UNIVERSITY OF TWENTE.

MSc. Business Administration

May 2017

A maturity model for improving data quality management

Onur Kirikoglu

First supervisor: Dr. A.B.J.M. (Fons) Wijnhoven

Second supervisor: Dr. ir. S.J.A. (Sandor) Löwik

Colophon

Date	:	31 May 2017
Version	:	5.5
Project Reference	:	Master Thesis
Status	:	Final version

Author	:	Onur Kirikoglu
Student number	:	s1620576
E-mail	:	o.kirikoglu@student.utwente.nl
Education	:	MSc. Business Administration
Track	:	Business Information Management (BIM)
Institute	:	University of Twente School of Management and Governance Enschede, The Netherlands
University supervisors	:	Dr. A.B.J.M. (Fons) Wijnhoven Dr. ir. S.J.A. (Sandor) Löwik

Synopsis

In this research, a maturity model and supporting scorecard are developed, which can be used to determine the current or desirable state of maturity and with this evaluate data quality management within the firm.

Acknowledgements

I have had difficult times during the course of my master thesis and therefore would like to thank some people. Firstly, Fons Wijnhoven for his support, theoretical insights and providing guidance during the course of this study. Secondly, Sandor Löwik for his feedback which made me think about certain subjects differently. In addition, I appreciate the support and patience from both my supervisors and cannot express my gratitude. Lastly, I would like to thank my parents, friends and family for their support in completing my thesis.

Enschede, 31 May 2017

Onur Kirikoglu

Abstract

Nowadays, firms cannot afford any mistakes with the data in their firm. The reason for this is that the data within the firm should provide additional value and not hurt the firm. Therefore, firms should be aware of the quality of their data and take possible actions to improve. This study focuses on providing firms with a tool for determining their state of maturity and evaluate data quality management. Specifically, this thesis investigates maturity models and analyzes different data quality management principles. The reason is that, these models are a fitting tool in assisting organizations in indicating the organization's current or desirable state with regards to a specific topic of concern. During the course of this study, a maturity model and supporting scorecard were developed. These were both applied at nine different firms in the form of case studies and showed the practical applicability. On the basis of the results of this study, it can be concluded that the maturity model and scorecard are well applicable to firms that are looking at ways to improve their data quality management processes. Finally, the maturity model and supporting scorecard proved its usefulness within practice with the help of the conducted case studies and presented some insights in the business processes of the firms using it.

Keywords: maturity models, scorecard, data quality management, business improvement

Contents

List of abbreviations

DBMS	Database management system
CMM	Capability maturity model

List of tables

List of figures

1. Introduction

Poor quality consumer data costs U.S. businesses **$611 billion** a year (Eckerson, 2002). According to a research conducted by Veritas (2015), where they identified the value of data within firms, they found out that the majority of data within firms is neglected, not used or redundant. Within the research 1,475 respondents were covered across 14 countries. The results of this research showed that 14% of all the data within firms is critical data, 32% is redundant data (data that is not relevant for the business) and the remaining 54% is dark data (data that is not used within the business). Furthermore, Veritas mentions that redundant data could cost organizations **$891 billion** by 2020 (Veritas, 2015).

This is shocking because data available within firms and insights gained from them could widely benefit business performance (Chang, 2014; Mantha, 2014; McAfee & Brynjolfsson, 2012; Merino, Caballero, Rivas, Serrano, & Piattini, 2015). Firms use systems to increase their overall business performance and optimize their services. However, these systems usually generate unneeded data but they cannot be missed within firms (Romero & Vernadat, 2016). Furthermore, Tayi & Ballou (1998) mention that "*data is viewed as a key organizational resource and should be managed accordingly*" (p. 54). In other words, the data available within firms should be managed properly so that this data is always ready for use. With this in mind, the correctness of data is also very important. Graham (2015) points out that when creating reports, managers would easily oversee duplicate data and therefore communicate more revenue than actually made. Redman (1996) states that "*Errors in data can cost a company millions of dollars, alienate customers, and make implementing new strategies difficult or impossible*." (p. 99). We can say that firms cannot tolerate any mistakes with regards to their data in their enterprise systems.

Managers are often not able to find the most accurate and useful data within the systems of their firm (Redman, 1996). Redman (1996) found that managers are unware of the quality of the data they use and perhaps assume that data stored within the enterprise systems is correct. Additionally, they found that poor data may cause managers to ineffectively implement strategies (Redman, 1996). Data quality within this research is defined as the "fitness for use" of the data collected, so in what sense the data meets the requirements of the users (Cai & Zhu, 2015). Employees whom manage the data are more aware of its value and meaning, then an employee who is accessing/using it (Tayi & Ballou, 1998). Moreover, Tayi & Ballou (1998) state that the value of a given set of data can be correct, but it could quickly be misinterpreted. This

misinterpretation can occur because there are no standard rules and procedures. Employee A could register the data in a different way than employee B. This will lead to employees making assumptions when viewing the data. Tayi & Ballou (1998) also mention that firms generally assign low priority to data quality. Low qualitative data could not only affect the competitiveness of the firm, but also hurt the firm from within; the trust among employees could be affected, because of constantly receiving invalid information (Ryu, Park, & Park, 2006). Xu et al. (2002) state that data quality is critical to an firm's success, however not many firms take action to solve these issues.

In existing literature, many aspects of data quality and data management are discussed (Batini, Cappiello, Francalanci, & Maurino, 2009; Cappiello, Francalanci, & Pernici, 2004; Cong et al., 2007; Haug, Zachariassen, & van Liempd, 2011; Lee, Strong, Kahn, & Wang, 2002; Ryu et al., 2006; Tayi & Ballou, 1998). The increase of enterprise systems and the direct access to information by managers and employees have increased the need for, and awareness of, high quality data within firms (Lee et al., 2002). Furthermore, Chengalur-Smith, Ballou & Pazer (1999) mention that often managers must make decisions without thinking about the imperfections of the data found in their systems.

Nowadays, more and more systems share and exchange data in order to form an interconnected IT landscape. In addition, Marsh (2005) states that "*data has always been 'wrong', but now the effects of it are much more visible and the consequences more serious."* (p. 105). In the past firms were mainly working with one system. Currently, an error within one connected system could assure that the data within all the other systems are also affected. This means that inaccurate, incomplete and inconsistent data cannot be neglected so easily, because it could directly affect for example the sales numbers.

To contribute to science, a maturity model for data quality management is developed which aims to create awareness for firms. The model creates points of discussion for firms and may guide firms in making plans for improving data quality management. In this research, the following central research question is treated:

"*What is the usefulness of a maturity model for determining the state of data quality management in organizations?"*

To guide this research, the concept of maturity models is used. Specifically, the advantages and disadvantages of data quality management and the impact it can have on firms are treated. Maturity models are widely applied within different types of research and don't always have the same topic of interest. The reason for using the maturity model concept is because this type of model is a fitting tool in assisting organizations in indicating the organization's current or desirable state with regards to a specific topic, in this case data quality management (De Bruin, Freeze, Kaulkarni, & Rosemann, 2005; Pöppelbuß & Röglinger, 2011). In addition, the theories of data quality, databases and data management are treated. The reason for using these is that the maturity model developed within this research is focused on data quality management and the theories add to defining the context of the maturity model. As a limitation to this study, other possible topics that could influence data quality management (e.g. environmental factors, competition etc.) are beyond the scope of the study. Since this study aims to add data quality management within the subject of maturity models and therefore introduces new insights in this topic, it can be classified as explorative research (Babbie, 2012; Crossman, 2016).

This paper is structured as follows. First, a literature search is treated wherein relevant topics are treated. Specifically, data quality management, its quality characteristics and the different approaches to managing data. Moreover, the concept of maturity models are discussed, including discussions about the first maturity model introduced to science by Paulk et al. (1993). Subsequently, the data management maturity model of Ryu et al. (2006) is presented. Additionally, the research problem and model are introduced, where maturity models are critically assessed, design principles of Pöppelbuß & Röglinger (2011) are applied and the initial model for this study is developed. Second, the methodology for how this research is conducted is discussed. Third, data analysis is performed wherein the results of the data collected are treated. Thereafter, the results are presented, which includes the optimized maturity model for data quality management. Lastly, a conclusion is drawn, a discussion is given, contributions are discussed, the limitations of this research are presented and some directions with regards to future research are included.

2. Literature search

This chapter establishes and discusses data quality management. It starts with the methodology of this literature search (2.1). Additionally, data quality and the quality characteristics are presented (2.2). Subsequently, data management is treated and discussed (2.3). Furthermore, the concepts of maturity models are analyzed (2.4). Moreover, the research problem and the model of this study are elaborated (2.5). Lastly, a conclusion of the literature search is given (2.6).

2.1 Methodology for the literature search

The first part of this research is based on a critical review of relevant literature. The literature search was conducted between September 2016 and November 2016. In order to gather relevant articles several search engines were accessed: Google Scholar, ICT Services & Archive (LISA), ScienceDirect and Scopus. No publication date limits or language restrictions were used. A few keywords were applied for the search process: 'data quality' OR 'data quality management' OR 'quality management' OR 'data improvement' OR 'data enhancement' OR 'data management' OR 'database management' OR 'databases' OR 'data maturity model' OR 'master data management'. In some cases, these keywords were used in combination of each other. Additionally, the collected articles were reviewed on the impact score of the journal and number of times the article was cited.

With the literature search a total of 104 articles and 13 books were found. The articles and books provided further insights and points of discussion for this thesis. For finding new articles the same sources were applied. The articles were categorized based on publication data and books were separated from articles.

2.2 Data quality and quality characteristics

Data quality has many definitions, for example, "the perception or assessment of data's fitness to serve its purpose in a given context." (Rouse, 2005), "fitness for use; the consumer viewpoint of quality because the consumer determines whether the data is fit for use." (Tayi & Ballou, 1998; Wang & Strong, 1996), "the capability of data to be used effectively, economically and rapidly to inform and evaluate decisions." (Karr, Sanil, & Banks, 2006). By critically looking at these definitions of data quality, the definition of Tayi & Ballou (1998) and Wang & Strong (1996) will be used as the central definition of data quality within this research. The definition *"fitness for use, by data consumers"* defines the usability of the data and the effective usage of it within firms. In

addition, Cai & Zhu (2015) propose that the judgment of data quality depends on the data consumers (the ones that use the data). Moreover, Strong et al. (1997) define a data quality problem as any difficulty that renders data completely or largely unfit for use. Cai & Zhu (2015) have identified several challenges of data quality: (1) There are many types of data sources that bring different data types and complex structures, this causes difficulty with data integration between different systems; (2) The amount of data within systems has a lot of volume, therefore it is difficult to judge the quality in a given time; (3) Data within firms change very fast and thus the 'timeliness' of data is very short. This may cause outdated or invalid information.

Eckerson (2002) states that although firms understand the importance of high quality data, most of them are blind by the true business impact of defective or inadequate data. Furthermore, Eckerson (2002) mentions two most common problems caused by poor data quality: (1) extra time required to reconcile data and (2) loss of reliability in the system or application because of existing errors. According to Strong et al. (1997) the characteristics of high-quality data consists of four categories: intrinsic, accessibility, contextual and representational aspects. In addition, Pipino, Lee, Wang, Lowell Yang Lee & Yang (2002) also mention some dimensions that could provide an overview of the characteristics of data quality. Table 1 presents these dimensions and elements as categories and characteristics, the characteristics are shown in alphabetical order.

Table 1: Data quality characteristics (Cai & Zhu, 2015; Pipino et al., 2002; Strong et al., 1997).

Categories	Characteristics	Description
1. Availability	a. Accessibility (Strong et al., 1997)	"Accessibility" refers to the amount of access employees have to the data within the firm. Can the employees find the data needed or do they always need to put effort in getting it?
	b. Timeliness (Cai & Zhu, 2015; Pipino et al., 2002)	"Timeliness" refers to the availability of data within a given time and if the data is updated regularly. Furthermore, if data retrieval and processing to release, meet the requirements.
2. Usability	a. Credibility	"Credibility" refers to the amount of maintenance that is performed in order to check the correctness of the data.
3. Reliability	a. Accuracy	"Accuracy" means that data within the firm is correct and precise, and no errors are found in the gathered information.
	b. Consistency	"Consistency" means that the same data within different systems should be identical and no differences should be present.
	c. Integrity	"Integrity" means that the data is clear and meets the given criteria.
	d. Completeness (Pipino et al., 2002)	"Completeness" means that no data is missing in the information that has been gathered. For example, no orders can be excluded from the total amount of sales.
	e. Free-of-Error (Pipino et al., 2002)	"Free-of-Error" refers to the extent in which data is correct and contains no errors. Furthermore, data with no errors is regarded as reliable.
4. Relevance (Pipino et al., 2002)	a. Fitness	"Fitness" refers to the usability of the data retrieved by employees and if this retrieved data meets the users' needs.
5. Presentation (Strong et al., 1997)	a. Reliability	"Readability" refers to the extent in which the data is clear and easy to understand.

