership	1
Data Protection	5
Maintenance	1
Overall	0

The overall master data management maturity level of E.ON Energy is 0. The overall number sums up from the maturity models of the five key topics. This leads to a rather low result since the maturity is measured as the absolute minimum of levels for one focus area. So having one capability on the first level not implemented leads to a maturity level of 0. But this also means that implementing that particular capability will lead to a notable raise in maturity.

The single maturity levels are as follows. For the key topics ‘Data Model’, ‘Usage and Ownership’ and ‘Maintenance’, the maturity is 1. ‘Data Quality’ sums up to a maturity of 0. The fifth key topic ‘Security’ reaches maximum maturity of 5. The following paragraphs after the matrix will go more into detail about the single maturity levels.

Table 20 *Maturity Matrix for E.ON Energy Trading*

	A	B	C	D	E
Data Model					
Definition of Master Data	Implemented	Implemented	Implemented	Missing	Missing
Master Data Model	Implemented	Implemented	Implemented	Missing	Missing
Data Landscape	Implemented	Missing	Implemented	Implemented	Missing
Data Quality					
Assessment of Data Quality	Missing	Implemented	Implemented	Missing	Missing
Impact on Business	Implemented	Implemented	Implemented	Implemented	Implemented
Reasons/Sources for poor Quality	Implemented	Implemented	Missing	Implemented	Implemented
Improvement	Missing	Implemented	Missing	Implemented	Missing
Usage & Ownership					
Data Usage	Implemented	Implemented	Missing	Missing	Missing
Data Ownership	Implemented	Missing	Missing	Missing	Missing
Data Access	Implemented	Implemented	Implemented	Missing	Implemented
Data Protection					
Data Security	Implemented	Implemented	Implemented	Implemented	Implemented
Maintenance					
Storage	Implemented	Missing	Missing	Missing	Missing
Data Lifecycle	Implemented	Implemented	Missing	Missing	Implemented

Data Model:

In this key topic, the maturity level is 1. This means that all capabilities on the first level are implemented. That implies that the company has a basic understanding of master data in some entities. Additional, attempts are made in order to obtain an overview about the master data situation within the organization. In line with this, the firm owns overviews about systems using master data.

On the second maturity level, there are two of three capabilities implemented. EET has started to streamline definitions of the terms around master data to ensure a

common understanding. The different business units can define master data from their perspective. The first missing capability in this area is the overview of systems with read/write access on master data. This is not implemented yet.

The following maturity level consists of fully implemented capabilities. The maturity level of the whole key topic is still 1, because there is one missing capability on the level before. E.ON owns a master data definition that is based on different views. EET is thriving towards a global definition that is valid for the whole organization. The departments within EET know about different parties' definitions of master data. Furthermore, the company knows whether data is stored redundantly.

The forth maturity level is not implemented except one for capability. EET can neither provide an official companywide definition of master data nor a master data model that is agreed upon internally. However, there is an inventory of all data sources and by which systems they get accessed. Redundancies are found and there are plans in place to resolve them.

The final stage of maturity is not reached for this key topic. E.ON does not have interfaces for data exchange for other companies in the E.ON group. There are no rhythm and responsibility established for maintaining the master data model regularly. Redundancies are not fully solved yet nor is the data logic scalable and superfluous systems are substituted.

Data Quality:

The maturity for this key topic is 0. This value is that low because two capabilities on the first level are not implemented. The company does not have a feeling about the quality of data and in which regards the quality adds value for the company. Furthermore,

there are no processes in place to figure out the sufficiency of data quality. Two other capabilities are implemented however. The employees at EET are aware of the fact that bad data quality affects the business' reputation. Furthermore, they know about different reasons causing bad master data in the firm.

The following maturity level is already completely implemented. EET has clearly stated which aspects are part of data quality from their point of view and need to be measured. Additionally, they know that issues in data quality impact the business as direct monetary loss. They also know which reasons for poor data quality are relevant for EET. On top of that, E.ON knows how important good quality is for efficiency and effectiveness.

For maturity level three, the result is 50-50. The data quality is defined for different stakeholders requiring different levels of accuracy. EET can state how much monetary deficits are caused by poor master data quality. However, there are no patterns investigated where bad quality origins in and there is no benchmarking tool in place to measure sufficient or insufficient quality.

On level four, all but one capabilities are implemented. Master data quality is not measured objectively and it is not defined for each piece of master data which quality is has. On the other hand, EET knows precisely how much and how poor master data impacts the reputation of the firm. The people know the sources and origins of bad master data and to what consequences they will lead. There are measures in place to improve data quality.

On the highest maturity level, there are two capabilities implemented and two missing. There is no regular assessment of data quality and there is no benchmarking

system according to which the quality is assessed. Still, the organization can classify in financial arguments how poor master data quality impacts the business from a monetary and a reputational perspective. Furthermore, they know reasons for bad quality and where the weak spots in the data setups are.

Usage & Ownership:

In this key topic, the maturity level is 1. This means all capabilities of the first level are implemented. EET knows who uses which kind of master data. The data elements have a dedicated owner and access is granted following a defined process.

On the following maturity level, one capability is missing. The ownership of data does not follow logically consistent roles. Purpose, usage and content are not defined by the owner. However, it is public if every employee has access to needed data and the employee knows where to obtain it. Data access is denied to unauthorized people.

The third level counts one implemented capability. Every employee has access to data he needs in order to fulfill his job. The employee is possibly not given access to every single source of data that he might need. These sources are possibly not communicated to him. Responsible people for parts of data are not communicated throughout the company.

The next maturity level does not contain any implemented capabilities. Data repositories are not maintained regularly and there are no data stewardships established. The access to unnecessary data is not restricted to the users in order to avoid confusion or possible access right issues.

On the final maturity level, two capabilities are not implemented. The employees do not necessarily use all possibilities they have and might be reluctant to use some

useful systems or applications. Data stewardship as a concept is not promoted within EET and data quality standards are not defined and adhered to. It needs to be considered that every employee knows to which sources he has access and what he can find there to help with his task.

Data Protection:

In the area of data protection, the maturity level sums up to 5. All capabilities are implemented in this key topic. All technical requirements for data protection are fulfilled for the master data. Access to data must be requested and will only be activated after request. There are clear rules at EET for who might get access to certain areas of master data. There are passwords in place for systems with data access that need to adhere to security standards. Those passwords need to have a certain quality and must be changed regularly. EET's employees have a raised awareness for data protection and e.g. do not leave their computers unlocked while unattended.

Maintenance:

For this key topic, the maturity model is 1. Both capabilities for the first level are implemented. This implies that the master data is stored in a persistent and performance-oriented way and the organization knows about the lifecycle of their data and that data and its structure changes over time.

On level 2, one capability is implemented. EON treats data as an organizational asset, but does not check the data logic regularly for up-to-datedness.

The third maturity level is not at all implemented, resulting in EET not having tools automatically checking for redundancies and duplicates. Also, there are not guidelines of how to treat data throughout the lifecycle.

On the forth level, no capability is implemented as well. The data base logic is not regularly check for performance and up-to-datedness. Furthermore, there is no single source of truth concept established for every data element.

The final maturity level contains one implemented capability. The entering, changing and deleting of data is automatically logged to facilitate auditing and decrease maintenance effort. However, the data is not stored innovatively to enable forecasting and analyses like BI solutions.

The total relation of implemented capabilities to missing ones is approximately 3/5 to 2/5. Almost 60 percent of the capabilities are already implemented.

Table 21 *Comparison Total Number of Capabilities*

Capabilities Total	Implemented	Missing
65	38	27
100 %	58,5 %	41,5 %

For every maturity level, the amount of implemented and missing capabilities is shown in the next table. Every maturity level consists of 13 capabilities to be implemented. There are 13 focus areas which all have 5 maturity levels, therefore the number 13. The numbers in the table indicate the total amount of implemented or missing capabilities and the italic ones in brackets stand for the percentages. There is a correlation between rising maturity level and missing capabilities. This means that on a higher maturity level, fewer capabilities are achieved.

Table 22 *Maturity per Level*

Maturity per Level	Total	Implemented	Missing
1	13 *(100%)*	11 *(84,6)*	2 *(15,4)*
2	13 *(100%)*	10 *(76,9)*	3 *(23,1)*
3	13 *(100%)*	7 *(53,8)*	6 *(46,2)*
4	13 *(100%)*	5 *(38,5)*	8 *(61,5)*
5	13 *(100%)*	5 *(38,5)*	8 *(61,5)*

The next table shows the implementation degree per key topic and also split up per capability. There is an overall tendency to be seen from more implemented capabilities towards less implemented ones throughout the rise of maturity levels. This can be derived from the previous table. The first maturity level is with 84,6 % implemented and the last two one have a less than half as high implementation quote.

The key topic ‘Data Model’ is 60 percent fulfilled. The key topic contains 3 focus areas, ergo 15 capabilities.

Table 23 *Maturity Levels per Key Topic*

Key Topic	Total Capabilities	Implemented (Levels)	Missing
Data Model	15	9 (3,2,3,1,0) *(60%)*	6 (0,1,0,2,3) *(40%)*
Data Quality	20	13 (2,4,2,3,2) *(65%)*	7 (2,0,2,1,2) *(35%)*
Usage & Ownership	15	7 (3,2,1,0,1) *(53%)*	8 (0,1,2,3,2) *(47%)*
Data Protection	5	5 (1,1,1,1,1) *(100%)*	0 (0,0,0,0,0)
Maintenance	10	4 (2,1,0,0,1) *(40%)*	6 (0,1,2,2,1) *(60%)*

6.2 Internal Findings

This chapter discusses the results in terms of the interests of the case company. All resulting suggestions are presented in the following chapter.

6.2.1 Definition of master data.

It was a goal to find a shared definition of master data for the organization. This definition was jointly achieved by a team of people from different departments with both business and IT focus. The target was to have a shared definition that everyone agrees with and that can be used to solve ambiguity concerning the terminology. Basing on preceding literature research, the following aspects were agreed upon. Master data includes key information on the operational business. This data serves as the basis for transactional processes, operations and reporting purposes. It is in general used by several

systems and user groups and processes. It may include reference data, which is only used by single entities, but under circumstances very relevant for the whole business.

Having this definition in place helps the different business units and people talk about the same things. Furthermore, building upon this, it is possible to establish a master data model. Afterwards, it was therefore possible to implement the master data model.

Some observations were made when working on this topic. It is possible to figure out the single data elements that are used in the systems. However, this is very complicated and long-lasting because there are loads of different data entities to be entered. Additionally, their designations differ throughout the different systems and the different business units. So, the company needs to define whether they need the deep data-entity granularity or if they rather need a more high-level one which can serve as a guideline and where one can sort the deeper elements in. In this case, the decision was made to stay with one model that is not on a data element level. It would have been so much more effort to figure this out and there is no urgent direct need for this. So, if there was the need, it could still be analyzed and mapped to the existing model.

6.2.2 Risk assessment.

The area risk assessment includes assessing reasons for wrong master data as well as the impact. For E.ON Energy Trading, the assessment of risks brought the following results.

A big chunk of reasons for poor master data is the wrong entering. This includes simple typing errors, wrong data types or errors in the zip codes or phone numbers. A company-specific example would be if traders entered the wrong information. In one system, for internal deals the traders have to enter a buy and a sell deal separately. If they

make a typing error, the other people working with this data have a lot of effort to find the deal. Other systems do this automatically to decrease that effort.

The other big source for errors is those when it is assumed data is correct when it is not. Then assumingly correct data is entered into the systems and then the wrong data spreads. If that data gets stored in different sources, it will be very difficult to extinguish the errors.

Of course, this reminds of the chicken or egg dilemma. Often it is not clear which problem was there first, once it spread. It is therefore not that important as long as the existing problems get solved and new ones get avoided by the means of intelligent measures.

One problem that came across during this thesis work was about price curves. Price curves are curves that are used to calculate prices for a product. These curves base on public calculations from providers like Bloomberg or Reuters and get enriched with an internal calculation. There are two types of price curves; internal ones which are used for internal calculations and external ones for pricing. Due to the complexity of the business and the organization's size, there are countless numbers of price curves. Unfortunately, these curves are not documented well and there is no owner, so nobody can be sure where the curves come from or what purpose they serve. Furthermore, there was no understandable way of naming the price curves, so knowledge could get derived from them.