2.3 Data management

Databases are used to store, manipulate, and retrieve data in practically every kind of organization, including business, health care, education, government and libraries (Hoffer, J. A., Prescott, M.B., Topi, 2008). Databases are used by everyone in their daily life, whether it is writing a message on Facebook or storing your invoices in a finance software. If the systems used within the firm are arranged properly, then the customers of the firm can also access these systems and add new data to the databases or change existing data within the databases. Hoffer, Prescott & Topi (2008) state that many organizations have incompatible databases that were developed to meet immediate needs, rather than based on a planned strategy or a well-managed evolution. Furthermore, Hoffer, Prescott & Topi (2008) mention that "*much of the data is trapped within older systems, and the data are often of poor quality."* (p. 45). Hoffer, Prescott & Topi (2008) define a database as an organized collection of logically related data. Databases are of any size and complexity. Nowadays, databases store any kind of data, such as documents, maps, photos, sound and video segments. Hoffer, Prescott & Topi (2008) define data as stored representations of objects and events that have meaning and importance in the users' environment. Furthermore, Hoffer, Prescott & Topi (2008) define information as data that has been processed in such a way as to increase the knowledge of the person who uses the data. A way to convert data into information is to summarize them and create a report (Hoffer, J. A., Prescott, M.B., Topi, 2008).

Data only becomes useful when put into context. The primary mechanism for providing context for data is metadata. Hoffer, Prescott & Topi (2008) define metadata as data that describes the properties or characteristics of end-user data, and the context of that data. Some of the properties that are described include data names, definitions, length (or size), and allowable values. For example, the name of a student, the name of a course or the number of a section. Metadata enable database designers and users to understand what data exists, what the data mean, and what the fine distinctions are between seemingly similar data items (Hoffer, J. A., Prescott, M.B., Topi, 2008). The management of metadata and data is equally important, because without a clear meaning data can be confusing, misinterpreted, or erroneous. Furthermore, metadata is not only applicable within databases but also within documents, for example the date of creation, the owner of the document or other similar specifications.

A database management system (DBMS) is a system that is used to create, maintain, update, store, retrieve and provide controlled access to user databases (Hoffer, J. A., Prescott, M.B., Topi, 2008). A DBMS also enables users and programmers to share data among diverse applications, however the application in concern should support data sharing. Designing a database accordingly is fundamental to creating a database that meets the users' needs. Data models are used to capture the relationship and nature among data. Hoffer, Prescott & Topi (2008) state that the effectiveness and efficiency of a database is directly associated with the structure of the database. A data model exists of objects, also known as entities. Examples of entities are Customers, Invoices, Cases and Orders. Information about each of these entities is referred to as an instance, for example the name of a customer or her given ID. A well-structured database establishes the many relationships between entities that exist in organizational data so that the desired information can be retrieved. Relational databases establish the relationship between entities by a common field, for example the ID of an order and the ID of a customer. With the help of a relational database, it is possible to create a relation between the different entities. For example, customers and their orders.

The use of databases is not a standard within firms, mostly databases exist and grow with the help of a specific system. When no databases are used, firms tend to store their data within files. However, Hoffer, Prescott & Topi (2008) mention that storing data within files do have some disadvantages: (1) Program-Data dependence: when changes occur within a file that is used within several systems, then all systems need to be updated with the new changed file in order for the data to be up-to-date; (2) Duplication of data: the different files may contain the same customer data, e.g. orders and invoices may be two different files with the same customer data; (3) Limited data sharing: because the data is not stored within databases, employees cannot access every separate file within different applications. Therefore, employees need to request data from each other. (4) Lengthy Development times: every new addition ensures that developers need to start from scratch by designing new file format descriptions and afterwards connect these to the corresponding applications and (5) Excessive Program Maintenance: the previous points described create a heavy load of maintenance. Furthermore, Hoffer, Prescott & Topi (2008) mention that over 80 percent of the development budget might be used for maintenance, leaving little opportunity for new developments. However, this does not mean that these cannot occur with the implementation of databases. A database requires some form of maintenance in order to tackle these disadvantages, creation of a database will not directly solve these.

The database approach underlines the integration of system applications and enables sharing of data throughout the organization, or at least across major divisions within the organization (Hoffer, J. A., Prescott, M.B., Topi, 2008). The advantages of the database approach are almost the opposite of working with files. Table 2 presents the advantages mentioned by Hoffer, Prescott & Topi (2008).

Table 2: Advantages of using databases.

Advantage	**Description**
Program-data independence	Data descriptions are stored in a central location and therefore it allows to change data without adjusting the system that processes it.
Planned data redundancy	The data of a particular order is located in one table, this may ensure that additional information about orders are not included within other tables, but gathered from the orders table.
Improved data consistency	By eliminating data redundancy, the opportunities for data inconsistency are reduced. If done correctly customer information will only be stored once and adjusted at the same location and not entered twice.
Improved data sharing	A database is designed as a shared corporate resource (Hoffer, J. A., Prescott, M.B., Topi, 2008). To access the database the administrator can grant employees privileges in accessing the different kinds of information.
Increased productivity of application development	A major advantage of implementing databases is that it greatly reduces the cost and time for developing new business applications (Hoffer, J. A., Prescott, M.B., Topi, 2008).
Enforcement of standards	The database administrator should establish and enforce data standards. These standards will include naming conventions, data quality standards, and processes for accessing, updating and protecting data. Hoffer, Prescott & Topi (2008) state that the most common source of database failures happen because of failure in implementing database administration.
Improved data accessibility and responsiveness	With the help of a relational database, programmers can retrieve and display data accurately, even when the data is spread over different departments.
Reduced program maintenance	Within a database environment, data are more independent of the applications that use them. With this in mind, program maintenance can be reduced in a modern database environment.
Improved decision support	Some databases are designed for decision support applications. For example, to support customer relations or inventory management.

Besides the advantages of database systems, Hoffer, Prescott & Topi (2008) also recognize several costs and risks: (1) New, Specialized Personnel: organizations that adopt the database approach need to hire specialized personnel to implement and manage these databases and because of the rapid changes in technologies these employees need to be trained on regular basis; (2) Installation, management cost and complexity: A multiuser database system is a large and complex system, has high initial cost, requires trained staff and has annual maintenance and support costs; (3) Conversion costs: the costs of converting old file handling systems to a database are often high. These costs are measured in terms of money, time and organizational commitment; (4) Need for explicit backup and recovery: The data within databases should be available anytime, and during casualties a backup should always be available. Based on the importance of the data the frequency of backups can be set and (5) Organizational conflict: conflicts on data definitions, data formats and coding may occur, handling these issues require organizational commitment. Lack of commitment may cause bad decision making that threatens the well-being of the organization (Hoffer, J. A., Prescott, M.B., Topi, 2008).

Two different types of database management exist: (1) the management by systems: that ensures that the data is stored in the right place and available when needed and (2) management by users: that assure that the data is correct by performing daily checks and ensure that new data enters the different systems correctly. The second type can potentially lead to poor data quality because, mistakes can be made when inserting the data in the 'correct' systems. In addition, users of the data could willingly neglect addition of 'new information' because of the potential harm for their own career. Neglecting this data, may cause missing and incomplete data. If all the data is not present, managers will make assumptions based on the data they at that point possess.

Hoffer, Prescott and Topi (2008) mention that using default values or not permitting empty values could be a solution for missing data. However, missing data is not completely unavoidable. Babad and Hoffer (1984) mention that *"when data are missing, lost, or incomplete, procedures may work incorrectly; computations may lose reliability, and their results may have to be interpreted as estimates (if they are even valid at all)."* (p. 748). Additionally, Babad and Hoffer (1984) define a set of procedures for handling missing data:

1. Substitute an estimate for the missing value: when calculating the total monthly product sales, take the mean of existing monthly sales indexed by total sales of that month. This estimate should be marked so that the user knows it is an estimate;

2. Track missing data so that special reports and other system elements cause people to resolve unknown values quickly: this can be done by setting up triggers in the database definition. These triggers are routines that trigger when a special event occurs;
3. Perform sensitivity testing so that missing data are ignored unless knowing a value might significantly change results: monitoring thresholds would be of case, for example when the compensation of an employee based on sales almost reaches a limit that would make a difference in his/her compensation.

Missing and erroneous data affect the following data quality characteristics mentioned in table 1 (see page 12): Timeliness, Credibility, Accuracy, Completeness, Free-of-Error and Fitness. Timeliness is affected because the data within the firm is not updated regularly and therefore contains errors or missing data. If a lot of missing or erroneous data is present within the firm one may regard the amount of attention given to maintenance as low and therefore affecting the usability of the data. When errors exist within the data the accuracy of the data is low and therefore not reliable in making decisions. If a lot of missing data is present within the firm the data of the firm cannot be regarded as complete and therefore affecting the characteristic completeness. Furthermore, if a lot of errors are also present within the firms then the characteristic Free-of-Error is affected and therefore also the reliability of this data. Lastly, the characteristic Fitness is affected because the data cannot be used because it contains errors or the totals do not represent good values because of missing data. By looking at the previous mentioned, it can be concluded that missing and erroneous data can negatively affect a firms' performance and therefore taking these into consideration is important.

2.4 Maturity Models

Since the introduction of the capability maturity model (CMM) of Paulk et al. (1993), different types of maturity models are introduced by many researchers. For example, the Process Maturity Model developed by the Rummler-Brache Group (De Bruin et al., 2005) and the Project Management Maturity Model developed by the Office of Government Commerce, UK (De Bruin et al., 2005). The goal of a maturity model is to reach a certain level of maturity. Paulk et al. (1993) define the differences between immature and mature organizations: *"In an immature organization, processes and tasks are improvised by practitioners and managers during a project. Even if a process has been specified, it is not correctly followed or enforced"* (p. 19) and *"A mature organization has the organization-wide ability to manage development and maintenance. Managers communicate well with staff and work activities are carried out according to plan."* (p. 19). The CMM is based on the idea of continues improvement and aid organizations in prioritizing its improvement efforts. The CMM proposes five different maturity levels, in which achieving each level of maturity establishes a different component in a software process, resulting in an increase in the process capability of an organization (Paulk et al., 1993). Each maturity level forms a foundation for the next. The five maturity levels are described below (Paulk et al., 1993):

1. At **level 1: Initial**, an organization typically does not provide a stable environment for developing and maintaining software. These kinds of firms have difficulties with commitment of staff and this can result in crisis. During a crisis, projects typically abandon planned procedures. Focus is given to individuals, not organizations.
2. At **level 2: Repeatable**, policies for management and procedures to implement those policies are established. New projects are based on experiences with similar projects. Project standards are defined and the organization ensures that they are faithfully followed. Level 2 organizations are *disciplined* because project planning and tracking are stable and earlier successes can be repeated.
3. At **level 3: Defined**, a typical process for developing and maintaining software across the organization is documented, including the software-engineering and management processes. A defined process contains a coherent, integrated set of well-defined software-engineering and management processes, both are stable and repeatable. This process capability is based on a common, organization-wide understanding of activities, roles, and responsibilities in a defined process (Paulk et al., 1993).

4. At **level 4: Managed**, an organization sets quantitative quality goals for both products and processes with well-defined and consistent measurements. An organization-wide process database is used to collect and analyze the data available from a project's defined processes. The risks involved in moving up the learning curve are known and carefully managed. When limits are exceeded, managers take action to correct the situation.
5. At **level 5: Optimizing**, the entire organization is focused on continuous improvement. The organization has the means to identify weaknesses and strengthen the process proactively, with the goal of preventing defects. At level 5, waste is unacceptable; organized efforts to remove waste result in changing the system by changing the *common* causes of inefficiency. Reducing waste happens at all maturity levels, but it is the focus of level 5. Improvement occurs both by incremental advancements in the existing process and by innovations in technologies and methods (Paulk et al., 1993).

Paulk et al. (1993) mentions several attention points with regards to the maturity levels. For example, level 1 organizations often miss their deliver date based on their initial schedule by a wide margin. More mature organizations should be able to meet the targeted dates with increased accuracy. Moreover, as maturity increases, the variability of actual results around targeted results decreases. In addition, similar projects in mature organizations, should be delivered within a smaller range. As organizations mature, costs decrease, development times shorten and productivity and quality increase. Development times in level 1 organizations are time consuming, because of rework needed. This rework is mostly needed because of incorrectly following the defined standards and procedures.

Skipping a maturity level is not advised, because each maturity level lays the foundation for achieving the next and is therefore counterproductive. Paulk et al. (1993) state that processes without a proper foundation fail and provide no basis for future improvement. Specifically, achieving higher levels of maturity is incremental and requires a long-term commitment to continues process improvement and should therefore be conducted accordingly. The CMM identifies the characteristics of an effective software process, but the mature organization addresses all issues that are essential for a successful project, including people, technology and process.

In addition to the CMM of Paulk et al. (1993), a maturity model for data quality management has been developed by Ryu et al. (2006). This maturity model focuses on data quality,

as this topic is regarded as important because it is the basis of an information system. Moreover, quality plays an important role within businesses and acts as one of the powerful metrics to gain competitive advantage (Ryu et al., 2006). In order to increase the competitiveness of organizations this maturity model has been introduced. The maturity model helps to appraise firms' levels of data quality management and to acquire better quality. The four defined maturity levels are described below (Ryu et al., 2006):

1. At **level 1: Initial**, the data structure quality is managed through the rules defined in the database system catalogue. It is the early stage of data management.
2. At **level 2: Defined**, data is managed through the logical and physical data models. If the data structure is modified or remodeled, it should refer to the data model. The modification should be returned as a new input to the database.
3. At **level 3: Managed**, data is managed through data standardization. This stage is focused on the management of metadata (information about the data within the database) that selects all corporate data and standardizes various attributes, schema, domain and data model (Ryu et al., 2006). Level 3 also enables sharing and reusing of standardized data through standardization of metadata. With the help of data standardization, technical standards for data are set and all the data and their data types are correctly registered within the databases.
4. At **level 4: Optimized**, data management is data management through data architecture management. Ryu et al. (2006) mention that "*This level is to define the enterprise standard architecture model, which is the optimized data management stage to manage the data, data model, and data relationship on the basis of the defined enterprise standard architecture model."* (p. 194). The enterprise standard architecture model aids firms in making analysis of their firm, create the right planning and implement these correctly. This architecture model applies principles and practices to guide organizations through their business. Fischer, Aier &Winter (2007) mention that the architecture model should always remain up-to-date and reflect to the current state of corporate structure and processes.

2.5 Research problem and model

In this research, a maturity model for data quality management will be developed. This maturity model aims to determine the state of maturity of an organization and with this provide an overview of the current state of the organization by looking at data quality management principles. With this in mind the maturity model of Ryu et al. (2006) will be critically analyzed and looked at from a new angle. In the current maturity model firms cannot directly categorize themselves within a specific maturity level, additional effort is needed for categorization within this model. The firm using the model needs to sit together with some technical personnel, to discuss the certain aspects of each maturity level and classify themselves accordingly. The maturity model of this study focuses on providing an easy tool that can be used by any employee that is included within the business processes of the firm, for example managers. However, the maturity model is not applicable to any type of organization, for instance a bakery around the corner would not easily start in applying this maturity model within her firm. But, firms that are dealing with a decent amount of data and if data is critical in their business processes, this model would be suitable.

Poppelbuß & Röglinger (2011) have defined a set of design principles that maturity models as design products should meet, because there is no comprehensive understanding of relevant principles. Poppelbuß & Röglinger (2011) have created a checklist which researchers that are involved in the design of a maturity model can use. This checklist is used in the development of the maturity model of this study. In addition, several purposes of use for maturity models are mentioned:

- **Descriptive**: this purpose of use applies when the maturity model is used for assessing the current capabilities of the firm with respect to certain criteria. The model hereby is used as a diagnostic tool (Pöppelbuß & Röglinger, 2011).
- **Prescriptive**: this purpose of use applies when the maturity model is used to indicate how to identify desirable maturity levels and provides some guidelines for improvements.
- **Comparative**: this purpose of use applies when the maturity levels of similar business units and organizations can be compared with each other.

The purpose of use for the maturity model of this research can be categorized as ‘descriptive’ because the model acts as a tool for diagnosing the current state of an organization with regards to data quality management. Poppelbuß & Röglinger (2011) have defined some basic principles that

could be applied in the design stage of a maturity model. In addition, some principles with regards to the purpose of use, in this case 'descriptive' are also mentioned. Nevertheless, Poppelbuß & Röglinger (2011) state that *"We do not require each maturity model to meet all design principle. Instead, the framework serves as a checklist when designing new maturity models."* (p. 5). With this in mind, the design principles will be followed accordingly. Table 3 presents the checklist applied for designing the maturity model of this study.

Table 3: Design principles 'checklist' (Pöppelbuß & Röglinger, 2011).

1. Basic (design principles)		
1.1	**Basic information**	
	a. Application domain and prerequisites for applicability	x
	b. Purpose of use	x
	c. Target group	x
	d. Class of entities under investigation	x
	e. Differentiation from related maturity models	x
	f. Design process and extent of empirical validation	x
1.2	**Definition of central constructs related to maturity and maturation**	
	a. Maturity and dimensions of maturity	x
	b. Maturity levels and maturation paths	x
	c. Available levels or granularity of maturation	x
	d. Underpinning theoretical foundations with respect to evolution and change	
1.3	**Definition of central constructs related to the application domain**	x
1.4	**Target group-oriented documentation**	
2. Descriptive (design principles)		
2.1	**Intersubjectively verifiable criteria for each maturity level and level of granularity**	x
2.2	**Target group-oriented assessment methodology**	
	a. Procedure model	x
	b. Advice on the assessment of criteria	x
	c. Advice on the adaptation and configuration of criteria	x
	d. Expert knowledge from previous application	

The design principles that are applied in the design process of the maturity model will be checked within the checklist that is given in Table 3. Poppelbuß & Röglinger (2011) have categorized the design principles into different groups, these are applied in this study accordingly.

- **1.1**: (a) The domain where this maturity model will be applied is the field of data quality management, because data quality management is nowadays not the focus of many firms, while it of huge importance (Cai & Zhu, 2015; Strong et al., 1997). The prerequisites for applicability of this model are that the firm using the model should be highly dependable on their data, cannot afford mistakes in their data, should work with critical data within their business processes and must have some knowledge of the firms' business processes. (b) The purpose of use for the maturity model is 'descriptive' because the model acts as a tool for diagnosing the current state of an organization. (c) The target group of the maturity model are mostly managers that have a good overview of the business processes of the firm or employees that also possess this same view. It is targeted towards managers because not every employee on the workplace may have the needed knowledge about the firms' business processes. (d) The entities selected to be investigated for this study are diversified, from insurance, software and consultancy firms to government agencies. This diversification is applied, so that different views towards the maturity model of this study could be gathered. (e) The model in this study provides a new angle to the model of Ryu et al. (2006), which is more focused towards databases of the firm and the management of these and has a technical nature. The difference of the maturity model in this study is that it focuses on individuals that are known with the firms' business processes and with this can categorize themselves within a certain maturity level. (f) Lastly, the maturity model will be subject to empirical validation with the means of interviews that will be conducted with a variety of individuals. In addition, within the interviews an overview of data quality management will be gathered and afterwards the developed maturity model will be discussed and the model usage will be tested.
- **1.2**: (a) In the maturity model of this study, the decision had been made to include five stages (levels) of maturity. The reason is that the founding maturity model (CMM) has defined five maturity levels, it has been cited 2790 times and therefore used as the foundation for the model of this study. In addition, the five level provide a good overview of consequent steps for reaching the final stage of continues improvement. The following

dimensions are applied: (1) person dependent and basic, (2) policies, standards and procedures, (3) defined and stable, (4) managed and standardized, (5) continues improvement. Table 4 provides elaboration on these dimensions.

Table 4: Maturity stages of the maturity model.

Maturity levels	**Definitions**
Level 1: Person dependent and basic.	Many tasks with regards to data quality management are performed by one individual, this causes uncertainty within firms. Because, when the employee of the firm is not present, the firm will lose some knowledge and influence. In addition, the systems of the firm are not maintained but solely used.
Level 2: Policies, standards and procedures.	The firm develops policies, standards and procedures, so that these can be followed by the individuals within the firm. These ensure that the organization can repeat earlier success because these are defined and can be followed over again.
Level 3: Defined and stable.	This level of maturity is reached when the firm applies every small change to their data structure and reflects this change towards their data model. This creates new input possibilities within the system of the firm and helps managers to perform more effectively. Additionally, the firm ensures that the employees are educated well and possess the knowledge and skills they require.
Level 4: Managed and standardized.	All the data within the firm is standardized. This enables sharing and reusing of standardized data through standardized metadata. Data standardization ensures accuracy and integration of the information that enters the systems of the firm. This process also makes is easier to analyze and ensure reliability of the data. Furthermore, the organization sets quantitative goals for both products and processes with well-defined measurements.
Level 5: Continues improvement.	The final stage of maturity aims towards continuously improving. Strengths and weaknesses are known and can be identified. The main focus in this level is to reduce waste, however this also takes places in other levels but is not the main focus. The enterprise standard architecture model of the firm is defined, this provides a basis for successful development and execution of a strategy.

(b) To these maturity level corresponding maturation paths (processes) could be related. These processes could provide some meaning of reaching the next maturity level. From level 1 to level 2 the path is defined as a *disciplined process*, because firms need to discipline themselves and optimize their ways of working. In order to achieve maturity level 2, it is not needed for tasks to be person independent, because firms of small size are mostly bond to one individual taking care of certain tasks. From level 2 to level 3 the path is named

the *standard consistent process*, because policies, standards and procedures are defined. This ensures consistency throughout the organization. From level 3 to level 4 the maturation path is called the *predictable process*, because measurements are made and therefore predications about future trends can be drafted. The process is both stable and measured, this ensures that during circumstances, managers can take actions to correct the situation. From level 4 to level 5 the path is defined as the *continuously improving process*, because firms following this process should always aim to improve. Firms can identify their strengths and weaknesses, with the goal of preventing mistakes. Additionally, teams in level 5 organizations aim to determine causes of events and evaluate these to prevent reoccurring errors in the future.

- **1.3**: The domain where the maturity model would be applicable are firms whom are dealing with a decent amount of data within their business processes and these data are critical in the daily processes of the firm. This excludes firms like bakeries who are not really focused on their data, but rather towards generating enough turnover before the end of the month. However, government agencies cannot afford mistakes within their data and therefore need the quality of their data to be at a certain level and could use the maturity model of this study to determine their current state and build up upon that. Some central constructs that could apply to this application domain are: data quality, usability and organizational performance.

  2.1: The criteria that are defined for each maturity level originate from the description of each level, so that consistency between both can be achieved. Table 5 shows the criteria defined for each maturity level. Probably some additional criteria could be defined, however during the analysis of the theory only these criteria are included.

Table 5: Criteria for each maturity level.

Maturity level	**Criteria**
Level 1	dependency, maintenance, competence
Level 2	repeatability, disciplined
Level 3	effectivity, educating, consistency
Level 4	repetition, reusability, measured, predictable
Level 5	improving, reducing waste

Poppelbuß & Röglinger (2011) also mention within their checklist the procedure within the maturity model, so how you get from one level to the other. Additionally, some advice about the selected criteria needs to be included. Moreover, expert knowledge from previous application is not available for this study, because a new not previously applied model will be used. In the next paragraph, the maturity model for this study will be presented and the procedure and advice on assessment of criteria will be discussed.

Maturity model for Data Quality Management

The maturity model of this study provides a quick overview of each maturity level and its criteria. Each maturity level is shown in a column, within the column the characterizing descriptions of each level are summed up. In addition, the maturation paths for each advancement are illustrated with an arrow below the description of each maturity level. Lastly, the criteria for each maturity level are given. Table 6 illustrates the maturity model of this study.

The maturity model provides a good overview of each maturity level. Firms can use this model to think about their current or desirable state with regards to data quality management. It is important to note that each level is forms the foundation for the next level. The contribution of data quality management for this model is that it characterizes some aspects of data quality within. In addition, each maturity level given a certain amount of attention to data quality. For instance, in level 2 organizations have defined some policies, standards and procedures, but do not reflect changes in the data structure towards their data model. By not doing this the firm misses new input possibilities and therefore also new information towards the systems, this refers back to the data quality characteristic *completeness*, because the data sets are not complete and some are missing. However, the maturity model does not provide a guided indication of a maturity level, the firm has to think about each maturity level and categorize themselves based on their thoughts and opinions. Therefore, a scorecard is developed besides the maturity model who provides an overview of each maturity level. For the scorecard, factors based on the descriptions and criteria of each maturity level are created. Each factor will be scored with a fitting Likert scale, because the levels of maturity are incremental, which means that the higher the level the better the firm should perform. The scorecard and maturity model are created separately, because otherwise it would be too complicated to comprehend and the scorecard should act as a supporting tool for the maturity model and not replace it. Table 7 illustrates the supporting scorecard.

Table 6: Maturity model for determining the level of data quality management.

Maturity **level 1**: Person dependent and basic.	Maturity **level 2**: Policies, standards and procedures.	Maturity **level 3**: Data model optimizations.	Maturity **level 4**: Managed and standardized.	Maturity **level 5**: Continues improvement.
1.) Focused on individuals, therefore tasks are person dependent. **2.)** Knowledge is lost when individual is not available. **3.)** Systems in the firm are not maintained regularly.	**1.)** Policies, standards and procedures are defined or updated. **2.)** Earlies success can be repeated, because these are defined accordingly.	**1.)** Changes in the data structure of the firm are reflected to the data model. **2.)** The firm educates her employees so that new knowledge and skills can be acquired.	**1.)** The data within the firm is standardized. This ensures sharing and reusing. **2.)** The firm sets quantitative goals for both products and processes with well-defined measurements.	**1.)** The firm is continuously improving. **2.)** The firm can identify their strengths and weaknesses. **3.)** The main focus of the firm is to reduce waste. **4.)** The enterprise standard architecture model of the firm is defined, this can help in the successful development and execution of a strategy.
	Disciplined process →	Standard consistent process →	Predictable process →	Continuously improving process →
dependency, maintenance, competence	repeatability, disciplined	effectivity, educating, consistency	repetition, reusability, measured, predictable	improving, reducing waste

Table 7: Scorecard for determining the current state of maturity.