The impact that the above mentioned problems cause, can be divided into three different categories. Those categories were developed with expert consultants on the topic and help classify the impact. The next table visualizes this.

Table 24 *Impact and Criticality*

Impact	Criticality for business
Regulatory	high
Financial	medium
Economic	high

The first kind is of a regulatory nature, it can ruin the entire business. If problems with master data results in violation of laws or regulations, the organization might get fined or have to defend itself in a lawsuit. In the worst case, the business must be shut down. This could be the case for instance if sensitive data was accessed by unauthorized people or published or lost.

Another field is the one of financial impacts. Direct financial impact is referred to when invoices are wrong; a delivery gets shipped to a wrong destination or in wrong quantities and this happens due to wrong data. This results in monetary losses, the organization will have to pay for the inconveniences. Certainly, the company does not appreciate this and it decreases its efficiency, but is does not substantially harm the business. Therefore the impact is not highly critical, but still critical to harm if it happens too often.

The final group is the economic impact. This one is highly critical and it is harming the whole business. If some things are constantly going wrong because of the wrong data, it will harm the whole organization. This could be wrong invoices as well as wrong payments or shipping, but in a high frequency. The customers will lose their trust and change to the competition. The reputation will irreversibly be damaged and the press

will publish these findings. Then, no other organization will be making business with them.

7 Discussion

This chapter will discuss the findings of the research. It will as well address the MD3M development as the practical results. It will give conclusions from an academic point of view and provide advice to the case company basing on the model's findings.

7.1 Evaluation of the MD3M and its Improvement Potential

In this chapter, the MD3M will be evaluated. It will be investigated whether the setup is practical and able to provide the desired insights.

7.1.1 The maturity levels.

The MD3M consists of five levels of maturity. The levels' descriptions are kept rather broad because they are used to describe the maturity level of very distinct capabilities situated in different topics. The level aims at being able to describe all capabilities properly.

The first level is called 'initial'. It describes a first awareness on the topic of MDM for the different focus areas. The second one displays a 'repeatable' state, meaning that insular measures have been initialized in different departments, but they have no connection to other projects. The third 'defined process' level indicates first collaborations and awareness on an inter-divisional level. The fourth level contains 'managed and measurable' processes that are in place in the organization. Processes are defined. The fifth one is 'optimized'. On this level, the efficiency has been improved due to MDM measures. The attention is on a tactical level.

The levels proved to be of an appropriate granularity. The levels cover enough information to be helpful in assessing the maturity and providing a status update on the

current situation in an organization. Additionally, they are not too specialized that the interviewees would get lost in details that they could not answer without consulting specialists for every capability. Experts with a good overview about master data activities can answer the questions without too much effort. This is also an important factor. If answering consumed too much time, the organizations would be reluctant to participate because the benefits would only show after a lot of input. For this, the resources might not be available.

So, concluding, it appears that the granularity is on a good level and helps both practitioners and academics in assessing and comparing master data maturity levels in different organizations.

7.1.2 The capabilities order

This subchapter investigates whether the ordering of capabilities makes sense. This subchapter deals with the question if the capabilities are correctly ordered along the maturity levels.

It is tough to state levels of maturity for such a complex topic. However, the process of maturity level deployment included a feedback loop. First, the levels were defined based on thorough literature research and comparison of best practices. Then the levels were validated by experts. This procedure ensures a certain degree of quality and academic relevance.

However, to fully exclude doubts and to possibly fine-tune some arrangements, it would be advisable to have it validated with more experts from different experts from different industries and companies. Furthermore, it could be a good idea to apply the

model at more case companies not only to assess the maturity levels of those, but also to validate the results.

When comparing different maturity levels, it should be expected that on average, the number of implemented capabilities decrease with rising level of maturity. The implemented maturity model at hand already indicates this. For level 1, 84,6 % of the capabilities are implemented, on the second level 76,9 %. The third level has an implementation rate of 53,8 % and the fourth and fifth both reach 38,5 %. This result gives valid reasons for the assumption that the levels are ordered logically.

7.1.3 The capabilities

This chapter is dedicated at analyzing whether the capabilities within the key topics are relevant. These will be discussed grouped into the key topics. As with the capabilities order, the capabilities themselves were developed based on literature and best practices and then validated by experts. So, it can be assumed that these are reliable.

The key topic ‘Data Model’ consists of 60 % implemented capabilities. Overall, there is a development to be seen that indicates a tendency towards less implemented capabilities with rising maturity. This justifies the impression that the ordering is reasonable. This key topic contains focus areas about modeling the data and defining relevant terms. These are implemented to a point where the organization obtains much status quo information but is not yet focused on future maintenance and improvement of these models. However, this logically is the following step after realizing, assessing and improving a current situation.

Table 25 *Overview Implemented Capabilities for Key Topic Data Model*

Key Topic	Total Capabilities	Implemented (Levels)	Missing
Data Model	15	9 (3,2,3,1,0) *(60%)*	6 (0,1,0,2,3) *(40%)*

‘Data Quality’ contains 65 % implemented capabilities. There is a rough tendency toward decreasing implementation on higher levels. The focus area ‘Impact on Business’ if fully implemented and the reasons are fully but one. There is a high criticality to be seen here. A company must know what the impact of certain issues is. The assessment and improvement of insufficient data suffers from less attention. Incorrect data can manually be improved, so it is not necessarily a highly critical problem. However, it slows down efficiency and effectiveness.

Table 26 *Overview Implemented Capabilities for Key Topic Data Quality*

Key Topic	Total Capabilities	Implemented (Levels)	Missing
Data Quality	20	13 (2,4,2,3,2) *(65%)*	7 (2,0,2,1,2) *(35%)*

The key topic ‘Usage & Ownership’ consists of roughly the same amount of implemented and not implemented capabilities. Still, the amount of implemented capabilities is decreasing. This key topic contains capabilities about effective organization and access of data. Within ‘Data Access’, all but one capability are

implemented. Having a decent and complying data access policy is more critical than having lean master data responsibility concepts in place.

Table 27 *Overview Implemented Capabilities for Key Topic Usage & Ownership*

Key Topic	Total Capabilities	Implemented (Levels)	Missing
Usage & Ownership	15	7 (3,2,1,0,1) *(53%)*	8 (0,1,2,3,2) *(47%)*

For data protection, the case company implemented all capabilities and therefore reaches full maturity in this key topic. This leads to the conclusion that the area security is a very important topic that is highly critical to be implemented.

Table 28 *Overview Implemented Capabilities for Key Topic Data Protection*

Key Topic	Total Capabilities	Implemented (Levels)	Missing
Data Protection	5	5 (1,1,1,1,1) *(100%)*	0 (0,0,0,0,0)

‘Maintenance’ contains mainly efficiency elements. This key topic is only implemented by 40 %. Still, there is a slight tendency towards decreasing implementation levels. This area again shows that efficiency matters stay behind security ones. The data is stored in a secure way, but it is not taken care of a technical, up-to-dated way that helps decrease effort.

Table 29 *Overview Implemented Capabilities for Key Topic Maintenance*

Key Topic	Total Capabilities	Implemented (Levels)	Missing
Maintenance	10	4 (2,1,0,0,1) *(40%)*	6 (0,1,2,2,1) *(60%)*

The above mentioned observations show that the capabilities are of different criticality and some capabilities on higher levels are implemented, even though other lower capabilities are not yet. This leads to the conclusion that first critical factors like security need to be considered before going into efficiency issues.

7.1.4 General conclusions.

Whilst developing the maturity model and applying at the case company, it appeared that there are three big improvement areas, namely knowledge management, process management and data landscape management. Knowledge management is about managing what the people in the organization know and making sure that nothing gets lost. In general, it is important to have a functioning knowledge management policy in place and create an atmosphere in which sharing is more appreciated than employees gathering knowledge for themselves. Process management is all about creating feasible, lean processes that guide the employees and do not create any unnecessary overload, but still considering all necessary steps. The last field data landscape management is about the set up of the data. This needs to be in form and reflect the reality in a sensible manner.

The key topic ‘Data Model’ is about the streamlining definitions. Different business units have different understandings of the same key terms due to different perspectives on the same topic and possible ignorance of others’ requirements, which is a state to be

terminated. Actions in this field clearly belong to the field of knowledge management which is about sharing knowledge across an organization. Furthermore, it belongs in the category of process management because clearly defined processes could help with this problem. A structured data model delivers much precious information about the data landscape in an organization which is important to keep a good overview in order to see improvement potential.

The key topic ‘Data Quality’ needs processes and knowledge management in order to become mature. In order to figure out the actual quality of data elements, there must be processes in place to reach all business units and to have a reliable path to walk along. Knowledge management is definitely needed when trying to figure out the quality of data. Since quality here is defined as fitness for use, the fitness can only be evaluated if every business unit shares their requirements. It is also relevant that business units share information in order to define the impact that quality can have on the whole business.

Within ‘Usage & Ownership’, all three aspects are covered. There must be processes in place for granting or denying access to systems or data. Everyone must know how he can get access and the ones deciding must have processes to know what to do. Knowledge sharing is important because this helps defining which user groups need access to which data. Additionally, a data landscape overview is beneficiary for users to having an overview about existing sources of information, so nothing is set up or saved redundantly.

In the field of ‘Data Protection’, process management is an important factor. To keep data safe and secure, there must be processes in place that must be adhered to. Only when this is given, data is secured against unauthorized access or other violations.

Concerning ‘Maintenance’, Data landscape has an important role. This key topic is about how to store the data and maintaining the data lifecycle. Therefore, information on storage or data lifecycle, which are by definition close to the data layer, belong fully to the field of data landscape.

These three improvement areas can be approached either reactively or proactively. The reactions of the MD3M’s results are clearly focused on improvement of the organizational setup towards the management of master data. Actions can be classified into two categories from the point of time, at which the snapshot is taken. Both approaches are important because the old mistakes and inefficiencies need to be resolved but the roots must be extinguished in order to avoid upcoming mistakes. First, there can be reactions to the situation. Beyond that, there are proactive improvement possibilities. An example for reactive would be correcting errors in data that are already in the systems. Proactive measures are changing the way the data are entered; e.g. implement a system where elements to be inserted are not written down manually, but chosen from drop-down lists in order to avoid spelling mistakes.

The following table presents the action categories and the improvement areas in a matrix.

Table 30 *Improvement areas and action categories*

	Reactive	Proactive
Knowledge Management	- assess the knowledge sharing culture	- provide a forum to share knowledge
Process Management	- revise existing processes	- setup new improved processes
Data Landscape Management	- improve existing data	- improve data setup to avoid mistakes

7.2 Advice for E.ON Energy Trading

This chapter investigates advice for E.ON Energy Trading in terms of improving their master data management.

EET's master data management maturity level is 0. The company is strong in direct security-related issues. All topics about saving data securely and having reasonable access rights are fully or almost fully implemented (Data Protection, Impact on Business, Data Access, Awareness of Quality Gaps). These focus areas are directly related to security issues.

The weaknesses definitely lay in the efficiency related areas. The data is stored safely and securely but there is no emphasis on innovative ways of storing to maybe facilitate forecasting. There is a master data model in place, but there is no process established to keep it up to date. Also, there is no ubiquitous overview about access to data and systems. There is no high quality awareness among the employees with

understanding for the processes behind the data. Little communication and action is done about data sources, access and usage. The data storage is functional but not innovative.

7.2.1 Improving the MDM maturity.

In order to improve maturity, the organization should raise awareness and a feeling among the employees for the importance of good quality data. On top of this, the organization should figure out definite areas where the data is insufficient. This would leverage the maturity level to 1. Improving these aspects does not include major structural changes in the organizational architecture, so it should be easily implemented.

Following on that, there are only three capabilities to be implemented to reach maturity level 3. The organization should take the time to investigate the systems and their reading and writing access on data. Building upon this, EET can give responsibilities for data to consistent people or departments. This certainly requires some amount of effort. The organization needs to define who should be having the responsibility for which kind of data because of what reason. This might require some degree of process reengineering and could therefore mean a certain effort. Furthermore, the organization should regularly check whether the data logic is still depicting the real life scenario. Starting with implementing the missing capabilities on this level, the focus shifts towards process and organizational reengineering for efficiency increase.

To reach a maturity level of 4, the organization will need to put effort in. Only five of thirteen capabilities are implemented on this level. The organization needs to work towards one shared definition of master data focusing on the sharing aspect. On top of that, an objective measurement of master data quality is needed within EET. Additionally, they should focus on the key topic ‘Data Usage and Ownership’. There

should be a concept in place for keeping data repositories up to date to keep them useful. Furthermore, EET should introduce data stewardship in alignment with new responsibilities for data areas.