Factors	Maturity **level 1**	Maturity **level 2**	Maturity **level 3**	Maturity **level 4**	Maturity **level 5**
The maintenance of applications is person dependent.	None or 1 person	2 or 3 persons	Team	Division	Enterprise
The data provided to the systems is mainly done by employees.	1 person	2 or 3 persons	Team	Team and enterprise systems	Enterprise systems
The data within applications is managed by applying updates.	Never updated	Yearly updated	Monthly updated	Weekly updated	Always updated
Standards and procedures are defined within the firm, that can be repeated in the future.	None defined	Roughly 25% defined	Roughly 50% defined	Roughly 75% defined	All defined
New gained data is stored within the systems of the firm, so that incomplete data can be avoided.	Not stored	Roughly 25% stored	Roughly 50% stored	Roughly 75% stored	Always stored
The data model of the firm is updated when new metadata arises.	Never updated	Yearly updated	Monthly updated	Weekly updated	Always updated
Corporate data within the firm is standardized (standz.).	Not standz.	Roughly 25% standz.	Roughly 50% standz.	Roughly 75% standz.	All data standz.
The firm is aiming to continuously improve.	Never improving	Rarely improving	Often improving	Very often improving	Always improving
The firm can identify their strengths and weaknesses with regards to data management.	None identified	1 or 2 identified	2 or 5 identified	5 or more identified	All possible identified
The main focus of the firm is to reduce waste.	Never focused on reducing waste	Rarely focused on reducing waste	Monthly focused on reducing waste	Weekly focused on reducing waste	Always focused on reducing waste

The factors within the scorecard all have equal impact towards the final result (current maturity level) of the scorecard. Each factor has been elaborated in more detail in Table 8. The usage of the scorecard is advised as follow: (1) for each factor the scale closest to reality of the firm needs to marked, preferably with a yellow marker, (2) the maturity level that has been marked the most can

be regarded as the current maturity level of the firm. However, there could be scenario's that more than one maturity level is marked. In this case, the lowest maturity level should be preferred, before advancement to a higher level is made. The factors in the lower maturity level should be improved upon first. (3) It should be noted that other factors besides the one mentioned could also have an influence on the maturity level and data quality management within the firm. These factors are currently unknown and therefore not included within the scorecard. The choice for ten factors is also related to the fact that some are unknown and based on the descriptions of each maturity level these were regarded as sufficient to determine the current state of maturity. (4) The factors influence the data quality of the firm, for example when procedures are defined the data will reach the correct systems and stored correctly. This ensures that correct analysis can be made.

Table 8: Scorecard factors.

Factors	**Elaboration**
The maintenance of applications is person dependent.	The firm has assigned an individual who is responsible for the maintenance and only that individual possesses the knowledge.
The data provided to the systems is mainly done by employees.	This is aimed towards, how much the tasks within the firm are automatized versus manual tasks.
The data within applications is managed by applying updates.	This factor refers towards the up-to-dateness of the data and how complete the available data is.
Standards and procedures are defined within the firm, that can be repeated in the future.	If this is achieved, employee can approach documents wherein is stated how they should perform a certain task. These documents should always be available, and therefore repeatable.
New gained data is stored within the systems of the firm, so that incomplete data can be avoided.	New retrieved information is directly stored within the systems and not documented somewhere not approachable by others.
The data model of the firm is updated when new metadata arises.	Previously not available data is made available within the systems, so that these can be included in future registrations.
Corporate data within the firm is standardized (standz.).	By standardizing the data, accuracy and integration of the information within the systems can be established.
The firm is aiming to continuously improve.	This factor is aimed towards removing repetitive tasks by improving for example procedures.
The firm can identify their strengths and weaknesses with regards to data management.	The firm is aware of mistakes within their processes and know how they could solve these. The firm also knows their strengths and know how to operate based on these.
The main focus of the firm is to reduce waste.	The focus is given on removing redundant data and stick to the usable data within the firm.

2.6 Conclusion of the literature search

Firstly, data quality is discussed and the definition '*fitness for use, by data consumers*' is used as the central definition of data quality within this research. Furthermore, two problems caused by poor data quality are mentioned: extra time needed to reconcile data and loss of reliability because of existing errors. In addition, several data quality characteristics are presented and briefly discussed. The presented characteristics are: *Accessibility*, *Timeliness*, *Credibility*, *Accuracy*, *Consistency*, *Integrity*, *Completeness*, *Free-of-Error*, *Fitness* and *Reliability* (2.2).

Second, data management as a separate topic has been discussed (2.3). Within this section of the literature search, the principles of managing data accordingly are discussed. Additionally, examples of poor data management are shown and the advantages and disadvantages (costs and risks) of using databases are noted. For instance, databases provide improved data sharing, because employees can easily access the systems and gather the data needed. Furthermore, implementing a good database system within the firms needs someone with that knowledge, this could be costly depending on the task that needs to be conducted. Lastly, some procedures for handling missing data are mentioned. For instance, using estimates when the data is not available.

Lastly, the capability maturity model of Paulk et al. (1993) and the maturity model for data quality management of Ryu et al. (2006) are discussed. The maturity levels of each model are elaborated (2.4). Additionally, the design principles by Poppelbuß & Röglinger (2011) that maturity models as design products have to meet are discussed (2.5). With these in mind, the model of this research has been categorized as *descriptive* because the model acts as a tool for diagnosing the current state of an organization with regards data quality management. Furthermore, the maturity model of this research is illustrated (table 6) and a scorecard that acts a tool for determining the current state of maturity is developed (table 7).

In the next chapter, the methodology of this study will be discussed. Particularly, the type of research, reliability and validity, data collection and data analysis are elaborated in more detail, so that the actions taken in this research are treated.

3. Methodology

This chapter will discuss in depth how this research was conducted. First, the choice for qualitative research is presented (3.1). Second, the reliability and validity of this research are discussed (3.2). Third, how data was collected for this research is explained (3.3) and, as last data analysis is treated (3.4).

Since this study aims to add a new angle towards data quality management within the subject of maturity models and with this introduce new insights in this topic, it can be classified as explorative research (Crossman, 2016). Furthermore, within every research the discussion whether qualitative or quantitative methods are more valid is treated, therefore both methodologies will be briefly discussed.

3.1 Qualitative vs. Quantitative research

Quantitative research is an approach for testing objectives (theories) by investigating the relationship among different variables (Creswell, 2013). Mostly the researcher defines a hypothesis and tests this hypothesis with statistical procedures through fixed variables. Quantitative research follows deductive reasoning: this implies reasoning from one or more statements in order to reach a certain conclusion. Furthermore, quantitative research is suitable for relatively large samples (Silverman, 2013). Interviews within quantitative research are mainly structured and are conducted with the help of surveys (Edwards & Holland, 2013).

Qualitative research is an approach for exploring and understanding the meaning of individuals with regards to a specific problem (Creswell, 2013). Within this type of research the researcher develops questions and procedures in order to gather additional information for his/her research. Qualitative research follows inductive reasoning: this implies reasoning in providing strong evidence for the truth of a drafted conclusion and also generalizes the observations made. Semi-structured and unstructured interviews are mostly used within qualitative research settings. Edward & Holland (2013) mention that these types of interviews are characterized by increasing levels of flexibility and lack of structure.

Furthermore, validity and reliability of qualitative research are discussed in section 3.2. Qualitative research is regarded as subjective because it is based on personal opinions or experiences of individuals and is not regarded as objective because it is not based on facts. Within

qualitative research generalization of the findings is not the main goal, it aims to understand and explore certain cases and contexts (Bryman, 2013).

The research method chosen depends mainly on the central research question of a study. Since this study aims to add data quality management within the subject of maturity models and therefore introduces new insights in this topic, it can be classified as explorative research. A qualitative research design will be applied within this study, wherein insights in the topic of data quality management will be collected.

3.2 Reliability and Validity of this research

Reliability is aimed towards whether the results of a study are repeatable by different researchers or by the same researcher at a different time and place, and if these would give the same results. Achieving reliability is no easy task within qualitative studies, because the results of the study may change based on the perception of each researcher. Reliability can further be divided to external and internal reliability. External reliability is the degree to which the findings of the study can be replicated (Bryman, 2013). For this study, the findings of the study can be replicated in some way. However, identical replication could probably not occur, because of a 'open' interview structure. This means that the interviewee would probably not answer exactly the same, because he or she gives an answer based on their thoughts at that exact moment. Nevertheless, also each individual will have a different look at a certain subject. In addition, Bryman (2013) states that *"true replication of a qualitative research is difficult because it is unstructured and often reliant upon the qualitative researcher's ingenuity and hardly any standard procedures can be followed."* (p. 405). Internal reliability is aimed towards whether the observers within a research agree about their findings and if an independent researcher would come to the same conclusion (Bryman, 2013). It is difficult to reach internal reliability for this study, because like mentioned before each research has their own point of view which regards to the topic of concern. Therefore, reaching the same conclusion could be rather difficult.

Bryman (2013) states that *"the most important criterion of research is validity and that validity is concerned with the integrity of the conclusions that are generated from a piece of research"* (p.47). There are different types of validity: measurement validity, internal validity, external validity and ecological validity. *Measurement* validity is mainly used within studies with a quantitative setting with some measures. This study is of qualitative research and does not include

any specific measurement and therefore could be regarded as relevant. *Internal* validity refers to the causality of two or more variables and whether there is a good match between the observations and the ideas researchers develop (Bryman, 2013). Internal validity does not directly apply to this study, because no causal relation between two or more variables are tested. The researcher does not test if different approaches to maturity models have different outcomes for maturity levels. *External* validity refers to the generalizability of the findings of the study beyond the specific research context (Bryman, 2013). Achieving generalizability is no easy task within qualitative research, because of the relatively small sample sizes within the study. The people that are generally interviewed within this type of research are not representative of the whole population. Additionally, Bryman (2013) mentions that the findings within qualitative research are to generalize to theory rather than to populations. *Ecological* validity refers to the applicability of the findings of a study to people's everyday lifestyle. This type of validity is discussable, because each researcher may look differently at what exactly influences an individual's life. For this topic, the firm could drastically change their procedures or processes to improve data quality management within the firm. And by looking at this scenario, employees could be affected towards their lifestyle, because of more workload within the firm. Therefore, ecological validity could be applicable in some sense for this study.

3.3 Collecting data

In qualitative research, several main research methods are applied, such as ethnography/participant observation, interviews and focus groups. For this study, interviews seem to be a suitable method because they allow open ended questions to a small sample and make it possible to gather insights or experiences of individuals. Within qualitative research two types of interviews are generally used; semi-structured and unstructured. In comparison to quantitative research, interviews within qualitative research are less structured and have more flexibility, the researcher can deviate from any schedule and gathering detailed answers is possible (Bryman, 2013).

For this study, the semi-structured interview type seems to be a suitable method because the flexibility of semi-structured interviews makes it possible for the interviewer to ask in depth questions (Miles & Gilbert, 2005). In addition, the flexibility of this research enables the researcher to gather in depth answers to questions by asking follow up questions. A structured interview would restrict the researcher to go deeper into certain topics while conducting the interview. Furthermore,

the questions within the interview guide could be asked in any order assumed most suitable for the current situation. Additionally, questions not listed within the interview guide could be asked based on the answers picked up by the researcher. No questions will be neglected for no valid reason.

The available literature was studied intensively, before conducting any interviews. The insights gained from the literature were used as inspiration to define the questions for the interviews. In addition, an informed consent with the principles of ethics in mind was constructed which can be found in Appendix II. This informed consent is discussed with the interviewee which in turn needs to agree to these in order for the interview to take place. The informed consent acts as additional information about the research in order for the interviewee to make an informed decision about whether or not they wish to participate in this research (Bryman, 2013).

An interview guide was developed to act as a tool in guiding the interviews and to ensure that all the planned subjects for this study were covered. Possible misunderstandings about the questions were taken into consideration, because of the different backgrounds of the interviewees. Accordingly, clarifying questions were asked whenever deemed necessary the check if the interviewee understood the questions and if their given answer matches the question correctly. The interview guide can be found in Appendix IV.

For the purpose of improving the initial maturity model and scorecard developed within section 2.5 of this paper, nine respondents were interviewed as shown in table 9. To ensure anonymity, the company names were replaced by incremental alphabetical letters, based on the order of the interviews. During the course of these interviews information with regards to data quality management was collected and the maturity model and scorecard were tested as a form of case study. Each interviewee, if willing, applied the maturity model to their organization. This ensured that the researcher could get practical insights about the applicability of the model. The respondents for this research were selected through purposive sampling. In this sampling approach, the goal of the researcher is to sample participants in a strategic way, so that the participants are relevant to the research question of this study (Bryman, 2013). In addition, convenience sampling is also applied within this study. I have been in the field of business performance and have gained several personal contacts within this process, which could be of value for this study. Finally, snowball sampling is applied within the end of each interview, in order to identify potential interviewees. An interview request letter was send to the individuals, which can be found in Appendix III. The respondents were chosen based on the following criteria: (1) the respondent

should have a role within the firm that is associated with using, arranging and analyzing the data, (2) the firm of the respondent should be dealing with critical data among their business processes, (3) the respondents should not solely be the users of the data but also the ones that maintain them and (4) the respondent should have some level of experience in the topic of data quality management and should not be inexperienced.

To limit bias, highly knowledgeable respondents with regards to data quality management were selected (Bryman, 2013). The respondents were found through different channels. Some were personal contacts of the researcher and some were approached through the networks of these contacts. Others were contacted through LinkedIn, because the occupation of individuals can be found. The Dutch language was preferred for conducting the interviews, however if needed otherwise, the English language was preferred. All the interviews were voice recorded and promised anonymity. The interviewees gave permission to record, by agreeing to the informed consent defined within Appendix II. Six interviews took place at the interviewees workplaces, two were invited to my personal workspace and one was done with the help of FaceTime. Each interview took approximately between 40 and 55 minutes. The number of interviewees was decided based on the time and costs available for this research. This number excludes any dropouts that have occurred by for example withdrawing from this research because of personal reasons. The interviews were stopped because enough differentiated results were gathered with regards to the maturity model and the interviews were not providing any new and different insights.

Table 9: Respondents for this study.

Company	Occupation	Company characteristic	Company size
A	Business Developer	Insurance agency	50 - 100
B	Software Developer	Process digitizing	1 - 10
C	Business Intelligence Consultant	Business solutions	1 - 10
D	Business Development Manager	IT software development	150 - 200
E	Business Intelligence Consultant	IT consultancy	10 - 25
F	Managing Director	Full-service Internet agency	50 - 100
G	Chief Financial Officer (CFO)	Government agency	50 - 100
H	Portfolio manager	Government agency	500+
I	Consultant(s)	IT consultancy	25 - 50

3.4 Analysis

Before conducting analysis on the data collected, the interviews were transcribed. The interviews were transcribed without neglecting any important data and usefully summarized. In order to provide more detailed insights in how the transcribed interviews were turned into analyzable contents for this study, the following actions were taken: Firstly, the topics of each set of questions within the interview guide were grouped as a theme. Consequently, the themes were included in a coding agenda, which can be seen in Appendix V. In defining the columns for the coding agenda, the qualitative content analysis approach of Mayring (2000) was used. Second, after creating the coding agenda, a category system was defined wherein the statements of the interviewee were included. Some statements were directly quoted, but most were summarized and written in the researchers own words. For instance, company B mentioned that *"We don't really use any procedures or standards; mostly common sense is applied."*. Statements and answers on the questions within the interviews are included in the category system. The category system can be found in Appendix VI. Lastly, themes that provided additional value to the topic of this study were analyzed and discussed in more detail (see chapter 4). The collected statements, opinions and events were interchanged and combined within the different themes. Data that seemed less important was completely left out.

During the interviews data has been collected as complete as possible. The collected data has been discussed within the interviews, to verify the answers. However, the transcribed interviews were not presented to the interviewees. There was no particular reason to do so, if additional information was needed the interviewee could be contacted, but was not given a summary of the interview.

4. Analysis

Within this chapter, the interviews and their objectives are discussed (4.1). In addition, the interviews are analyzed to provide a practical view on data quality management (4.2). Furthermore, the points of improvement for the maturity model and scorecard gathered from the case studies are presented (4.3).

4.1 Interviews

The interviews were conducted at nine different firms, some firms are active in a similar field, but none of the firms are correlated with each other. Some firms of similar type were interviewed, so that differences between the firms could be gathered. The interviewed firms ranged from large to small firms. This variety was applied so that the applicability of the maturity model and scorecard could be tested in firms of different size and if this would have some effects. By identifying possible effects, the model and scorecard could be adjusted for better fit in practice, so that the size of a firm would not negatively or positively affect the maturity level. Table 10 illustrates the interviewed firms in more detail.

Table 10: Detailed overview of respondents.

Company	Company size	Located in	Contact type	Length of interview	Number of participants
A	50 - 100	The Netherlands	face-to-face	37 minutes	2
B	1 - 10	The Netherlands	face-to-face	43 minutes	2
C	1 - 10	The Netherlands	face-to-face	57 minutes	2
D	150 - 200	The Netherlands	face-to-face	42 minutes	2
E	10 - 25	The Netherlands	face-to-face	72 minutes	2
F	50 - 100	The Netherlands	face-to-face	45 minutes	2
G	50 - 100	Turkey	facetime	106 minutes	2
H	500+	The Netherlands	face-to-face	45 minutes	2
I	25 - 50	The Netherlands	face-to-face	98 minutes	3

The first objective of the interviews was to gain insights in the principles firms applied for data quality management within practice, so that a comparison with theory could be made. For instance, discussions with regards to experiences of poor data quality were reviewed. Furthermore, the interviews confirmed any assumptions the researcher had with regards to data quality management within practice. The second objective was to receive feedback on the maturity model and scorecard of this study. The feedback in turn lead to improvements for better fit within practice. During each interview fitting questions regarding the topic of interest were presented, that would reflect back to the maturity model. Furthermore, the maturity model and scorecard were tested accordingly by performing a case study, on the organization of the respondent that was interviewed. Additionally, conclusions could be drafted for what was working well and bad within the model and scorecard. This was also the overall objective of the interviews; to enhance the applicability of the model and scorecard of this study, this data collection of different perspective was successful.

4.2 Interview results

In this section, the themes used within the interviews will be discussed. The themes consist of several questions; however, some will be combined in order to decrease the length of this section. The topics *data quality within practice* (4.2.1) and *applicability of maturity model within practice* (4.2.2) are discussed in detail and statements from the respondents are stated. Not all the data that are collected with the help of the interviews are discussed in detail, an overview of all can be found in Appendix VI.

4.2.1 Data quality within practice

The focus of this paragraph are the actions that firms take to manage their data and possible steps they take for improving data quality. In addition, any principles or standards firms are using are treated. For instance, **company A** mentioned that *"We don't have any principles or standards defined. Our employees just use their common sense to perform their daily work processes"*. This statement was reoccurring within other interviews, which in turn could be a cause for poor data quality. **Company C** stated that *"Our customers were not paying that much attention to managing data quality. This proved when we did analysis on their data and found the same city registered with three different names."* This company in turn changed the steps for registration and only provided the firm with pre-filled fields for this kind of situations. This ensured that the employees

cannot freely fill-in the data, but rather had to select the data from a pre-filled dropdown. Nevertheless, changes like these could improve data quality. Furthermore, **company E** noted that *"We don't have any quality procedures, only responsibilities of what needs to be done. Our firm is not internally giving additional attention to managing data quality."*. This confirms the assumption of the researcher that not many additional attention is given for managing data within practice, but rather tasks are just executed. In addition, firms assume that data management is not really important until it proves useful. Firms do not want to spend time in improving their data quality management processes, but in the long run this proves otherwise. Nonetheless, this does not mean that all firms are not giving any attention to data quality management. More and more firms are seeing the benefits of good data quality management. For instance, **company H** mentioned that they are gradually increasing the attention given to good data quality. Nowadays, more techniques and software are available to manage the data within the firm and more firms are starting to apply these so that mistakes within their data can be avoided.

Reasons for collecting data, could be for instance to provide input for management; create 'monthly' reports; aid in decision making and to track the business processes of the firm. If the users don't collect the data, then most of the time the systems provide insights within specific software. However, in most cases the firms have to arrange these data and provide input for the different kinds of software. For example, **company B** was providing services to their customers and with this service the customers were using the databases of the firm. The company could arrange the data and their sources accordingly, because they eventually maintained it. The company mentioned *"We collect our data from our own MYSQL-databases and make connections by performing queries."*. This is not always the case, when you provide services to your customers, then you can arrange the data correctly and present them with regards to your own standards. However, when you need to rearrange the customers' data sources, then you can experience some difficulties, like mentioned by **company E** *"We have difficulties in combining multiple information sources, because the data is different within every system. For example, within system A, credit numbers are saved as 103 etcetera, but within system B, these are noted as 103-1. This ensures problems in making comparisons and combining the data sources."*

Many firms still work with Excel, because it is easy to access and everyone has it installed on their working stations, **company C** mentioned that *"Firms are working too much with Excel. There are many lists that need to be imported to the systems, and additional columns need to be*

added to their databases.". Whereas, **company A** and **company F** note that firms are relying too much on manual work and do not put any effort in automatization. In addition, **company F** added *"Employees don't take any action for making improvements, they just repeat their work over and over again, because they are accustomed to it."*. This shows that little initiative is taken by employees and maybe this could be one of the causes of poor data quality management.

When comparing the findings within practice to the theory, the conclusion can be made that these are in close connection with each other. The reason for this is that challenges of data quality match with practice, for example data integration with different systems is stated as a challenge of data quality by Cai & Zhu (2015) and this challenge repeats itself within practice. In addition, the problem that extra time is needed to reconcile data caused by poor data quality mentioned by Eckerson (2002) has also been noted by the respondents within the interviews. These show that the theory of data quality management and the reality within practice are in close connection to each other and maybe firms should start considering these theories. Next, the applicability of the maturity model and scorecard are discussed.

4.2.2 Applicability of maturity model within practice

The maturity model was presented to and discussed with each interviewee. These discussions ensured that valuable new information could be gathered in order to improve the maturity model of this research. Furthermore, the scorecard was applied to the firm of the interviewee which indicated the maturity level of the corresponding firm. With the application of the scorecard and discussion of the maturity model, points for improvements were mentioned, these are discussed in depth in this paragraph.

Firstly, it was mentioned if a similar model was used by the interviewees within their field of practice. Most of the interviewees mentioned that they did not use any similar maturity model, at least not with regards to data quality management. However, one interviewee told that he has used a similar model and also was included within the development of this model. The model was called the Prince2 Maturity Model, where a set of requirements were defined within. The performance of the firm was defined by the extent of which the firms followed these. This model was created digitally, and provided immediate feedback on the level of maturity.

Secondly, methods, techniques or software used for data management were treated. Within this subject, the same kind of answers were gathered, firms were not specifically using any for data

quality management. The firms were more using software for visualizing the data, but not to maintain it. One of the firms is using a document management system with regards to arranging all their data, but does not follow any other data management practices. Furthermore, **company C** mentioned *"We are not using any models to manage our data, it is mostly done with common sense."*. This shows again that no procedures, but more common sense is used within practice.

Thirdly, the maturity model and the scorecard were discussed. At this part of the interview, the maturity model and the maturity levels were explained and afterwards the scorecard was filled in by every firm, in order to gather in depth information for improvement. The scorecard and points of improvement gathered for each of the factors will be discussed first and the applicability of the maturity model are elaborated afterwards. Table 11 presents the factors and corresponding remarks. The interviewed companies are only mentioned by their letters.

Table 11: Feedback of each factor in the scorecard.

Factors	**Remarks**
The maintenance of applications is person dependent.	- The scores are not really that clear. What is a division, team or enterprise? (A, D) - The scores are not applicable for small firms, they could never reach a division/enterprise level (F). - Scores are not really applicable, because within practice applications are maintained by individuals and not whole divisions. Maybe common practices as scores could be applied (H).
The data provided to the systems is mainly done by employees.	- This factor should be reworded towards automatization, the current description does not say much. You should be focused on the limitation of manual entry (I).
The data within applications is managed by applying updates.	- Updates are really application dependent, you cannot say that you do them within a timeframe (H).
Standards and procedures are defined within the firms, that can be repeated in the future.	- Small firms do not always define standards and procedures, because it is a time-consuming task (C).
New gained data is stored within the systems of the firm, so that incomplete data can be avoided.	- Instead of percentages, maybe mention a range from not stored to stored (E).

The data model of the firm is updated when new metadata arises.	- Some elaboration needs to be given for metadata, not known by non-technical individuals (A, F).
Corporate data within the firm is standardized (standz.)	- Not directly know what standardized means (D).
The firm is aiming to continuously improve.	- Not all types of firms aim to continuously improve, but this should not affect this factor (G).
The firm can identify their strengths and weaknesses with regards to data management.	- Difficult to identify the strengths and weaknesses, maybe different scores can be used (B). - The numbers given are not that applicable, maybe percentages could be used instead (F, G).
The main focus of the firm is to reduce waste.	- Difficult to know where you stand, not always aware which data is stored where (B).