For achieving full master data maturity, the organization needs to establish interface which can help interactions with frequently contacted counterparties. The data model and the data sources inventory needs to be regularly revised to keep them up to date. There also need to be regular objective quality assessments along the internally defined benchmarks. On top of that, it needs to be evaluated if users are able to seize the functionalities of the systems and actually do it. If no, this needs to be enforced. Data ownership must be included in the roles of the employees and integrated in their work life. Finally, the organization might consider implementing high level BI solutions to enable data forecasting.

7.2.2 Further advice on improvement potential.

The advice that could furthermore be derived from the research is structured into three management areas and two action levels. Along the areas knowledge management, process management and data landscape management and the distinction between reactive and proactive it is described what the organization can improve.

Knowledge Management.

EET should develop a better sense for knowledge sharing. Knowledge sharing is very important. If knowledge is not shared and a person leaves, the knowledge is lost. Also if other people depend on special knowledge, but do not know that someone has it, it is worthless. A practical example is the one with the price curves that was mentioned before. Price curves are master data because they are static formulas. The actual

instantiations are fed with dynamic prices, but the general formula itself is static. Had there been a better knowledge management for maintaining them, it would not have been so difficult to understand the price curves and the naming of them. Therefore, a knowledge sharing culture must be in place. It is important to convince the people of the advantages of shared knowledge rather than keeping knowledge to themselves to be considered as experts. A good understanding of other business units' daily work is beneficial. Regular knowledge sharing events can be established, e.g. business units within a logical group explain their daily work to each other and people with expert knowledge share it with people for whom it is of interest. Furthermore, if tasks get allocated to other people, an intensive handover should take place. The old owner should present all relevant information to the new task owner. There should be a guideline for that, so this will not be forgotten or left out due to time issues.

One topic that came up which is a good example for the problem of knowledge getting lost. Price curves are calculations on prices published by certain providers for special trading products. On these prices, somehow calculated margins are added. These are used either for internal or external pricing purposes. These curves are used by the traders when conducting a deal. They are also used throughout the whole trade lifecycle which makes it complicated. The curves are neither consistently named, so their purpose cannot be derived, nor is there a proper documentation. So, none of the involved parties can be definitely sure that he chose the right one. This can lead to errors in pricing deals which can end in critical situations. Once it happened, it is very difficult to resolve the issue. The impact of wrong pricings can be very critical because agreements might not be adhered to, which would lead to legal consequences. It turned out that the people who set

up those price curves are either unknown or have already left the company. Getting this issue solved will be a lot of effort and involve different parties. There is no process place to manage the knowledge in this area. This impacts the topic process management and also data landscape management because they data landscape is confusing and creates overload. It thus is advisable to first invest the work to assess the existing price curves and figure out their purpose. As a next step, it would be good to implement a naming convention for those curves and similar objects. The convention would have a legend explaining why it is called like that and what it means. This decreases the maintenance effort and questions around those curves can easily be answered. It is taken care of that the knowledge is being shared and cultivated.

Process Management

Along the different trading systems, the processes massively differ for the setup of products. All these systems contain master data. These systems are differently structured because they are proprietary solutions from different vendors. Different departments set up analog data while following different processes. Many business units are involved like, business, IT, legal and risk departments. This fact comes with certain risks; due to the complicated structure, knowledge might get lost if someone leaves the company. The procedures are so different that this is no common knowledge for the majority of staff. Furthermore, it is possible that responsibilities are given to people who do not have the right to work with them. This implies audit threat especially if it is about sensitive data. The different setup leads to confusion among and unnecessary overload for the employees. Additionally, knowledge is likely to get lost. If one expert within a specific area leaves the company, the knowledge leaves with him. It is difficult to keep track of all

knowledge to share when the processes do not follow a similar structure. Therefore, it is strongly recommended to establish one standard process for similar tasks and clearly document the necessary process steps and responsibilities within. For the organization, all previously involved parties should be involved as well to not lose any valuable knowledge. This includes both reactive and proactive steps, reactive is the assessment of the existing situation and proactive is planning and implementing of improvement methods.

Data Landscape

EET has to deal with a relatively high amount of complex trading systems. All of them serve a different purpose and therefore are relevant for the current point of view. Cutting off a system does imply a lot of planning and many risks. The systems all rely upon a different data logic, ergo need different databases and possibly interfaces to be able to interact. These systems however, use mainly the same data. Therefore, it would be perfect to use one underlying database system. Unfortunately, this is rather infeasible due to the complexity.

Putting a certain effort in the data landscape will nevertheless pay off with good data quality and leaner processes. A big improvement would be avoiding errors automatically. For this, there are several measures to be taken. The first one which happens in the background is duplicate recognition. The databases should regularly be checked for redundancies based on heuristics. These, of course, need to be checked manually or at least approved. But a tool could go through all the data and compare similar ones, for instance, if two entries only differ because of transposed digits.

Furthermore, the goal would be to automate as many entries as possible. Units should be predefined, like volumes, weights etc. This measure will avoid calculation mistakes because of wrong units or conversions. Additionally, regular expressions should be used in order to make sure that the data is entered in the right format. Regular expressions mathematically express patterns which occur in certain texts that need to follow restrictions (phone numbers, email addresses, postal addresses…) For instance, an email address always needs to contain the letters @ and .(name@example.com). So, it can be validated whether the email address entered complies with those regulations by automatically checking for the necessary elements. Furthermore, the endings can only consist of a number of predefined endings (.com, .nl, .de, etc.). It can furthermore be checked whether this applies as well. With those lists in the background entries can be checked for plausibility. Also postcodes, phone numbers and many other things follow these patterns.

Moreover, there is the possibility of implementing select lists whenever possible. A list is offered and a value is selected by the user instead of typing it in. After the first selection, other selections are possible depending on the preceding choice. An example would be addresses. First, the country is chosen from a list. Then, the correct address format is chosen and a new list for the city is provided. After entering the street, the zip code can automatically be generated. This approach brings the benefits of avoiding both spelling mistakes and wrong entries because of non-existing streets etc. Also the addresses adhere to the country's standard.

Furthermore, the data quality should be assessed regularly with automatic means. This needs to be reviewed manually and degradations will be seen soon. Additionally,

data should always be entered into only one source. There must be a single source of truth. If the data needs to be stored at different places simultaneously, the second place should be transitively fed from the original source.

Table 31 *Improvement Matrix for EET*

	Reactive	Proactive
Knowledge Management	- handover of tasks	- knowledge sharing initiatives - improve understanding of other business units' activities
Process management	- assess current processes	- streamline processes for similar activities across different systems (like data setup)
Data Landscape Management	- Assess data quality per system along objective criteria	- Duplicates recognition and resolution - Single source of truth - Server sided validation of entries - Dropdown lists - Relation checking

Concluding, it can be said that E.ON Energy Trading is well on track regarding security measures. However, the efficiency parts can be improved. This can be done in some areas within the areas of knowledge management, process management and data landscape management. Improvements in these areas will help making the working steps

faster and more efficient. Less knowledge will get lost and people will have a better understanding about activities that take part cross-divisional.

8 Evaluation and Limitations

This chapter covers the academic validity of this research and describes limitations and further research opportunities.

8.1 Evaluation

To ensure validity of this research, it will be evaluated along Yin's criteria (Yin, 2003). Yin proposes for single case study research to 'Construct Validity' through the usage of multiple evidence sources and through establishing a chain of evidence (Yin, 2003). This takes place when collecting data. The MD3M is based on several sources of academia and models from practice. These were investigated and compared to serve as a basis for the developed model. Additionally, it is suggested to have the key informants review the draft study. This was conducted via evaluation loops with the experts that were consulted. For ensuring 'External Validity', it is advised to make use of theory in single case studies (Yin, 2003). The research design is based on a thorough literature study and a comparison of existing models. For ensuring 'Reliability', it is recommended to demonstrate that the study can be reconducted and achieve the same results. This study is thoroughly documented, so a researcher would be able to conduct this research at another case company and is likely to achieve the same results. The results are extensively documented, so can serve as input for a case study database. Since this is the first study in this particular field, there cannot be a database filled with equivalent studies.

The adherence to the described criteria shows the validity and soundness of this research.

8.2 Limitations and Further Research

One possible limitation of the research at hand is the small amount of case companies. This might lead to restricted generalizability. Therefore, a higher amount of case studies at different companies would uncover possible shortcomings. Thus, applying the MD3M at different companies would give comparable results. The experts could be asked for feedback and show if there are any aspects left out. It would be particularly interesting whether there are differences for companies of a different size. Additionally, the more frequent application could help uncover possible inconsistencies regarding the capabilities.

Another improvement opportunity could be the feedback by experts. More experts could have been consulted to comment on the model and give possible improvement ideas. Alternatively, there could have been another approach of developing the matrix criteria. Experts could be asked to develop their own criteria in which maturity is to be achieved. The downside of this approach that it is very time consuming. Developing a model that covers all or at least most possible factors will take time. Then, cross-checking would have been needed. There, the experts could have discussed the developed approaches and derive one optimized one together. Again, this would be very time consuming because it must be based on research before.

Another approach could have been to derive the maturity model first from practical opinions and then cross-check it with academia. For this, experts from different companies could present their ideas which would be compared and then brought in relation to existing literature on the topic.

A possible different approach could be to develop different maturity models for different sizes of organizations. Now, for SMEs, the approach is rather complicated compared to their size. Maybe it would have been better to investigate those groups apart from each other and develop a lighter version for organizations which are not as complex as multinational ones. This could maybe be developed in a dynamic questionnaire. If certain size criteria apply, there will automatically be another model with fewer capabilities.

The improvement matrix was developed as a result from the analysis. It could be investigated whether this is a valid concept and could be elaborated on to help practitioners structure their improvement approaches in the field of master data.

Furthermore, an interesting aspect would be if the different capabilities have different weights. If some are more important than others, it could lead to another matrix and other parts to emphasize. The aspect of importance was not considered in this research.

Finally, the results could have been discussed with experts who could have been asked to draw own conclusions in order to find possible hidden conclusions that the author might not have found.

9 Conclusions

Information and their efficient use is a big competitive factor. Organizational assets do not anymore only consist of physical products to process or sell. The information that organizations obtain are valuable. Therefore, it is really important for organizations to integrate the internal data in a sensible way. The businesses' pace in increasing steadily in terms of innovation cycles which leads to more complexity. The companies need to have short decision cycles based on thorough data. If the organizational landscape changes, e.g. due to mergers, it results in much effort to integrate the different counterparties' data. Synergies may arise from good data management, if different business units benefit from other units' work.

Master data is the subset of organizational data, which is foundational for all processes in an organization. Without master data, it would not be possible to invoice because invoices base on master data. That makes master data a particularly important area of organizational data. Therefore, a decent master data management helps coping with the masses and the exploitation to generate further value. From a corporate perspective, the aim of this research was therefore to have a possibility to assess an organization's master data management maturity to evaluate the status quo and compare with others. The academic objective was to add knowledge to the area of master data management maturity because little research was conducted on this topic, even though the incentives are there.

Deriving from the above mentioned motivation, the questions to be answered were stated as: "How can a company's current state in master data management be measured to identify potential improvement areas?"

The research was based on the “Design Science in Information Systems” approach by Hevner et al. (2004). This approach includes information systems and the environment which is also important for improving the processes where it is utilized. This approach was well suited for the context because it included grounded research and the development of an artifact and valid evaluation.

The literature research took place in interdisciplinary fields. The basic terms of the field of knowledge management were introduced, namely data, information, knowledge and wisdom. The concepts were regarded from an organizational point of view. From there, the research was narrowed down towards master data. The term was defined and distinguished from the general understanding of data. The distinction from general data and the importance in an organizational context were described. Then, the focus was laid on data quality and benchmarks. Challenges for data quality approaches were displayed

The following chapter dealt with master data management. There was a closer look at the impact of insufficient MDM. Then some practical models were investigated and compared.

The forth chapter described the full development process of the MD3M. The model was developed along the guidelines of Hevner et al. (2004) and Becker et al. (2009). The elements like maturity levels and attributes, key topics, focus areas, capabilities, dependencies and influential factors are deployed. Then, an assessment questionnaire was developed.

The following chapter emphasizes on the practical application of the MD3M. For this, the case company is introduced and their background and motivation for a MDM analysis are described. In the next step, the maturity model is applied at the case

company. The questionnaire was answered by internal experts on the field of master data management.

The results are presented in the sixth chapter. The maturity model of the case company and the influential factors are presented. The maturity levels are analyzed for each key topic and maturity level. Furthermore, additional findings about the master data management of the case company are depicted. From the results, criticality groups were developed in which MDM issues can be assorted.

The seventh chapter discusses this research's results. The choice of the maturity levels is critically discussed, as well as the order of the capabilities and the capabilities themselves. The setup appeared to be promising. Then, the concept of three improvement areas and the improvement matrix were discussed. After the general discussion on the usability of the MD3M, the conclusions and advice for the case company were described. Finally, limitations and further research were presented.

During this research, a method was developed to assess organizations' master data management maturity along an objective benchmarking system. This tool can be used from different organizations to compare them with each other to figure optimization potential. The results have turned out to be as expected which proved the applicability and validity of the model. Further research can be done in some areas regarding the application at several other companies in different industries.

This research's goal was to provide an overview about existing master data maturity model and then derive a model which can be used to assess an organization's master data management maturity. This will help the organization to position them and see if they are underperforming and where they could improve. In order to do so, there

was an extensive literature study conducted covering relevant terms and aspects in the field. Then different maturity models were compared with each other. From this, criteria were developed in which a company should achieve maturity when willing to perform efficiently. These criteria were grouped into the five key topics Data Model, Data Quality, Usage & Ownership, Data Protection and Maintenance. These five groups are subdivided into one to four focus areas, in which sub-maturity can be achieved. Additionally, dependencies between single capabilities were identified and influential factors which do not apply to all groups of organizations. The goal was to have a maturity model that is applicable for all sizes of companies and all different industries. Therefore these were necessary, so the model can be applied even though some things apply for one kind of organization, but do not for another one. This maturity can be assessed by answering a questionnaire which was developed in accordance with the matrix. Both were validated with experts.

The model was applied at the case company E.ON Energy Trading and their maturity was analyzed. Following, the results were interpreted to give improvement advice to the case company.

Summarizing, this model has been proven to be a good possibility to analyze an organization’s master data maturity and benchmark it against other organizations. This helps an organization to see whether it has a maturity that is adequate to its company size and whether there are weak spots to work on.

References

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

Anderson, R., & Moore, T. (2007). The Economics of Information Security: A Survey and Open Questions. *Technology*, 1-27.

Becker, J., Knackstedt, R., & Pöppelbuß, J. (2009). Developing Maturity Models for IT Management. *Business & Information Systems Engineering*, *1*(3), 213-222. doi:10.1007/s12599-009-0044-5

Bekkers, W., & Van De Weerd, I. (2010). *SPM Maturity Matrix*. Technical Report No. UU-CS-2010-013. Sciences-New York.

Beniger, J. R. (1986). *The Control Revolution: Technological and Economic Origins of the Information Society* (1st ed., p. 493). Cambridge, MA: Harvard University Press.

Bernard, R. (2007). Information Lifecycle Security Risk Assessment: A tool for closing security gaps. *Computers & Security*, *26*(1), 26-30. doi:10.1016/j.cose.2006.12.005

Borghoff, U. M., & Pareschi, R. (1997). Information Technology for Knowledge Management. *Journal of Universal Computer Science*, *3*(8), 835-842.

Borgman, C. L. (2000). The Premise and the Promise of a Global Information Infrastructure. *From Gutenberg to the Global Information Infrastructure: Access to Information in the Networked World* (1st ed., pp. 1-34). Boston, Massachusetts: MIT Press.

Bowker, G. C., Baker, K., Millerand, F., & Ribes, D. (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In J. Hunsinger, L. Klastrup, & M. Allen (Eds.), *International Handbook of Internet Research* (1st ed., pp. 97-117). Dordrecht: Springer Science + Business Media.

Butler, D. (2011). Oracle MDM Maturity Model. Retrieved December 12, 2011, from http://blogs.oracle.com/mdm/entry/oracle_mdm_maturity_model

Casassa Mont, M., & Beato, F. (2007). On Parametric Obligation Policies: Enabling Privacy-aware Information Lifecycle Management in Enterprises. *Eighth IEEE International Workshop on Policies for Distributed Systems and Networks POLICY07 (2007)* (pp. 51-55).

Chisholm, M. (2008). What is Master Data? Retrieved August 18. 2012 from [1] http://www.b-eye-network.com/view/6758

Davenport, T. H., & Prusak, L. (2000). *Working Knowledge - How Organizations Manage What They Know*. Boston, Massachusetts: Harvard Business School Press.

Dizard, W. P. (1985). *The coming information age: An overview of technology, economics, and politics* (p. 224). New York: Longman Higher Education.

E-ON Energy Trading. (2011a). E.ON Markets & Products. Retrieved a from http://www.eon-energy-trading.com/cms/en/1299.jsp

E-ON Energy Trading. (2011b). E.ON Energy Trading facts & figures. Retrieved b from http://www.eon-energy-trading.com/cms/en/eet_facts.jsp

EDUCAUSE Center For Applied Research. (2009). Data Stewardship, Security, and Policies. *ECAR Research Study 8* (pp. 75-91).

Energy Risk. (2010). *Energy Risk Glossary 2010*. London: Incisive Financial Publishing.

European Commission. (2003). COMMISSION RECOMMENDATION of 6 May 2003 concerning the definition of micro, small and medium-sized enterprises. *Official Jouirnal of the European Union*, *361*, 36-41.

Gordon, L. A., & Loeb, M. P. (2002). The economics of information security investment. *ACM Transactions on Information and System Security (TISSEC)*, *5*(4), 438–457. ACM. Retrieved from http://dl.acm.org/citation.cfm?id=581274

Hart, D. (2007). Ownership as an Issue in Data and Information Sharing: A Philophically Based View. *Australiasian Journal of Information Systems*, *17*(2), 23-29.

Henderson, J. C., & Venkatraman, N. (1993). Strategic alignment: Leveraging information technology for transforming organizations. *IBM systems journal*, *32*(1), 1829–1841. Wiley Online Library. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/cbdv.200490137/abstract

Hevner, A. R., March, S. T., & Park, J. (2004). Design Science in Information Systems Research. *MIS Quarterly*, *28*(1), 75-105.

Huang, K., Lee, Y., & Wang, R. (1998). Quality Information and Knowledge. *International Encyclopedia of the Social Behavioral Sciences*, *33*(5), 7453-7456. Upper Saddle River, NJ: Prentice Hall.

Hwang, H.-G., Yeh, R., Chen, H.-G., Jiang, J. J., & Klein, G. (2011). IT Investment Strategy and IT Infrastructure Services. *The Review of Business Information Systems*, *6*(2), 55-63.

IBM. (2007). The IBM Data Governance Council Maturity Model : Building a roadmap for effective data governance.

IT Governance Institute. (n.d.). Cobit 4.1 - Executive Summary Framework. *Reproduction*.

Jain, D., & Singh, S. (2002). Customer lifetime Value research in marketing: A review and future directions. *Journal of Interactive Marketing*, *16*(2), 34-46.

Jashapara, A. (2004). *Knowledge Management . An Integral Approach*. Prentice Hall.

Kahn, B. K., Strong, D. M., & Wang, R. Y. (2002). Information quality benchmarks: Product and service performance. *Communications of the ACM*, *45*(4), 184-193.

Kenessey, Z. (1987). The primary, secondary, tertiary and quaternary sectors o f the economy. *Review of Income and Wealth*.

Klimko, G. (2001). Knowledge Management and Maturity Models: Building Common Understanding. *Second European Conference on Knowledge Management* (pp. 279-288). Bled, Slowenia.

Kumar, S. (2010). MDM Maturity Model. *Information Management Newsletters, March 15, 2010*. Retrieved December 13, 2011, from http://www.information-management.com/newsletters/master_data_management_mdm_maturity_model-10017331-1.html

Loshin, D. (2008). MDM Components and the Maturity Model. *A DataFlux White Paper*.

Malzahn, D. (2008). ISO/IEC 15504 based assessments for Information Quality.

Müller, A., von Thienen, L., & Schröder, H. (2004). IT-Controlling: So messen Sie den Beitrag der Informationstechnologie zum Unternehmenserfolg. *Arbeitspapiere der Nordakademie*, (1), 99-122.

Nonaka, I., & Takeuchi, H. (1995). *The Knowledge Creating Company*. New York: Oxford University Press.

Otto, B., & Hüner, K. M. (2009). Funktionsarchitektur für unternehmensweites Stammdatenmanagement Inhaltsverzeichnis. *Data Management*.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data quality assessment. *Communications of the ACM*, *45*(4), 211-218. doi:10.1145/505248.506010

Redman, T. C. (1995). Improve Data Quality for Competitive Advantage. *Sloan Management Review*, *36*(2), 99-107.

Redman, T. C. (1998). Poor Data Quality on the Typical Enterprise. *Communications of the ACM*, *41*(2), 79-82.

Rehäuser, J., & Krcmar, H. (1996). Wissensmanagement im Unternehmen.

Rooney, P. (2003). Microsoft's CEO: 80-20 Rule Applies to Bugs, Not Just Features. Retrieved February 6, 2012, from http://www.crn.com/news/security/18821726/microsofts-ceo-80-20-rule-applies-to-bugs-not-just-features.htm;jsessionid=cg1Fs4+FXiPCiNdKFNyB2A**.ecappj01

Sacu, C., & Spruit, M. (2011). DWCMM: The Data Warehouse Capability Maturity. *Architecture*.

Sapient Global Markets. (2011). *Glossary of Financial Terms - A Sapient Global Markets quick reference guide to OTC derivative and commodity market trading terminology*.

Schneider, G., Geiger, I. K., & Scheuring, J. (2008). *Prozess- und Qualitätsmanagement: Grundlagen der Prozessgestaltung und Qualitätsverbesserung mit zahlreichen Beispielen, Repetitionsfragen und Antworten* (1st ed.). Zürich: Compendio Bildungsmedien AG.

Shaw, R. S., Chen, C. C., Harris, A. L., & Huang, H.-jou. (2009). The impact of information richness on information security awareness training effectiveness. *Computers & Education*, *52*, 92-100.

Simon, H. A. (1996). *The Sciences of the Artificial*. *MIT Press* (3rd ed.). Cambridge, MA.

Steenbergen, M. V., Berg, M. V. D., & Brinkkemper, S. (2007). An Instrument for the Development of the Enterprise Architecture Practice. *ICEIS*.

Steenbergen, M. V., Bos, R., Brinkkemper, S., & Van De Weerd, I. (n.d.). The Design of Focus Area Maturity Models.

Turban, E., Volonino, L., & Wetherbe, J. (2009). *Information Technology For Management: Transforming Organizations in the Digital Economy*.

Van De Weerd, I., & Brinkkemper, S. (2008). Meta-Modeling for Situational Analysis and Design Methods. In M. R. Syed & S. N. Syed (Eds.), *Handbook of Research on Modern Systems Analysis and Design Technologies and Applications* (pp. 33-54). IGI Global. doi:10.4018/978-1-59904-887-1

Van der Linden, V. (2009). The Data interdependency model: A research on the effects of data quality on organizations. Utrecht.

Wang, R. Y. (1998). A Product Perspective on Total Data Quality Management. *Communications of the ACM*, *41*(2), 58-65.

Wegener, H. (2008). Metadaten, Referenzdaten, Stammdaten. In B. Dinter & R. Winter (Eds.), *Integrierte Informationslogistik* (pp. 189-210). Berlin: Springer.

Winter, R., & Meyer, M. (2001). Organization of data warehousing in large service companies - A matrix approach based on data ownership and competence centers. *Proceedings on American Conference on Information Systems (AMCIS) 2001* (Vol. 6, pp. 1-7). Boston, Massachusetts.

Winter, R., Schmaltz, M., Dinter, B., & Bucher, T. (2008). Das St. Galler Konzept der Informationslogistik. In B. Dinter & R. Winter (Eds.), *Integrierte Informationslogistik* (pp. 1-16). Berlin: Springer.

Wolter, R. & Haselden, K. (2006). The What, Why and How of Master Data Management. Retrieved August 18 2012 from http://msdn.microsoft.com/en-us/library/bb190163.aspx

Yin, R. (2003). *Case Study Research - Design and Methods*. SAGE Publications (3rd ed). Thousand Oaks.