As you can see within the table above, many remarks were given for the first factor. This factor was the most discussed factor, because the scores were not representative of reality. In reality, firms do not want their applications to be person dependent, but they cannot assume that the whole organization will know the applications inside-out. Therefore, firms are aiming to define a set of standards and procedures, so that these can be followed when a responsible person leaves or a new employee enters the firm. Therefore, this factor will be adjusted so that it fits within the common practices that a firm aims to retain. In addition, like mentioned by **company E**, numbers will be added to each factor, because it is hard to address a certain factor when discussing it. Overall the scorecard was simple and usable with some additional explanations, **company H** mentioned *"The scorecard is clear and simple to fill in."*, **company A** stated *"The scorecard acts as an easy tool for determining which maturity level is fitting for the firm."* and **company B** noted *"The scorecard is clear; however, some factors could be noted in simpler terms so that every kind of employee can understand them."*.

Furthermore, some additional remarks were made by the interviewees. For instance, **company F** was really focused towards waste management and was trying to apply this within his firm, but did not succeed. He mentioned the following *"We always mention waste management, this needs to happen but it just doesn't. People just don't see the added value."*. The interviewee noted over and over again that employees were not willing to participate in waste management. This shows that some data quality issues could arise, because there is a lack of commitment by the employees in improving the data quality management of the firm.

In addition to the scorecard, **company I** proposed two new factors: (1) Is the firm actively searching for wrong definitions of data? (2) Is the firm focused on removing double entries of data? These factors are a good addition to the scorecard, so that erroneous data can be prevented within the firm by focusing on its removal. Next, the experiences of using the maturity model and scorecard are discussed.

Experiences with the maturity model and scorecard

After some explanation about the maturity model and scorecard, the interviewees were known with its use and principles and could provide their first opinion about it. Specifically, five of the interviewees mentioned the following: *"The maturity model provides a good overview of its different maturity levels, however without further explanations it is hard to know what it exactly represents."*. Nevertheless, the explanations are provided within this research and are presented before the maturity model so each user can be informed by reading these first.

In addition, two interviewees mentioned entirely changing the maturity model and how it is structured currently. In the current version of the maturity model, each level is shown within a column in the table. However, these companies mentioned maybe developing a stairs kind of construction, so that you directly see which is the highest level and what 'steps' you need to take in order to reach the highest maturity level. With this structure, you can see that each step brings you closer to better data quality management. Additionally, the criteria below each maturity level were questioned and it was not really known what the role of those were.

Moreover, the applicability of the maturity model and scorecard for each size of firm was questioned. For instance, small firms and large firms would probably score differently and it would also differ in usage. Some of the factors noted are currently not applicable for small firms, therefore some changes will be made in order to create a better fit for each firm size.

Lastly, a good remark was given to the maturity model. Currently the maturity level of the firm is defined without first thinking about which maturity level is fitting for the firm. Therefore, **company H** proposed to first define the desirable maturity level for the firm. Afterwards, when the desired maturity level has been chosen, the firm can start to fill in the scorecard, so that the current maturity level can be identified. This approach ensures that the firm first thinks about their desired state of maturity, before looking at their current state. Nevertheless, being in the highest maturity level does not directly mean that this is the best level to be at. Some firms have a standard they

need to reach and higher than that could be financially hurting the firm. **Company F** mentioned that *"Models like these can help firms to see in what situation they are situated. We for example can see where we are and where we aim to be. There are some moments, where you think you are doing good, but when using the scorecard, you notice that actually there is little focus given on that factor."*. Moreover, with the help of the model and scorecard, certain points can be discussed, you do not have to 'create' some factors in order to start a discussion about data quality management. **Company I** mentioned that *"Not only can it be used to figure out in which maturity level you are situated, but can also be used as a measuring tool for calculating the amount of time you will lose when starting a new project and with this determine the complexity of the project. We can further point our customers to this maturity model and mention that they need to improve on some points before conducting any analysis on their data."*. This shows that the maturity model and scorecard can be used in many ways. Furthermore, **company H** stated the usefulness of this model and scorecard: *"People always tell that their data is on a high level, but they just give an estimation and do not know it exactly. If someone would ask me, I would also say that we are at 75%, but this does not really say something. With the help of your maturity model and scorecard, you can define in which maturity level you stand and work further based on this estimation."*.

5. Results

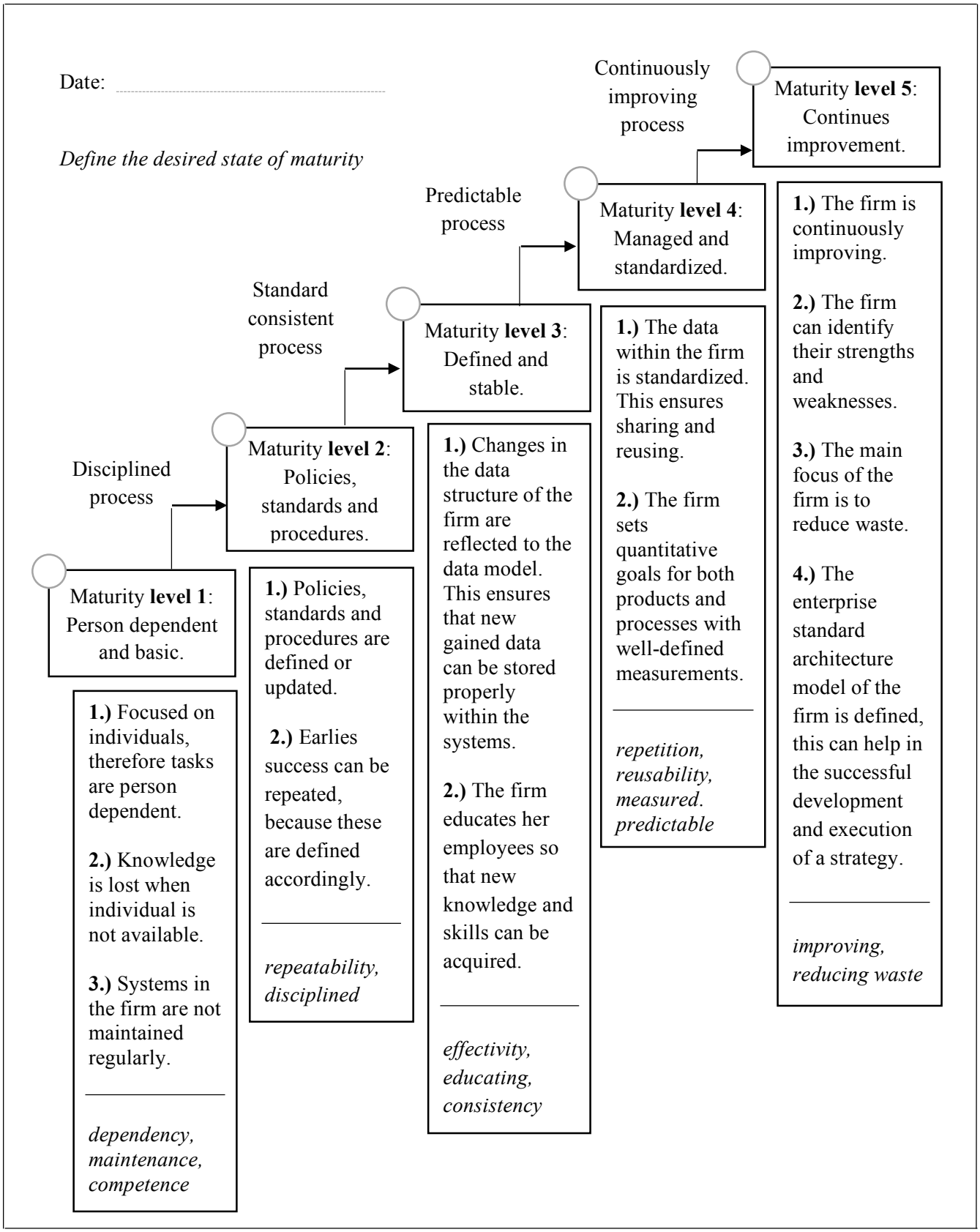
Within this chapter, the data collected from the interviews will be reflected towards the maturity model and scorecard. The improved maturity model and scorecard are presented and discussed in more detail (5.1). Furthermore, the improved maturity model has been tested (5.2) and conclusions are drawn (5.3).

5.1 Improved maturity model and scorecard

The maturity model was providing too much information whereas the user did not directly know where to focus on and what to comprehend from the maturity model. Furthermore, the text within each maturity level was not easily comprehensible, therefore the decision was made to optimize the text presented in the model. Moreover, some simple additions were made, like the date of using the maturity model, because these proved useful when testing the first version of the maturity model within the case studies. Lastly, the whole structure of the maturity model is changed, so that some consequent steps of improvement can be shown more effectively and where one immediately can see which level is the highest.

In order to effectively use the maturity model, instructions for usage are defined. The firm that is using this model should analyze the maturity model and define the desirable state of maturity. Firstly, the user should note the date of usage, so that in the future a comparison can be made with the different maturity models used. Secondly, when the firm has defined a maturity level, the user should mark the desired maturity level with the help of the circles provided on the top left of each maturity level. Lastly, the firm should note any attention points they came up with when analyzing the maturity model. These notes could help in making decisions when scoring the factors within the included scorecard for defining the current state of maturity. Figure 1 shows the new improved maturity model of this study.

Figure 1: Improved maturity model for determining the level of data quality management.



Now that the maturity model is improved and illustrated accordingly, the scorecard can be presented and discussed. For the scorecard, the same structure will be used and some factors will be worded differently, so that the comprehensibility can be increased. Based on the remarks gathered from the interviews, as shown in table 11, some changes will be made. Firstly, each factor has been reconsidered and when needed reformulated. Many feedback was received with regards to the structure of each factors, sometimes it was not clear or was formulated incorrectly. Therefore, some will be changed for better fit for determining the maturity level. Secondly, also the scores for each factor will be given a second thought, in order to create a better relation with the corresponding factor. This is also important, so that firms apart from their size, can also make good decisions when scoring a factor and a factor should not be scored based on the size of the firm. Furthermore, a firm regarding of their size should be able to reach the maturity level preferred and no score should prevent them from reaching the desired maturity level. Nevertheless, this maturity model and scorecard are not fit for use by for example farmers, a bakery and a shoe store. This should be kept in mind, firms that work with critical data within their business processes could apply this maturity model in order to determine the current state of maturity and with this evaluate their level of data quality management.

Firstly, if any notes are made in the maturity model, these should be considered and used in scoring each factor. The scores within the scorecard should be marked with preferably a yellow marker. Secondly, after each factor has been marked accordingly based on the current state of the firm, an indication of the current state of maturity can be given. After filling in the scorecard, the firm should check which maturity level is marked the most. For instance, if maturity level 2 is marked the most, then the current state of maturity is level 2. However, when factors reside within lower levels, than the firm should choose the lower level as current state of maturity, and should not advance to a higher level of maturity before tackling the factors in the lower maturity levels. Furthermore, when maturity levels are equally marked, the lowest maturity level should be preferred and the factors there should be improved before advancing to the next level of maturity. Table 12 illustrates the improved scorecard of this study.

Table 12: Improved scorecard for determining the current state of maturity.

Factors	Maturity **level 1**	Maturity **level 2**	Maturity **level 3**	Maturity **level 4**	Maturity **level 5**
1.) The firm is not dependent on certain individuals, because of common practices defined.	None defined	Roughly 25% defined	Roughly 50% defined	Roughly 75% defined	More than 75% defined
2.) The firm has automated (autom.) their procedures as much as possible.	None autom.	Roughly 25% autom.	Roughly 50% autom.	Roughly 75% autom.	More than 75% autom.
3.) The applications used by the firm are updated when needed, so that issues can be prevented.	Never updated	Yearly updated	Monthly updated	Weekly updated	Always updated
4.) Standards and procedures are defined, so that these can be repeated in the future.	None defined	Roughly 25% defined	Roughly 50% defined	Roughly 75% defined	More than 75% defined
5.) New gained data is stored within the systems, so that incomplete data can be avoided.	Not stored	Roughly 25% stored	Roughly 50% stored	Roughly 75% stored	More than 75% stored
6.) The data model of the firm is updated when new metadata arises. *Metadata is information about a certain set of data.*	Never updated	Yearly updated	Monthly updated	Weekly updated	Always updated
7.) The firm has standardized (standz.) their procedures for storing data.	Not standz.	Roughly 25% standz.	Roughly 50% standz.	Roughly 75% standz.	More than 75% standz.
8.) The firm is focused on continuously improving.	Never improving	Rarely improving	Often improving	Very often improving	Always improving
9.) The firm can identify their strengths and weaknesses with regards to data management.	None identified	Roughly 25% identified	Roughly 50% identified	Roughly 75% identified	More than 75% identified
10.) The main focus of the firm is to reduce waste of data and take actions accordingly.	Never focused	Yearly focused	Monthly focused	Weekly focused	Always focused
11.) The firm is actively searching for any wrong definitions of data.	Not searching	Rarely searching	Often searching	Actively searching	Always searching
12.) The firm is focused on removing double entries of data	Not focused	Rarely focused	Often focused	Actively focused	Always focused

5.2 Testing the improved maturity model and scorecard

For the post-test of the model and scorecard, two participants were approached. The first participant was not familiar with the maturity model and scorecard, and the second had used the first version of the maturity model before. The post-test between the two participants showed some clear improvements with regards to the first version of the model.