Youssef, L., Butrico, M., & Da Silva, D. (2008). Toward a Unified Ontology of Cloud Computing. *Grid Computing Environments Workshop (GCE)* (pp. 1-10). Austin, Texas.

Appendix A: Questionnaire Business Impacts of MDM

Please indicate if poor Master Data Management leads to the mentioned business impacts by ticking the appropriate checkbox. It will only take you three minutes, but would help me a lot to make my research more empiric. Thank you very much for your help!

1) Lost Sales Opportunities

Possibilities of acquiring new customers or making existing customers buy new products are not seized because of poor data management.

- ☐ Yes
- ☐ No
- ☐ Don’t know

2) Customer Service Costs

Costs that are spent for correcting wrong data; both actually changing the data set and finding the correct values.

- ☐ Yes
- ☐ No
- ☐ Don’t know

3) Customer Dissatisfaction

Customers are dissatisfied because of errors in data quality concerning them. As a result they terminate relationships and possibly advise others to do the same.

- ☐ Yes
- ☐ No
- ☐ Don’t know

4) Lost Revenue

Costs for wrong invoices or inability to bill customers due to errors in the data.

- ☐ Yes
- ☐ No
- ☐ Don’t know

5) Operational Inefficiencies

Inefficiencies in processes because of poor data can have many impacts including insufficient resource planning or the inability to react to changes in time, increased workload etc.

- ☐ Yes
- ☐ No
- ☐ Don’t know

6) Delays in System/Project Deployment

Projects or new system introductions get delayed or cancelled because of bad data quality.

- ☐ Yes
- ☐ No
- ☐ Don’t know

6) Regulatory Compliance

This implies costs that result from a company's inability of adhering to regulatory compliances. It can result in liability risks or even the whole company's life.

- ☐ Yes
- ☐ No
- ☐ Don't know

7) Poor Decision Making

Making wrong decisions because of wrong or outdated data serving as a source for business intelligence and decision support systems.

- ☐ Yes
- ☐ No
- ☐ Don't know

8) Lost Business Opportunities

The missed chance of buying resources at a cheaper price because of an insufficient overview about the market and the prices.

- ☐ Yes
- ☐ No
- ☐ Don't know

9) Employee Morale

If bad data quality leads to frustrated employees either because they cannot work as productive as they would like to or because they have to correct faulty data instead of working on their primary task.

- ☐ Yes
- ☐ No
- ☐ Don’t know

10) System Credibility

Low trust in general data will bring departments to setting up own data sources, which leads to multiple data sources with different values.

- ☐ Yes
- ☐ No
- ☐ Don’t know

Appendix B: Result of the Questionnaire per Respondent

This table shows the results of the questionnaire in Appendix A, precisely depicted for every respondent.

Table 32 *Questionnaire results*

	R1	R 2	R 3	R 4	R 5	R 6	R 7
Lost Sales Opportunities	Yes	No	Yes	yes	yes	yes	yes
Customer Service Costs	Yes	Yes	Yes	DK	yes	yes	yes
Customer Dissatisfaction	Yes	No	DK	yes	yes	yes	DK
Lost Revenue	Yes	Yes	Yes	yes	yes	yes	DK
Operational Inefficiencies	Yes	Yes	Yes	yes	yes	yes	yes
Delays in System/ Project Deployment	Yes	Yes	Yes	yes	yes	yes	DK
Regulatory Compliances	Yes	No	Yes	DK	yes	yes	no
Poor Decision Making	Yes	Yes	Yes	yes	yes	yes	no
Lost Business	Yes	No	No	yes	yes	yes	DK

Opportunities							
Employee Morale	Yes	Yes	Yes	yes	yes	yes	DK
System Credibility	yes	yes	yes	yes	yes	yes	no

Appendix C: Semi-structured interview for defining the matrix

This interview guideline is organized as semi-structured. In the first part, the model is being presented to ensure a thorough understanding of the matrix. Then, firstly some general questions about the model are asked and after that, there are open questions aiming at finding downsides of the model and improvement possibilities.

Part 1: Presentation of the structure (estimated time: 10 minutes)

Explain the concept of maturity matrix. What is the maturity matrix, what is it supposed to be measured? How is the logical structure of the matrix, what does capability mean etc. This explanation includes a document which shortly explains the concepts of

- maturity matrix
- key topics
- focus areas
- capabilities
- dependencies
- Influential factors.

Part 2: Presentation of the MD3M (estimated time: 30 *minutes)*

In this part, the model is being presented and explained. The interviewee is given as much time as he needs to understand the model and read through it. The abbreviations in the questionnaire have the following meaning:

TD – totally disagree

D – disagree

U – undecided

A – agree

TA – totally agree

Part 3A: Questionnaire (estimated time: 5 minutes)

1) I understand the model and its purpose.

TD D U A TA

2) The model is suited to assess a company's maturity and to benchmark against other companies.

TD D U A TA

3) The model can help organizations find problems in their way of handling master data.

TD D U A TA

4) The model provides an objective way to compare companies.

TD D U A TA

Part 3B: Open Interview (estimated time: 30 minutes)

1) Would you improve the descriptions to make them more understandable? If yes, which ones?
2) Do you agree with the key topics structuring? If no, what would you change and why?
3) Do you agree with the focus areas structuring? If no, what would you change and why?
4) Do you agree with the capabilities structuring? If no, what would you change and why?
5) Do you think there are aspects missing that need to be included?
6) Do you think some considered aspects are irrelevant?
7) Do you agree with the dependencies? If no, what would you change and why?
8) Do you agree with the influential factors? If no, what would you change and why?

Appendix D: The capabilities description

DATA MODEL

This key topic deals with the data and the infrastructural and organizational view on it. It contains topics like what data is considered as master data, how the data is structured, which systems use what data, and where the data is stored.

Definition of Master Data

The goal of this focus area is to achieve a companywide definition of master data to reach consensus among all stakeholders.

Initial

Goal:	A basic understanding of master data exists within some units or within individuals.
Action:	Preliminary discussions about MDM and attempts to formulate the individuals' understanding of master data have to take place.
Prerequisites:	-
Note(s):	-

B: Repeatable

Goal:	First cooperative definitions have been made between single units. Discussions are held about the topic.
Action:	Cross-functional discussions about MDM are held. During these meetings, discussions about a common definition take place.
Prerequisites:	-
Note(s):	-

C: Defined Process

Goal:	The definition bases on more information from different departments and is a cooperative result. Fewer units have their individual understanding, but thriving towards a shared definition.
Action:	Support and encourage discussions about master data between different units. Encourage establishment of a project team on this topic to bundle all activities.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	There is an official definition of master data for the organization with regard to the special circumstances of the organization. This definition is known by all parties involved and can easily be found on a centrally accessible space.
Action:	Define the term master data with a team of responsible persons from different key topics including business and IT.

Prerequisites:	-
Note(s):	-

Optimized

Goal:	There are interfaces for data of different organizations that need to exchange data on a regular basis. Standard formats are established.
Action:	Define data exchange standards for the exchange between the companies
Prerequisites:	-
Note(s):	This capability is only applicable for organizations belonging to a group which have to interact with others and exchange company data.

Master Data Model

This focus area contains the construction of a master data model. This serves as an overview for the organization about master data objects.

Initial

Goal:	There are initial attempts to design a model. Probably, there are already some models focusing on data for a particular topic.
Action:	Raise awareness for other units' view on master data. The fact that one department only needs certain parts does not mean that others only need the same.
Prerequisites:	Definition of Master Data B
Note(s):	-

B: Repeatable

Goal:	The different departments can give an overview about master data and how it is interrelated relevant in their scope. There is no knowledge about the data model for the other departments.
Action:	Bring the different departments together to exchange the models from the different units.
Prerequisites:	-
Note(s):	-

C: Defined Process

Goal:	The different departments can give an overview about master data and how it is interrelated relevant in their scope. Some knowledge already exists about master data objects in other key topics.
Action:	Bundling of knowledgeable people in the project team to share that knowledge.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	An enterprise wide master data model was constructed and agreed upon throughout the different units which are concerned with master data.
Action:	Develop a plan to maintain the model and adapt it to upcoming changes.
Prerequisites:	-
Note(s):	-

Optimized

Goal:	The enterprise wide master data model is maintained regularly. A clear plan with the intervals and the responsibilities concerning the maintenance exists and is communicated throughout the relevant roles.
Action:	Keep the plan updated and integrate the maintenance into role descriptions of responsible persons.
Prerequisites:	-
Note(s):	-

Data Landscape

This field contains the data and its accessing systems. The organization should gain an overview about how these interact and document the data landscape.

Initial

Goal:	There is an overview about systems that use or access master data.
Action:	Try to make the overview complete and elaborate on it.
Prerequisites:	Definition of Master Data C, Master Data Model A
Note(s):	-

B: Repeatable

Goal:	There is a full overview about which systems have reading or writing access to data.
Action:	Investigate the overview for redundancies in the data.
Prerequisites:	-
Note(s):	-

C: Defined Process

Goal:	It is pointed out if data is stored and accessed redundantly.
Action:	Find sources and origins of redundancies.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	There is a consistent inventory of all data sources and by which systems they are used. Redundancies are pointed out and concepts are developed to resolve them.

Action:	Develop concepts to resolve redundancies and implement them. Which systems can be substituted and which should be kept?
Prerequisites:	-
Note(s):	-

Optimized

Goal:	There is a consistent inventory of all data sources and by which systems they are used. Redundancies are solved. The data logic is scalable. Superfluous systems are substituted.
Action:	Check the data logic regularly if it still matches the organizational structure. Check for redundancies regularly.
Prerequisites:	-
Note(s):	-

DATA QUALITY

The key topic **DATA QUALITY** *is dedicated at data quality in all regards. This includes ways to assess the data quality, finding ways to improve data quality and investigate the reasons for and impact of quality issues. Furthermore, the organization can assess what the most frequent and critical sources for problems are.*

Assessment of Data Quality

The organization must make clear how it defines data quality and how they want to assess it.

Initial

Goal:	There is a feeling about data being of good or bad quality
Action:	It can be estimated if the quality of data in a particular field is rather good or bad.
Prerequisites:	-
Note(s):	-

B: Repeatable

Goal:	It is clearly stated which aspects are part of data quality and need to be measured in terms of assessing data quality
Action:	Formalizing criteria for defining data quality objectively, maybe based on best practices
Prerequisites:	-
Note(s):	-

C: Defined Process

Goal:	Data quality is defined regarding the requirements of different stakeholders.

Action:	Different departments with different needs must communicate these. The specific data must always suffice the need of quality of the users.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	Data quality is measured objectively and for each piece of master data it is known which quality it has.
Action:	Executing quality measurement.
Prerequisites:	-
Note(s):	-

Optimized

Goal:	The data quality assessment is conducted regularly for every group of data.
Action:	Setting up a frequency for regularly or consistently measuring data quality
Prerequisites:	-
Note(s):	-

Impact on Business

This section describes the impact of poor master data quality on the particular business. The organization must find ways to monetarize the impacts of insufficient master data.

Initial

Goal:	The organization knows that quality issues in certain data will impact the business from a reputational point of view.
Action:	Analyzing data and impact of it on business from a reputational point of view (image of bad quality, unfriendliness, incompetency, etc).
Prerequisites:	Assessment of data quality A
Note(s):	-

B: Repeatable

Goal:	The organization knows that quality issues in certain data will impact specific parts of the business as direct monetary loss
Action:	Analyzing data and impact of it on business from a monetary point of view (lost sales opportunities, lacking up- or cross-selling, etc).
Prerequisites:	-
Note(s):	-

C: Defined Process

Goal:	The organization knows how bad master data impacts the business from a monetary perspective.

Action:	The organization can precisely estimate the monetary loss for different cases of data issues.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	The organization knows how bad master data impacts the business from a non-monetary perspective, i.e. loss in reputation, lacking customer retention etc.
Action:	The organization can realistically forecast the impact of reputational damages due to data issues.
Prerequisites:	-
Note(s):	This does not apply for non-profit, military or governmental organizations.

Optimized

Goal:	The organization can state how insufficient master data influences the business in monetary and non-monetary terms and can classify this in financial arguments.
Action:	The organization has a realistic overview of the impact of poor data and is able to state consequences for different scenarios.
Prerequisites:	-
Note(s):	This does not apply for non-profit, military or governmental organizations.

Awareness of Quality Gaps

The organization needs to figure out the areas in which there are quality gaps in order to improve these. This helps gaining a realistic impression about the standard of master data.

Initial

Goal:	The competence team is aware of the fact that there are different reasons for poor data quality.
Action:	The organization has assessed origins of poor data which apply for their business and knows about them.
Prerequisites:	-
Note(s):	-

B: Repeatable

Goal:	The organization can state which reasons for poor data quality occur in the organization.
Action:	The organization has assessed the reasons for poor data and knows where the threats are.
Prerequisites:	-
Note(s):	-

C: Defined Process

Goal:	There are patterns investigated about poor data quality.
Action:	The organization is able to state which reasons are responsible for which problems with data
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	The employees are aware of the reasons and sources of bad master data quality in their daily work and the consequences thereof.
Action:	Employees are made aware of consequences of processing poor quality data.
Prerequisites:	-
Note(s):	-

Optimized

Goal:	The organization is aware of different reasons for poor data and where they are existent in house. The company knows where the weak spots are and what the reason for that weakness is
Action:	The organization is fully aware of reasons for poor data and can trace them back.
Prerequisites:	-
Note(s):	Examples could be misspellings, but also wrong information being entered due to outdated or wrong information

Improvement

Here, the organization spots potential improvement areas in which the data quality can be ameliorated.

Initial

Goal:	The organization figures out areas in which the data quality is not sufficient
Action:	The organization has an overview about the degree of insufficiency of data quality in different areas and can state which ones are below expectations
Prerequisites:	Assessment of Data Quality D
Note(s):	-

B: Repeatable

Goal:	There is awareness of the importance of high quality data in terms of efficiency and effectiveness.
Action:	The employees know about the improvements in efficiency if the data was better
Prerequisites:	
Note(s):	

C: Defined Process

Goal:	The organization has a benchmarking system in place to assess whether the data quality is sufficient or not.
Action:	Implement and communicate a company-wide guideline to measure data quality.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	Improvement measures are installed to improve the data quality.
Action:	The guidelines must be implemented.
Prerequisites:	-
Note(s):	e.g. as little manual input as possible, drop down lists, logical checks (do zip codes match the city etc)

Optimized

Goal:	The organization regularly assesses the data quality along the benchmarking system and ensures that the data quality stays within the defined quality.
Action:	It must be a process in place which clarifies intervals in which data quality is being assessed.
Prerequisites:	-
Note(s):	-

USAGE & OWNERSHIP

This key topic is dedicated at figuring out who uses the data in which systems. Which employee has read/write access and is it clear why people are granted or denied access to certain data? The organization can find out if there are data ownership concepts implemented and see whether the historically grown way still displays the needs.

Data Usage

In this focus area, the emphasis lays on how and by whom data is used.

Initial

Goal:	The organization knows for the area of master data who is using which data.
Action:	An overview or rules exist, who has access to the data.
Prerequisites:	Definition of Master Data C, Data Ownership A, Data Access C, Data Protection B
Note(s):	-

B: Repeatable

Goal:	It is known if every employee uses the data he has. The employee knows where to get the needed data.
Action:	This can be observed with access logging and asking the employees. Furthermore, the employees can be asked whether they are aware of available data sources.
Prerequisites:	-
Note(s):	-

C: Defined Process

Goal:	Every source of data that an employee might need it communicated to him and he is given access to.
Action:	Employees are explicitly made aware of the data sources that he has access to and which may be of any use.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	Data repositories are maintained regularly and do not get outdated, ergo unusable.
Action:	There are working instructions in place to take care of outdated data repositories on a regular basis.
Prerequisites:	-
Note(s):	-

Optimized

Goal:	The employees use the possibilities they have and are not reluctant to use certain systems to obtain data from.
Action:	Offer the employees trainings for and communicate the benefits of the systems.
Prerequisites:	-
Note(s):	-

Data Ownership

There must be responsibilities and ownerships distributed for the different parts of master data. These owners must feels responsible and have certain tasks regarding the data.

A: Initial

Goal:	Data elements are owned by individuals/departments.
Action:	For all data sets, there are ownerships defined.
Prerequisites:	-
Note(s):	-

Repeatable

Goal:	Data elements are owned by logically consistent roles/departments. The owner defines usage, purpose and content of data
Action:	The ownerships for data are executed by logical roles who have knowledge about the topic.
Prerequisites:	-
Note(s):	-

Defined Process

Goal:	Responsible persons for data are openly communicated and known throughout the company. The data owner has defined responsibilities for treatment of the data.
Action:	The employees know who the responsible person is for the data and what his responsibilities are. This information is accessible for all stakeholders.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	Data stewards are established for chunks of data.
Action:	The concept of data stewardship is promoted within the organization and data stewards are assigned.
Prerequisites:	-
Note(s):	-

E: Optimized

Goal:	Data stewardship is promoted and fixed in the role description of the job. Data quality standards are defined and adhered to.
Action:	Data owner defined quality standards that are adhered to.
Prerequisites:	-
Note(s):	-

Data Access

This focus area consists of how is access to data given, who is given access and the reasons of granting or denying access to data. Furthermore it is about the employees' knowledge about their access to the data.

A: Initial

Goal:	There is a defined process how to get access to data.
Action:	A standardized procedure is available as a process description.
Prerequisites:	-
Note(s):	-

Repeatable

Goal:	Access to data is denied to unauthorized personnel.

Action:	The decision is taken upon clear documented reasons and guidelines.
Prerequisites:	-
Note(s):	-

C: Defined Process

Goal:	Every employee has access to the data he needs to fulfill his job.
Action:	Access is automatically granted for the needed systems and data sources.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	Every employee has access to the data he needs to fulfill his work and only this data. He does not have access to data that he either does not need or should not be seeing.
Action:	Employees are not overwhelmed with too much data access and too many sources from which they have to search their goal information.
Prerequisites:	-
Note(s):	-

E: Optimized

Goal:	Every employee knows which sources he has access to and what he can find there for his purposes.
Action:	Employees have a good overview of their access to repositories and know exactly where to find what.
Prerequisites:	-
Note(s):	-

DATA PROTECTION

This section is about the technical security of data; whether and how it is secured against possible incidents.

Data Protection

A: Initial

Goal:	The technical requirements for data protection are fulfilled
Action:	No access possibility from outside, default protection like RAID systems etc
Prerequisites:	Storage A
Note(s):	-

Repeatable

Goal:	Access to data must be activated on request.
Action:	To any kind of restricted data, access must be permitted by a responsible authority.

Prerequisites: Data Access B
Note(s): -

C: Defined Process

Goal: There are rules for which roles data access can be granted.
Action: Access will be granted only if predefined conditions apply.
Prerequisites: -
Note(s): -

D: Managed and Measurable

Goal: Passwords exist for systems with data access which have to adhere to common security standards and have to be changed regularly.
Action: The systems must require secure passwords and regular changes.
Prerequisites: -
Note(s): -

E: Optimized

Goal: Awareness for data protection must be raised among the employees.
Action: Make the impact clear; e.g. convince people not to leave their computers unlocked.
Prerequisites: -
Note(s): -

STORAGE

Here, the focus is on physical and logical storage and the treatment of master data in the data lifecycle. The main points to investigate are how the data is stored and how the data is treated during the lifecycle.

Storage

This section is about storing data in an efficient, persistent way.

A: Initial

Goal: The data is stored in a persistent, performing way
Action: The hardware should be not too outdated to ensure appropriate access times and no bottlenecks.
Prerequisites: -
Note(s): -

Repeatable

Goal: The data logic is regularly checked for up-to-datedness.
Action: The data logic should be checked against reality if the logic still depicts the real world situation.
Prerequisites: -
Note(s): -

C: Defined Process

Goal:	Automatic tools regularly check for redundancies and duplicates.
Action:	Tools for redundancy checks should be implemented to improve database performance and decrease manual effort.
Prerequisites:	-
Note(s):	-

D: Managed and Measurable

Goal:	The data base logic is regularly checked for persistence, performance and efficiency.
Action:	The data base logic should regularly be checked in terms of efficiency and performance. Is it close to being up to date or are there other solutions available which would increase performance?
Prerequisites:	-
Note(s):	-

E: Optimized

Goal:	The data is stored in an innovative way with possibilities of forecasting and analysis.
Action:	Upon the modern systems, there should be BI solutions implemented which can do analysis and forecasting.
Prerequisites:	-
Note(s):	-

Data Lifecycle

This focus area is concerned with master data being an asset that changes over time and has a lifecycle which needs to be considered.

A: Initial

Goal:	The organization is aware of the fact that data has a lifecycle and that data structure will change over time.
Action:	All stakeholders should be made aware that master data is not a static thing that stays unchanged forever. Even rather static data like master data will undergo changes over time.
Prerequisites:	-
Note(s):	-

Repeatable

Goal:	Data is considered as an organizational asset.
Action:	Data must be regarded as assets like plants and human capital.
Prerequisites:	-
Note(s):	-

C: Defined Process

Goal:	Guidelines must be established for treating data over the lifecycle.

Action:	Since master data will change eventually, the data logic must be set up scalable and adaptable.
Prerequisites:	Data Ownership C
Note(s):	-

D: Managed and Measurable

Goal:	For every data item, a single source of truth is established.
Action:	For the data, one single source of truth is needed where it is originated to avoid redundancies.
Prerequisites:	-
Note(s):	-

E: Optimized

Goal:	The entering, updating and deleting of data is automatically logged by the systems to decrease documentation effort and facilitate auditing.
Action:	Automatic maintenance of data decreases documentation effort and increases employee efficiency.
Prerequisites:	-
Note(s):	-

Appendix E: Overview of Maturity Levels

MD3M – Master Data Management Maturity Matrix				
Data Model				
Initial	Repeatable	Defined process	Managed and measurable	Optimized
A basic understanding of master data exists within some units or within individuals.	First cooperative definitions have been made between single units. Discussions are held about the topic.	The definition bases on more information from different departments and is a cooperative result. Fewer units have their individual understanding, but thriving towards a shared definition.	There is an official definition of master data for the organization with regard to the special circumstances of the organization. This definition is known by all parties involved and can easily be found on a centrally accessible space.	There are interfaces for data of different organizations that need to exchange data on a regular basis. Standard formats are established.
There are initial attempts to design a model. Probably, there are already some models focusing on data for a particular topic.	The different departments can give an overview about master data and how it is interrelated relevant in their scope. There is no knowledge about the data model for the other departments.	The different departments can give an overview about master data and how it is interrelated relevant in their scope. Some knowledge already exists about master data objects in other key topics.	An enterprise wide master data model was constructed and agreed upon throughout the different units which are concerned with master data.	The enterprise wide master data model is maintained regularly. A clear plan with the intervals and the responsibilities concerning the maintenance exists and is communicated throughout the relevant roles.
There is an overview about systems that use or access master data.	There is a full overview about which systems have reading or writing access to data.	It is pointed out if data is stored and accessed redundantly.	There is a consistent inventory of all data sources and by which systems they are used. Redundancies	There is a consistent inventory of all data sources and by which systems they are used. Redundancies

			are pointed out and concepts are developed to resolve them.	are solved. The data logic is scalable. Superfluous systems are substituted.
Data Quality				
Initial	Repeatable	Defined process	Managed and measurable	Optimized
There is a feeling about data being of good or bad quality	It is clearly stated which aspects are part of data quality and need to be measured in terms of assessing data quality	Data quality is defined regarding the requirements of different stakeholders.	Data quality is measured objectively and for each piece of master data it is known which quality it has.	The data quality assessment is conducted regularly for every group of data.
The organization knows that quality issues in certain data will impact the business from a reputational point of view.	The organization knows that quality issues in certain data will impact specific parts of the business as direct monetary loss	The organization knows how bad master data impacts the business from a monetary perspective.	The organization knows how bad master data impacts the business from a non-monetary perspective, i.e. loss in reputation, lacking customer retention etc.	The organization can state how insufficient master data influences the business in monetary and non-monetary terms and can classify this in financial arguments.
The competence team is aware of the fact that there are different reasons for poor data quality.	The organization can state which reasons for poor data quality occur in the organization.	There are patterns investigated about poor data quality.	The employees are aware of the reasons and sources of bad master data quality in their daily work and the consequences thereof.	The organization is aware of different reasons for poor data and where they are existent in house. The company knows where the weak spots are and what the reason for that weakness is
The organization figures out areas in which the data quality is not sufficient	There is awareness of the importance of high quality data in terms of	The organization has a benchmarking system in place to assess	Improvement measures are installed to improve the data quality.	The organization regularly assesses the data quality along the benchmarking