The first participant could use the maturity model and scorecard without any additional clarifications needed. However, some points within the maturity model and scorecard were not self-explanatory. Specifically, the short description of maturity level 3 was showing some signs of uncertainties. Whereas, it was not known what it meant by *"This ensures that new data enters the systems."*, this statement has been reformulated and verified with this participant and corrected accordingly. Furthermore, it was not really known without any further explanations how the participant would define the maturity level after filling in the scorecard. However, the instructions for the maturity model and scorecard clarify this, but this participant had not seen this before and therefore was unclear about it. The first participant did however not mention any other misunderstandings within the model and scorecard.

The second participant was already known with the model, but had some remarks regarding the definition of each maturity level. For a side note, this participant had also not read any mentioned instructions and used the scorecard regardless of this. Firstly, the participant had difficulties in selecting a desired maturity level, because he wanted to know the advantages and disadvantages of each maturity level before making any decisions. Information about each maturity level can be found in section 2.4, but these were not presented to this participant in order to test this model without any information provided. Lastly, the participant provided some good remarks and therefore, some statements were worded differently because of these. Furthermore, no uncertainties were established, and the post-test of the maturity model and scorecard was finished successfully.

5.3 Conclusion

The maturity model and supporting scorecard can be used for determining the current state of maturity. By determining the level of maturity, the firm can retrieve some indications about where additional attention is needed to improve data quality management. The maturity model includes some aspects of data quality management, like mentioned within the theory. For instance, the databases of the firm could be improved to ensure efficient data sharing within the firm.

The contribution the maturity model and supporting scorecard make for practice, are the self-consciousness that it creates for firms on the topic of data quality management. The model and scorecard should always be used in combination of each other, so that the scorecard can support any decisions made when using the model. For instance, a firm may consider its current state of maturity as level 4, but with the supporting scorecard they could get a better indication at reality. In addition, the firm can use these factors as a starting point in improving certain procedures in order to reach a higher level of data quality management. The firms using the maturity model and scorecard should not neglect the fact that some other variables may cause poor data quality within their firm and may not be covered by this study.

However, the improved maturity model and supporting scorecard revealed some shortcomings. Namely, depending on the business of the firm using the maturity model, the importance of certain factors in the scorecard may differ and the highest level of maturity may not be the best choice. Since, the maturity model and scorecard do not give an overall score, as requested by some users of the maturity model, it may show some shortcomings with regards to exactly defining the maturity level. However, due to the time restrictions and the scope of this study, it was not possible to develop a scorecard with scores from 1 to 10 for each factor. Specifically, not enough data could be collected for developing this kind of scorecard. In addition, for developing this kind of scorecard a lot more factors need to be defined which could potentially consist of some sub-factors, so that more topics of data quality management can be covered.

6. Conclusion and discussion

Within this chapter, a conclusion (6.1) is presented and the findings are discussed (6.2). Furthermore, the contribution to science and practice are treated (6.3). Lastly, the limitations and some guidelines for future research are included (6.4).

6.1 Conclusion

The aim of this study was to create a maturity model and supporting scorecard to determine the current state of maturity with regards to data quality management. The literature search aided in identifying the maturity model for data management by Ryu et al. (2006). However, this model showed some shortcomings with regards to determining the current state of maturity and no supporting tool was available in providing an indication of maturity. Furthermore, the literature discussed data quality and data management separately. Since, this research was focused on data quality management, these were combined and analyzed. This ensured that the data quality characteristics could be identified, for example availability, usability and reliability. Next, interviews were conducted to gain a deeper understanding in data quality management principles firms applied within practice. In addition, the maturity model and scorecard were tested with the help of case studies, by application of these to the firm of the interviewee. These case studies aided in designing an optimized maturity model and supporting scorecard. In addition, the interviews showed that the theory was in close connection with reality. For instance, that standards and procedures were simply not followed by firms which in turn caused poor data quality. The reasons for not following these were mainly due to the amount of time these consumed and employees were not easily willing to stop doing it their own 'easy' way. Based on the findings of this study, the following central research questions which was assigned to this study could be answered:

"*What is the usefulness of a maturity model for determining the state of data quality management in organizations?"*.

The maturity model and supporting scorecard, which are illustrated in figure 1 and table 12, can be used by firms to determine their current and desirable state of maturity. The maturity model mentions some criteria for each maturity level, so that the firm using the model can categorize themselves based on these. The maturity model and supporting scorecard indicate the level of data

quality management within a firm, by filling in the scorecard the firm can get some indication in the points that need some additional attention and what the firm should improve. The scorecard indicates the current state of maturity with the help of same factors that are defined based on the theory of data quality management. In this way, the maturity model shows its usefulness by helping firms in determining points of improvement and in so make fitting choices within their business processes, so in turn data quality management can be improved.

6.2 Discussion

This study focused on maturity models and the usefulness of these within practice. During this study, empirical evidence was found that poor data quality management may have some effect on the performance of a firm. So, the results of our study suggest that firms should pay more attention towards data quality management principles. The findings of these studies correspond with previous research, which have shown that a maturity model could be applied to determine the current or desirable state of maturity and evaluate data quality within a firm (Paulk et al., 1993; Ryu et al., 2006). The key contribution that this study makes, is that it enables firms to determine their state of maturity and provides some factors that indicate issues with regards to data quality management. This study showed that maturity models are well applicable within practice and could be used as a powerful tool for optimizing the business processes of the firm.

As a result of this study, the maturity model for improving data quality management was developed, which helps firms in determining their current or desirable state of maturity and evaluates data quality within the firm. For developing the maturity model and scorecard, the capability maturity model of Paulk et al. (1993) and maturity model of Ryu et al. (2006) were used. Maturity models are widely applied in the field of research, however this study identified its shortcoming. Namely, the maturity model focuses on a certain topic, e.g. data quality management. While, other factors that could be of influence are not discussed. For instance, influence of competition and environmental factors, are ignored. Nevertheless, these could also have some effect on the maturity of an organization. When evaluating the maturity model and supporting scorecard developed in this study, some points of critique can be given. Firstly, the descriptions of each maturity level may provide insufficient information in some cases. For instance, the firm could categorize themselves within several maturity levels, but may have difficulties in choosing the correct one because certain information is missing. Secondly, the scorecard is a valuable tool in

determining the current state of maturity, but the objectivity of the factors can be questioned because percentages do not directly reflect reality. The scores for the factors are not specific in a way that firms can define an exact value, but rather provides an estimation. Since the scorecard does not give an overall score which indicates the current maturity level, it can create obscurities for someone who uses the scorecard. For this reason, it might have been better to make the factors more objective, and that each factor should be scored by the user, in this way a total overall score can be created which in turn can result in an overall level of maturity. However, this was not possible due to the time restrictions and scope of this research. Namely, not enough data could be gathered for developing a scorecard that provides an overall score. Furthermore, it should be noted that the maturity model and scorecard are not applicable in any field of business, e.g. bakeries and an ice cream shop. Nevertheless, the maturity model and supporting scorecard remain a useful tool which helps firm in determining the current or desirable state of maturity and evaluate the data quality within the firm.

6.3 Contribution

This study contributes to science, with the introduction of a maturity model with supporting scorecard. There is a lot of research done on maturity models, however in academic literature the focus is not given on a supporting scorecard, which can indicate the level of maturity of a firm. So, this study contributes to science by introducing a different approach the determining the current or desirable state of maturity, and with this evaluate data quality within the firm. Moreover, this study confirmed some challenges within practice with regards to data quality management and showed that these are in close connection with theory. In this way, the study acknowledges the need for more research on the topic of data quality management, so that firms tackle their issues and prevent any mistakes in their data. In doing so, I hope other researchers show interest for this topic.

Next to the contribution to science, this study also contributes to practice. The maturity model and supporting scorecard developed within this research, help firms in identifying their current or desirable state of maturity and aid in evaluating data quality management. The results of this study act as a tool for firms to make improvement within their business processes which in turn should lead to improved data quality management. Besides, determining the maturity level and evaluate data quality, the model and scorecard also help in providing useful information about the firm and their processes, so that a different perspective of the firm can be created.

6.3 Limitations and future research

The same as each study, this thesis also has its limitations. Like mentioned by Bryman (2013) achieving generalizability is no easy task within qualitative research, because of the small sample sizes. The results of this study can only be generalized to a certain extent. The findings of this study can be generalized to firms that work with critical data within their business processes, for example a logistics company. Within the scope of this study, only nine respondents were interviewed, which influences the external validity. Furthermore, case studies were performed with only one person of the firm, which in turn single response bias might have occurred. However, to overcome these shortcomings, different types of data (e.g. interviews, company documents) were used, which contributed to the validity of this research. Lastly, to achieve generalization different types of firms were approached, because of the wide applicability of the model not every type of firm was covered. Therefore, it should be noted that these firms do not directly represent the population.

According to Bryman (2013), external reliability refers to the degree to which the findings of a study can be replicated, because generalization is not reached without any effort. Therefore, in order to solve the problems with regards to generalization, a future attempt should be made including more diverse companies besides the ones approached in this study. Furthermore, the type of study could be changed to a quantitative approach (e.g. surveys), this might ensure larger amounts of data collection and tackle the problem of generalization. Lastly, after replication has been achieved, the findings of this study might be more acceptable to generalization.

Additionally, due to the scope of this study, it was not possible to add more factors to the scorecard and with this gather more input by the firms in order to reach the current state of maturity. If more factors were included within the scorecard, it would have helped to indicate the state of maturity more precisely. Therefore, a possibility for future research is to include more factors in the scorecard. In addition, the maturity model of this study could be critically assessed, so that each maturity level provides sufficient information for categorization by firms. This could be achieved by extensively testing the maturity model within practice and including aspects from practice.

Also, the maturity model and scorecard could be extended by providing these with an overall score. The current approach probably needs some adjustments, so that a precise overview can be given to when a firm reaches a certain maturity level (e.g. a firm scoring between 25 - 40 is probably located in maturity level 2). After including scores within the maturity model and supporting scorecard, these should be tested extensively in practice. For instance, confirming

questions could be applied, instead of providing estimations (e.g. Are you applying and data management principles?). In this way, the researcher could add scores behind each question, rate the firm based on these scores and count these for determining the current state of maturity. By doing so, the objectivity of the factors can also be addressed. Therefore, it is interesting for future research to investigate the possibilities of including an overall score.

Appendices

Appendix I: Time schedule

2016

During the end of October and the start of November, no progress has been made with regards to this thesis.

The loss of my MacBook was the main reason.

	September		October			November			December
	23	28	2	6	26	10	17	25	16
Optimizing thesis, solely focusing on data quality management.									
Include data quality standards within thesis.									
Analyze additional theories with regards to continues improvement.									
Define an interview guide for data collection.									
Methodology adjustments and finalization.									
Make adjustments to thesis based on feedback from meeting of 16 November 2016.									
Start developing own maturity model for improving data quality management.									

2017

During the period of January 13 till February 24, no progress has been made with regards to this thesis.

Certain personal events were the cause.

	February	March					April			
	27	3	6	8	20	27	7	12	19	28
Finalize interview guide.										
Schedule interviews with respondents.										
Interviews										
Transcribing & Coding of data collected.										
Analysis of data collected.										
Send report to supervisors, for feedback.										
Start with writing conclusions of thesis.										

Appendix II: Interview agreement (PacificWorlds, 2016).

Informed consent:

Master thesis of Onur Kirikoglu
o.kirikoglu@student.utwente.nl

Thesis description:

This thesis focusses on improving data quality management within firms by developing a maturity model. This study treats the following central research question: "*How can firms determine the quality of their data and take actions for possible improvement?"*. To provide an answer to this research question, literature based of data quality and data management are treated.

Procedure and risks:

We would like to record the interview, if permission is given to record, the recordings will be used to write information within this thesis. The interview will be recorded only with the interviewees written consent, and preferably no personal identifiers should be used within the interview, in order to ensure anonymity. The interviewee is free in deciding not to answer any question, or to stop the interview at any given time. The recordings will become property of this thesis, after the finalization of this thesis the recordings will be destroyed.

The recordings will be saved until the end of this thesis and will be kept anonymous, without any reference to the identity of the human participant. The identity of the participants will further be concealed within any reports written from the results of the interviews.

There are no known risks associated with human participation within this study.

Benefits:

We hope that the results of this thesis will benefit the field of data quality and data management and provide greater insights in this area. We aim to be able to improve data quality management within firms by applying the maturity model created within this study.

Cost compensation:

The interviewees whom participate within this study will receive no payments or will have no payment for participation.

Confidentiality:

All the information gathered during this study will be kept strictly confidential until the interviewee signs a release waiver. No publications of reports for this thesis will identify the interviewees without their signed permissions and after review of the written reports. If you agree to join this thesis, please fill in the page below:

Informed consent for interviews

Master thesis of Onur Kirikoglu

I, ……. agree to be interviewed for the master thesis of Onur Kirikoglu which is being produced by Onur Kirikoglu of the University of Twente.