	efficiency and effectiveness.	whether the data quality is sufficient or not.		system and ensures that the data quality stays within the defined quality.
Usage & Ownership				
Initial	Repeatable	Defined process	Managed and measurable	Optimized
The organization knows for the area of master data who is using which data.	It is known if every employee uses the data he has. The employee knows where to get the needed data.	Every source of data that an employee might need it communicated to him and he is given access to.	Data repositories are maintained regularly and do not get outdated, ergo unusable.	The employees use the possibilities they have and are not reluctant to use certain systems to obtain data from.
Data elements are owned by individuals/ departments.	Data elements are owned by logically consistent roles/departments. The owner defines usage, purpose and content of data	Responsible persons for data are openly communicated and known throughout the company. The data owner has defined responsibilities for treatment of the data.	Data stewards are established for chunks of data.	Data stewardship is promoted and fixed in the role description of the job. Data quality standards are defined and adhered to.
There is a defined process how to get access to data.	Access to data is denied to unauthorized personnel.	Every employee has access to the data he needs to fulfill his job.	Every employee has access to the data he needs to fulfill his work and only this data. He does not have access to data that he either does not need or should not be seeing.	Every employee knows which sources he has access to and what he can find there for his purposes.
Data Protection				
Initial	Repeatable	Defined process	Managed and measurable	Optimized
The technical requirements for data protection are fulfilled	Access to data must be activated on request.	There are rules for which roles data access can be granted.	Passwords exist for systems with data access which have to	Awareness for data protection must be raised among the

			adhere to common security standards and have to be changed regularly.	employees.
Maintenance				
Initial	Repeatable	Defined process	Managed and measurable	Optimized
The data is stored in a persistent, performing way	The data logic is regularly checked for up-to-datedness.	Automatic tools regularly check for redundancies and duplicates.	The data base logic is regularly checked for persistence, performance and efficiency.	The data is stored in an innovative way with possibilities of forecasting and analysis.
The organization is aware of the fact that data has a lifecycle and that data structure will change over time.	Data is considered as an organizational asset.	Guidelines must be established for treating data over the lifecycle.	For every data item, a single source of truth is established.	The entering, updating and deleting of data is automatically logged by the systems to decrease documentation effort and facilitate auditing.

Appendix F: MD3M Assessment Questionnaire

With these questions, the influential factors are assessed.

Influential Factors	
Does your company belong to a group and your company needs to interact regularly with other internal members of the group and exchange data?	
Is your company a non-profit organizations, and/or a governmental or military organization?	
Does your company exceed a number of employees of approximately 250?	
Do the employees need to work with many different systems for executing their daily work and have to follow different processes when doing this?	
?	

This questionnaire contains one question for each capability.

Nr.	Capability	Statement	Answer
		Definition of Master Data	
1	A	Is there a basic understanding in your department or in other departments about the definition of Master Data?	
2	B	Have there been discussions between functional units about master data with the target of getting a common understanding?	
3	C	There is a shared definition of some departments in the organization of master data.	
4	D	There is one official definition of Master data in the organization which is communicated to everyone and the employees all know where to find it.	
5	E	There are standard interfaces for exchanging data between companies belonging to the same group	
		Master Data Model	
6	A	There are some initial - possibly incomplete and isolated - attempts to get an overview about the master data.	
7	B	Some departments who are highly related to master data can present a master data model of data relevant to their daily work. It covers their daily work but does not regard the other units.	
8	C	There are some models from different departments. There is already some knowledge about master data objects in other key topics and how the data relates.	
9	D	There is an enterprise-wide master data model which was developed from all relevant departments.	
10	E	The master data model is maintained regularly and the responsibilities for the maintenance are clear.	

Data Landscape			
11	A	An overview exists with information on systems that use or access master data	
12	B	The overview is complete.	
13	C	It has been investigated if there are redundancies in the storing and accessing of data.	
14	D	There is an overview of all data sources and systems and their interaction. Redundancies can be mapped to systems and sources.	
15	E	The overview gets maintained on a regular basis and redundancies are resolved if possible. Superfluous systems are substituted.	
Assessment of Data Quality			
16	A	There is a common estimation about the quality of the master data within the organization.	
17	B	The organization has formalized quality criteria that are important and need to be measured.	
18	C	There are quality requirements defined taking into account the requirements of different business units.	
19	D	A quality assessment has taken place in the organization and it is known which quality the data has.	
20	E	There are defined intervals in which quality assessment is conducted and changes in quality are monitored.	
Impact on Business			
21	A	The organization is aware of the reputational impact on the business if data quality is insufficient.	
22	B	The organization is aware of the monetary impact on the business if data quality is insufficient.	
23	C	The organization knows how much money gets lost due to insufficient data quality (e.g. Lost sales opportunities).	
24	D	The organization knows how insufficient data impacts the firm from a non-monetary perspective (e.g. Reputational, customer-retention).	
25	E	The organization can classify the impact of bad master data quality from both monetary and reputational aspects into financial arguments and can state how much money is lost.	
Awareness of Quality Gaps			
26	A	The organization knows about different reasons for quality issues in master data.	
27	B	The organization knows which reasons for bad quality are relevant in the organization	
28	C	The organization can precisely state which reasons for bad quality are involved at which source of data entering.	
29	D	The employees are aware of reasons and sources of poor master data quality and the consequences for the business.	

30	E	The organization can precisely state the weak spots in data setup (e.g. Entering information manually - especially foreign words or numbers - results in spelling mistakes).	
Improvement			
31	A	The organization precisely knows in which areas the data quality is not sufficient according to the defined quality requirements.	
32	B	The organization is aware of the increasing efficiency and effectiveness in daily work if the quality adheres to the requirements. This is relevant for both employees setting up data and those using the data.	
33	C	There is a company-wide benchmarking system in place to measure data quality objectively.	
34	D	Improvement measures are in place to increase data quality.	
35	E	There is a constant loop of monitoring and improving quality to ensure it has the required quality.	
Data Usage			
36	A	The organization knows who is using i.e. has access to what data in the organization.	
37	B	The employees know where to get required data. If is assessed if the employees use the provided data sources.	
38	C	For every source of master data it is communicated to the appropriate users that they have access and that the data contains relevant information.	
39	D	Data repositories get regularly maintained and do not get outdated.	
40	E	The employees are aware of the sources they have access to and are not reluctant to use any of them (e.g. because of ignorance of the usage)	
Data Ownership			
41	A	Data elements have an owner who is either an individual or a department.	
42	B	The data elements are logically owned by related roles/departments. The data owner defines purpose, usage and content.	
43	C	The responsible persons for master data are communicated throughout the organization. The persons have documented responsibilities.	
44	D	Data stewardships are established for data areas.	
45	E	Data stewardship is promoted within the organization and fixed in role descriptions of jobs.	
Data Access			
46	A	There is a protocol to obtain access to data for the employees.	
47	B	Unauthorized personnel are not given access to sensitive data.	
48	C	Every employee is given access to necessary data beforehand. He is automatically equipped with the main sources.	

49	D	The employees have efficient data sources and have access to the data they need and not much more.	
50	E	The employees know their sources and have a good overview about what they can find where.	
Data Protection			
51	A	The data is secured against external or default threats with up to date solutions.	
52	B	Data access is only activated on request which was granted by the responsible authority.	
53	C	There are clear rules communicated for granting access for certain roles.	
54	D	Access to data (especially sensitive data) is restricted with passwords that need to be changed regularly and adhere to common standards.	
55	E	There is awareness about data protection among the employees, e.g. the employees do not leave their computers unlocked when leaving their desks).	
Storage			
56	A	The data is stored in an efficient way. Loading does not take too long.	
57	B	The data logic is displaying the real world situation.	
58	C	Automated tools regularly check for redundancies and duplicates.	
59	D	The data base logic is regularly compared to the real world situation it is meant to depict.	
60	E	The data is stored with innovative solutions to enable data analysis and forecasting (BI solutions).	
Data Lifecycle			
61	A	Data is considered as an object that is undergoing a lifecycle and changes over time.	
62	B	Data is valued as an organizational asset that brings value to the organization.	
63	C	The data logic is scalable to treat data according to its position in the lifecycle.	
64	D	For every data item, a single source of truth is established.	
65	E	Maintenance labor like entering and updating is automatically logged by the systems.	

Appendix G: The Abbreviations of the MD3M

Table 33 *Abbreviations Capabilities*

Abbreviation	Focus Area	Capability
DMD-A	Definition of Master Data	A basic understanding of master data exists within some units or within individuals.
DMD-B	Definition of Master Data	First cooperative definitions have been made between single units. Discussions are held about the topic.
DMD-C	Definition of Master Data	The definition bases on more information from different departments and is a cooperative result. Fewer units have their individual understanding, but thriving towards a shared definition.
DMD-D	Definition of Master Data	There is an official definition of master data for the organization with regard to the special circumstances of the organization
DMD-E	Definition of Master Data	There is a company-wide definition of Master data containing which parts of the data belong to master data and why.
MDM-A	Master Data Model	There are initial attempts to design a model. Probably, there are already some models focusing on data for a particular topic.
MDM-B	Master Data Model	The different departments can give an overview about master data and how it is interrelated relevant in their scope. There is no knowledge about the data model for the other departments.
MDM-C	Master Data Model	The different departments can give an overview about master data and how it is interrelated relevant in their scope.
MDM-D	Master Data Model	An enterprise wide master data model was constructed and agreed upon throughout the different units which are concerned with master data.
MDM-E	Master Data Model	The enterprise wide master data model is maintained regularly. A clear plan with the intervals and the responsibilities concerning the maintenance exists and is communicated throughout the relevant roles.
DLS-A	Data Landscape	There is an overview about systems that use or access master data.
DLS-B	Data Landscape	There is a full overview about which systems have reading or writing access to data.
DLS-C	Data Landscape	It is pointed out if data is stored and accessed

		redundantly.
DLS-D	Data Landscape	There is a consistent inventory of all data sources and by which systems they are used. Redundancies are pointed out and concepts are developed to resolve them.
DLS-E	Data Landscape	There is a consistent inventory of all data sources and by which systems they are used. Redundancies are solved. The data logic is scalable. Superfluous systems are substituted.
ADQ-A	Assessment of Data Quality	There is a feeling about data being of good or bad quality for data items and that good quality data creates added value for the company.
ADQ-B	Assessment of Data Quality	It is clearly stated which aspects are part of data quality and need to be measured in terms of assessing data quality
ADQ-C	Assessment of Data Quality	Data quality is defined regarding the requirements of different stakeholders.
ADQ-D	Assessment of Data Quality	Data quality is measured objectively and for each piece of master data it is known which quality it has
ADQ-E	Assessment of Data Quality	The data quality assessment is conducted regularly for every group of data.
IOB-A	Impact on Business	The organization knows that quality issues in certain data will impact the business from a reputational point of view.
IOB-B	Impact on Business	The organization knows that quality issues in certain data will impact specific parts of the business as direct monetary loss.
IOB-C	Impact on Business	The organization knows how bad master data impacts the business from a monetary perspective.
IOB-D	Impact on Business	The organization knows how bad master data impacts the business from a non-monetary perspective, i.e. loss in reputation, lacking customer retention etc.
IOB-E	Impact on Business	The organization can state how insufficient master data influences the business in monetary and non-monetary terms and can classify this in financial arguments.
AQG-A	Awareness of Quality Gaps	The competence team is aware of the fact that there are different reasons for poor data quality.
AQG-B	Awareness of Quality Gaps	The organization can state which reasons for poor data quality occur in the organization.
AQG-C	Awareness of Quality Gaps	There are patterns investigated about poor data quality.
AQG-D	Awareness of Quality Gaps	The employees are aware of the reasons and sources of bad master data quality in their daily

		work and the consequences thereof.
AQG-E	Awareness of Quality Gaps	The organization is aware of different reasons for poor data and where they are existent inhouse. The company knows where the weak spots are and what the reason for that weakness is
IMP-A	Improvement	The organization figures out areas in which the data quality is not sufficient
IMP-B	Improvement	There is awareness of the importance of high quality data in terms of efficiency and effectiveness.
IMP-C	Improvement	The organization has a benchmarking system in place to assess whether the data quality is sufficient or not.
IMP-D	Improvement	Improvement measures are installed to improve the data quality.
IMP-E	Improvement	The organization regularly assesses the data quality along the benchmarking system and ensures that the data quality stays within the defined quality.
DTU-A	Data Usage	The organization knows for the area of master data that is using which data.
DTU-B	Data Usage	It is known if every employee uses the data he has. The employee knows where to get the needed data.
DTU-C	Data Usage	Every source of data that an employee might need it communicated to him and he is given access to.
DTU-D	Data Usage	Data repositories are maintained regularly and do not get outdated, ergo unusable.
DTU-E	Data Usage	The employees use the possibilities they have and are not reluctant to use certain systems to obtain data from.
DTO-A	Data Ownership	Data elements are owned by individuals/departments.
DTO-B	Data Ownership	Data elements are owned by logically consistent roles/departments. The owner defines usage, purpose and content of data
DTO-C	Data Ownership	Responsible persons for data are openly communicated and known throughout the company. The data owner has defined responsibilities for treatment of the data.
DTO-D	Data Ownership	Data stewards are established for chunks of data.
DTO-E	Data Ownership	Data stewardship is promoted and fixed in the role description of the job. Data quality standards are defined and adhered to.
DTA-A	Data Access	There is a defined process how to get access to data.
DTA-B	Data Access	Access to data is denied to unauthorized personnel.
DTA-C	Data Access	Every employee has access to the data he needs to

		fulfill his job.
DTA-D	Data Access	Every employee has access to the data he needs to fulfill his work and only this data. He does not have access to data that he either does not need or should not be seeing.
DTA-E	Data Access	Every employee knows which sources he has access to and what he can find there for his purposes.
DTP-A	Data Protection	The technical requirements for data protection are fulfilled
DTP-B	Data Protection	Access to data must be activated on request.
DTP-C	Data Protection	There are rules for which roles data access can be granted.
DTP-D	Data Protection	Passwords exist for systems with data access which have to adhere to common security standards and have to be changed regularly.
DTP-E	Data Protection	Awareness for data protection must be raised among the employees.
STO-A	Storage	The data is stored in a persistent, performant way
STO-B	Storage	The data logic is regularly checked for up-to-datedness.
STO-C	Storage	Automatic tools regularly check for redundancies and duplicates.
STO-D	Storage	The data base logic is regularly checked for persistence, performance and efficiency
STO-E	Storage	The data is stored in an innovative way with possibilities of forecasting and analysis
DLC-A	Data Lifecycle	The organization is aware of the fact that data has a lifecycle and that data structure will change over time.
DLC-B	Data Lifecycle	Data is considered as an organizational asset.
DLC-C	Data Lifecycle	Guidelines must be established for treating data over the lifecycle.
DLC-D	Data Lifecycle	For every data item, a single source of truth is established.
DLC-E	Data Lifecycle	The entering, updating and deleting of data is automatically logged by the systems to decrease documentation effort and facilitate auditing.

Appendix H: The Questionnaire Results for EET

Influential Factors	
Does your company belong to a group and your company needs to interact regularly with other internal members of the group and exchange data?	Yes
Is your company a non-profit organization, and/or a governmental or military organization?	No
Does your company exceed a number of employees of approximately 250?	Yes
Do the employees need to work with many different systems for executing their daily work and have to follow different processes when doing this?	Yes

Nr.	Capability	Statement	Answer
Definition of Master Data			
1	A	Is there a basic understanding in your department or in other departments about the definition of Master Data?	Yes
2	B	Have there been discussions between functional units about master data with the target of getting a common understanding?	Yes
3	C	There is a shared definition of some departments in the organization of master data.	Yes
4	D	There is one official definition of Master data in the organization which is communicated to everyone and the employees all know where to find it.	No
5	E	There are standard interfaces for exchanging data between companies belonging to the same group	No
Master Data Model			
6	A	There are some initial - possibly incomplete and isolated - attempts to get an overview about the master data.	Yes
7	B	Some departments who are highly related to master data can present a master data model of data relevant to their daily work. It covers their daily work but does not regard the other units.	Yes
8	C	There are some models from different departments. There is already some knowledge about master data objects in other key topics and how the data relates.	Yes
9	D	There is an enterprise-wide master data model which was developed from all relevant departments.	No
10	E	The master data model is maintained regularly and the responsibilities for the maintenance are clear.	No
Data Landscape			
11	A	An overview exists with information on systems that use or access master data	Yes
12	B	The overview is complete.	No

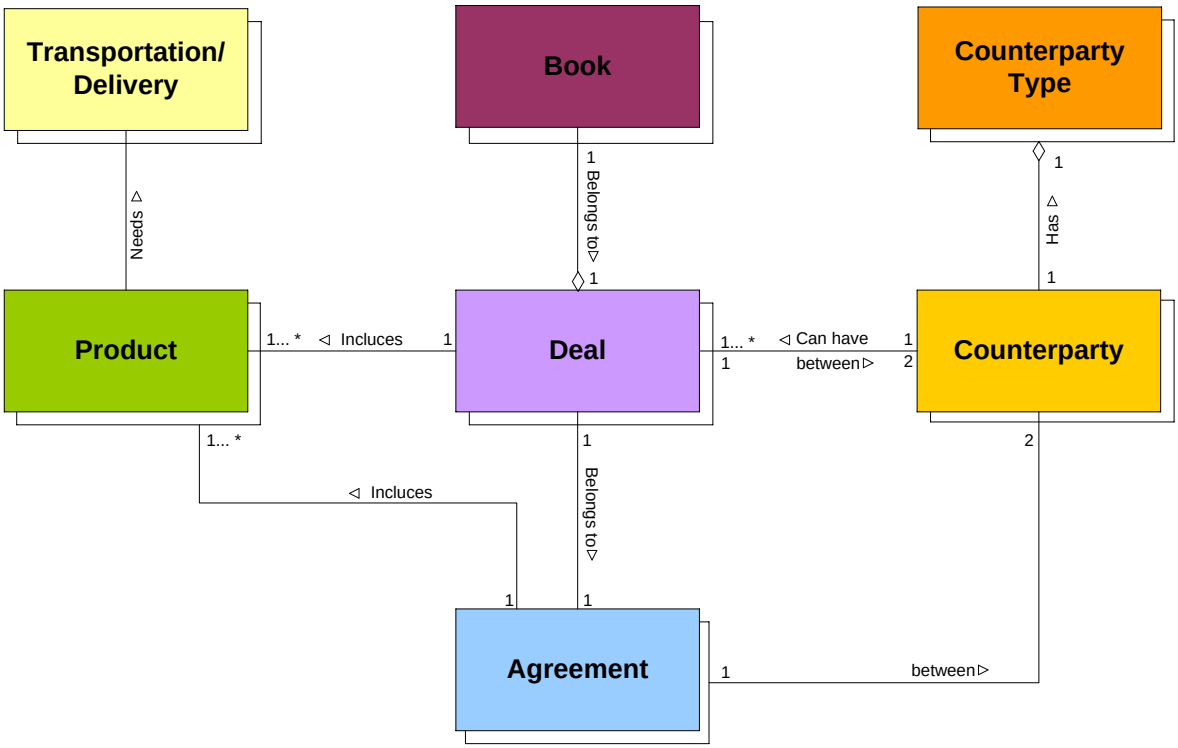
13	C	It has been investigated if there are redundancies in the storing and accessing of data.	Yes
14	D	There is an overview of all data sources and systems and their interaction. Redundancies can be mapped to systems and sources.	Yes
15	E	The overview gets maintained on a regular basis and redundancies are resolved if possible. Superfluous systems are substituted.	No
		Assessment of Data Quality	
16	A	There is a common estimation about the quality of the master data within the organization.	No
17	B	The organization has formalized quality criteria that are important and need to be measured.	Yes
18	C	There are quality requirements defined taking into account the requirements of different business units.	Yes
19	D	A quality assessment has taken place in the organization and it is known which quality the data has.	No
20	E	There are defined intervals in which quality assessment is conducted and changes in quality are monitored.	No
		Impact on Business	
21	A	The organization is aware of the reputational impact on the business if data quality is insufficient.	Yes
22	B	The organization is aware of the monetary impact on the business if data quality is insufficient.	Yes
23	C	The organization knows how much money gets lost due to insuffiecient data quality (e.g. Lost sales opportunities).	Yes
24	D	The organization knows how insufficient data impacts the firm from a non-monetary perspective (e.g. Reputational, customer-retention).	Yes
25	E	The organization can classify the impact of bad master data quality from both monetary and reputational aspects into financial arguments and can state how much money is lost.	Yes
		Awareness of Quality Gaps	
26	A	The organization knows about different reasons for quality issues in master data.	Yes
27	B	The organization knows which reasons for bad quality are relevant in the organization	Yes
28	C	The organization can precisely state which reasons for bad quality are involved at which source of data entering.	No
29	D	The employees are aware of reasons and sources of poor master data quality and the consequences for the business.	Yes
30	E	The organization can precisely state the weak spots in data setup (e.g. Entering information manually - especially foreign words or numbers - results in spelling mistakes).	Yes

Improvement			
31	A	The organization precisely knows in which areas the data quality is not sufficient according to the defined quality requirements.	No
32	B	The organization is aware of the increasing efficiency and effectiveness in daily work if the quality adheres to the requirements. This is relevant for both employees setting up data and those using the data.	Yes
33	C	There is a company-wide benchmarking system in place to measure data quality objectively.	No
34	D	Improvement measures are in place to increase data quality.	Yes
35	E	There is a constant loop of monitoring and improving quality to ensure it has the required quality.	No
Data Usage			
36	A	The organization knows who is using i.e. has access to what data in the organization.	Yes
37	B	The employees know where to get required data. If is assessed if the employees use the provided data sources.	Yes
38	C	For every source of master data it is communicated to the appropriate users that they have access and that the data contains relevant information.	No
39	D	Data repositories get regularly maintained and do not get outdated.	No
40	E	The employees are aware of the sources they have access to and are not reluctant to use any of them (e.g. because of ignorance of the usage)	No
Data Ownership			
41	A	Data elements have an owner who is either an individual or a department.	Yes
42	B	The data elements are logically owned by related roles/departments. The data owner defines purpose, usage and content.	No
43	C	The responsible persons for master data are communicated throughout the organization. The persons have documented responsibilities.	No
44	D	Data stewardships are established for data areas.	No
45	E	Data stewardship is promoted within the organization and fixed in role descriptions of jobs.	No
Data Access			
46	A	There is a protocol to obtain access to data for the employees.	Yes
47	B	Unauthorized personnel are not given access to sensitive data.	Yes
48	C	Every employee is given access to necessary data beforehand. He is automatically equipped with the main sources.	Yes
49	D	The employees have efficient data sources and have access to the data they need and not much more.	No
50	E	The employees know their sources and have a good overview about what they can find where.	Yes

Data Protection			
51	A	The data is secured against external or default threats with up to date solutions.	Yes
52	B	Data access is only activated on request which was granted by the responsible authority.	Yes
53	C	There are clear rules communicated for granting access for certain roles.	Yes
54	D	Access to data (especially sensitive data) is restricted with passwords that need to be changed regularly and adhere to common standards.	Yes
55	E	There is awareness about data protection among the employees, e.g. the employees do not leave their computers unlocked when leaving their desks).	Yes
Storage			
56	A	The data is stored in an efficient way. Loading does not take too long.	Yes
57	B	The data logic is displaying the real world situation.	No
58	C	Automated tools regularly check for redundancies and duplicates.	No
59	D	The data base logic is regularly compared to the real world situation it is meant to depict.	No
60	E	The data is stored with innovative solutions to enable data analysis and forecasting (BI solutions).	No
Data Lifecycle			
61	A	Data is considered as an object that is undergoing a lifecycle and changes over time.	Yes
62	B	Data is valued as an organizational asset that brings value to the organization.	Yes
63	C	The data logic is scalable to treat data according to its position in the lifecycle.	No
64	D	For every data item, a single source of truth is established.	No
65	E	Maintenance labor like entering and updating is automatically logged by the systems.	Yes

Appendix I: The Master Data Model of EET – Data Groups

Figure 15 *Data Groups for EET*



Appendix J: The Master Data Model of EET – Detailed Model

Figure 16 *Master Data Model of EET*