I approve that I have been told of the confidentiality of information collected for this thesis and the anonymity for my participation; that I have been given suitable answers to my questions with regards to the thesis procedures and other matters; and that I have been advised that I am free to withdraw my consent and to discontinue participation in this thesis at any given time without any prejudgment.

I agree to participate in one or more electronically recorded interviews for this thesis. I understand that the interviews and related materials will be kept completely anonymous, and that the results of this thesis may be published in an academic journal or book.

I agree that any information obtained from this research may be used in any way suitable for this thesis.

Date Signature of interviewee

…… …….

If you cannot obtain suitable answers to your questions or have comments or complaints about your treatment in this thesis, please contact:

o.kirikoglu@student.utwente.nl
University of Twente

Final consent form

Master Thesis of Onur Kirikoglu

Dear participant:

This form if filled in gives the interviewer the final approval to use the information gained from your interview in the master thesis of Onur Kirikoglu. A draft of any material created will be presented to the interviewee for review, correction or modification. The interviewee may grant rights to use the materials created within this thesis.

I, ……, hereby grant the right to use information from the recordings and or notes taken in interviews of me, to Onur Kirikoglu. I understand that the interview recordings will be kept by the interviewer and be property of this study until the finalization of this thesis and the information obtained from the interviews may be used in materials that could eventually be available to the general public.

Date	Signature of interviewee
……	…….
Date	Signature of interviewer
……	…….

Appendix III: Interview request letter (English and Dutch).

Subject:

Request to participate in research about data quality management.

Contents:

Dear ____:

I am a student at the University of Twente following the master program of Business Administration. I am currently conducting my research in the field of data quality management, wherein I am developing a maturity model to improve data quality management. I found/received your name from/through _____ and I would be really thankful if I could meet with you.

I would like to interview you in person and hope that you are available during the week of _____. I have a couple of questions with regards to data quality management within your firm/organization. The interview will take between 30-45 minutes of your time. I appreciate the time you took to read my request.

I hope you can provide me with an answer if you are available and would like to participate in my research. If you would like any additional information, feel fry to message me. Thank you so much and I am looking forward to meeting you.

Sincerely,

Onur Kirikoglu
06 48 12 34 87
o.kirikoglu@student.utwente.nl

Onderwerp:

Verzoek voor deelname aan onderzoek omtrent data kwaliteitsmanagement.

Inhoud:

Beste ____:

Ik ben een student aan de Universiteit Twente, ik volg de master Business Administration. Ik ben op dit moment bezig met mijn afstudeeronderzoek op het gebied van data kwaliteitsmanagement, waarin ik een maturity model ontwikkel voor het optimaliseren van data kwaliteitsmanagement. Ik heb u gevonden via _____ en zou u heel graag willen ontmoeten.

Ik zou u graag willen interviewen en hoop dat u beschikbaar bent in de week van _____. Ik heb een aantal vragen op het gebied van data kwaliteitsmanagement binnen uw organisatie. Het interview zal 30 à 45 minuten in beslag nemen. Ik waardeer de tijd die u heeft genomen om mijn verzoek te lezen.

Ik hoop dat ik een reactie kan ontvangen over uw interesse in de deelname voor mijn onderzoek. Als u verder nog vragen hebt, kunt u mij gerust een bericht sturen. Bedankt en ik kijk uit naar onze ontmoeting.

Met vriendelijke groet,

Onur Kirikoglu
06 48 12 34 87
o.kirikoglu@student.utwente.nl

Appendix IV: Interview guide (English and Dutch)

*Agreement to informed consent.

*The interviews are planned to take around 45-60 minutes.

*Ask permission to record the interview.

*Anonymity of the interviewee will be protected.

Introducing research topic

- *Explaining the topic and focus of my study.*

Before starting the interview

- Could you describe your firm in detail and the services they provide?
- What is your role within the firm and how long have you been working for your firm?

Starting the interview

- The focus of my study is on data quality management. What kind/types of data is your firm working/dealing with? For example, financial data.
- How is your firm storing data, documents, databases or both? *Maybe some experiences in your line of work on how ways that data is administrated?*
- How would you *define* the terms 'data quality' and 'data management'?
- Does your firm pay additional attention to managing data quality?
- Are there any principles or standards your firm follows for enhancing the quality of data? For example, maybe you are using some data management methods.

Data quality management

- What are your thoughts on data quality management, and what is the ideal situation?
- Do you think that data quality management within organizations is an issue? Is it difficult to manage data and does it challenge the firm?
- Have you had any experiences with data not meeting your requirements? For example, that you received erroneous data? If so, what were the consequences and what were the actions that you took to solve these?

- Data quality has a certain set of characteristics to consider, are these known for you?
- *The characteristics are: Free-of-Error, Fitness, Comprehensiveness, Accuracy, Completeness, Up-to-dateness, Applicability, Repetition, Reusability, Discrepancy and Timeliness.* Now that you heard these characteristics, are these applied within your organization? If so, how specific?

Data collection and visualization

- How is data collected within your firm and do you experience any difficulties with these?
- Do you use the collected data to make visualizations, for creating reports or to justify to management? If so, how do you create these visualizations?
- Are you satisfied with the current methods for data collection? *Why?*

Maturity model to improve data quality management

- *Introduce developed maturity model and discuss this model.*
- What are your thoughts and opinions about this maturity model? Have you ever used a similar model like this? If so, how did you approach in applying this model within your firm? How would you use this model?
- Are you using any methods, techniques or software to manage data within the firm? If so, which ones and what are the procedures to using these?
- *Discuss developed scorecard and ask for opinion. Test scorecard usage.*
- **Would you use this model and in which situation?**

Closing the interview

*Check if every question is treated.

*Ask for additional documentation if possible.

*Thank participant!

*Goedkeuring van de informed consent.

*De interviews zullen ongeveer 45-60 minuten in beslag nemen.

*Toestemming vragen voor het opnemen van het interview.

*Anonimiteit van de geïnterviewde wordt beschermd.

Introductie van onderzoek

- *Toelichten van onderwerp en focus van mijn studie.*

Voor het starten van het interview

- Zou je jouw bedrijf en haar diensten kunnen beschrijven?
- Wat is jouw functie binnen het bedrijf en hoe lang ben je al werkzaam binnen het bedrijf?

Starten van het interview

- De focus van mijn studie is data kwaliteitsmanagement. Met wat voor soort/types of data wordt er binnen het bedrijf gewerkt? Bijvoorbeeld, financiële data.
- How wordt de data binnen het bedrijf opgeslagen, documenten, databases of beide? *Misschien voorbeelden van manieren hoe data wordt opgeslagen?*
- Hoe zou jij de termen 'datakwaliteit' en 'datamanagement' definiëren?
- Wordt er binnen de organisatie aandacht gegeven aan het managen van datakwaliteit?
- Zijn er enige procedures of standaarden die jouw organisatie volgt voor het verbeteren datakwaliteit? Bijvoorbeeld, door het gebruik van data managementmethoden.

Data kwaliteitsmanagement

- Wat zijn jouw gedachten over data kwaliteitsmanagement, wat is de ideale situatie?
- Vind jij dat data kwaliteitsmanagement binnen organisaties een probleem is? Is het moeilijk om data goed te managen en brengt het enige moeilijkheden met zich mee?
- Is het ooit voorgekomen dat de data niet uw wensen had vervuld, bijvoorbeeld dat je foutieve data had ontvangen? Zo ja, wat waren vervolgens de gevolgen en was de acties die jij ondernam om het op te lossen?
- Datakwaliteit heeft een aantal gedefinieerde karakteristieken, zijn deze bij jou bekend?
- *De karakteristieken zijn: Foutloosheid, Geschiktheid, Juistheid, Volledigheid, Actueel, Toepasbaarheid, Herbruikbaarheid en Tijdigheid.* Nu je de karakteristieken hebt gehoord, worden deze toegepast binnen jouw organisatie? Zo ja, hoe specifiek?

Data verzameling en visualisatie

- Hoe wordt data binnen het bedrijf verzameld en ervaren jullie enige moeilijkheden hiermee?
- Gebruiken jullie de verzamelde data om visualisaties te maken, voor bijvoorbeeld rapportages of verantwoording aan management? Zo ja, hoe maken jullie deze visualisaties?
- Ben jij tevreden met de huidige methoden voor data verzameling? Waarom wel/niet?

Maturity model voor het verbeteren van data kwaliteitsmanagement

- *Introductie van ontwikkelde maturity model en discussie over dit model.*
- Wat zijn jouw gedachten en meningen over dit model? Heb jij ooit een soortgelijk model gebruikt? Zo ja, hoe heb je deze toegepast? Zou jij dit model gebruiken?
- Gebruiken jullie binnen de organisatie enige methoden, technieken of software voor het managen van data? Zo ja, welke en wat zijn dan de procedures?
- *Scorecard bespreken en mening vragen. Het gebruik ervan ook testen.*
- **<u>Zou jij dit model gebruiken en in welke situatie?</u>**

Sluiten van het interview

*Controleren of alle vragen zijn behandeld.

*Aanvullende documentatie vragen, wanneer noodzakelijk.

*Bedank geïnterviewde.

Appendix V: Coding agenda

The following table represents the coding agenda for the qualitative content analysis of the data collected from the interviews.

Category	**Definition of the category**	**Examples from the interviews for statements fitting into the category**	**Coding rules**
Theme 1: different types of data			
a. What types/kinds of data is the interviewee dealing with?	The aim of this question is to find out which data shows what kinds of challenges and difficulties. Furthermore, this makes it possible to ask further in depth questions during the interview.	"It was mainly about employees' performance within the front-office, f.e. amount of times called, time spend on each call, average time of each call" (I1)	The first variable gives us the types of data and the second variable should show further challenges with regards to those types of data.
b. How do you store data within your firm?	This question will identify the different methods for data storage within firms. This could possibly show what is going wrong.	In principle, the data is stored within databases. We use a lot of different company applications, line of business applications. Each of these applications has their own database. We aim to link all these databases together, so that all these databases are feed from one source. (I8)	Find out what is going wrong. How are firms storing their data? Furthermore, this category may provide further insights for data collection by firms.
Theme 2: the terms 'data quality' and 'data management'			
a. What does data quality and data management mean for the interviewee?	The aim of this question is to find out how different individuals define the terms of	For me 'data quality' is the care and duty you have to ensure that the data that you own, that these	This questions gives a good overview in how the different individuals define the terms.

	data quality and data management and if there are any differences.	are consistent, accurate and are up-to-date. That is one aspect, and the other is that data that you should not own, should not be present within the firm and other things that are concerning these data. And with this you come to 'data management', not only ensure that the data is accurate, but we are also within the social services, within this service you are forced to store certain amounts of data, and wherein also it is mandatory to delete this data after a certain amount of time. (I8)	Furthermore, this shows how each firm perceives these terms.
b. How does your firm achieve data quality?	What are firms doing to achieve data quality? The individual could also provide signs of lack of attention with regards to data quality.	"Within our own organization, we managed our data accordingly and therefore achieved qualitative data". (I3) "Our customers were not paying that much attention to managing data quality. This proved when we did analysis on their data and found the same city registered with three different names. For example, Amsterdam centraal,	Certain steps, procedures, methods could be given as answer to these questions. Furthermore, it will say what about data quality within the firm.

		Amsterdam ctr., Amsterdam center. This showed us that open forms were causing bad qualitative data, and preset entries should be used." (I3)	
c. Is your firm applying any principles or standards for the management of data quality?	The aim of this category is to find out the different kinds of ways firms manage the quality of their data. For example, by following certain steps defined within documents.	"We follow BPMN for process optimizations, these ensure standardized forms. The forms make it possible that user do not make simple entry errors, for example writing a city name different." (I3)	Within this category, it is important to ask further in depth questions so that finding the right fit with the procedures or standard can be found.
Theme 3: data quality management			
a. What are your thoughts on data quality management?	The aim of this question is to determine the opinion of firms about data quality management as a quick overview.	There are two things. The one are the simple registrations that should easily be solved, but are difficult enough. We have audits for our basis registration persons and building, at the audit we look how the quality is and a small margin of error is allowed. The very difficult point is with regards to making managerial decisions, because documents with regards to these decisions are stored on three to five different locations, which ensures difficulties in retrieving the right document. (I8)	Find out what firms think about the topic of this thesis and gather further insights.

b. Do you think that data quality management is an issue within firms?	This category extends category 3a and digs deeper in the opinions of the interviewee about data quality management.	Not directly, I am amazed at how some customers have arranged their procedures. We encountered a customer which had 10.000 products on their website, which she had written down in documents and not in databases. (I6)	The goal of this question is to determine issues that firms are currently facing and also gather some possible solutions from the interviewee's point of view.
c. What happens when data does not meet your requirements?	In order to understand the different actions each interviewee takes it is interesting to look at examples with regards to data quality issues.	Sometimes firms want to know something with regards to their data. However, they do not have this data and therefore making visualizations etc. is not possible. Sometimes relations between two datasets are just not possible and therefore the needed data cannot be gathered. (I9-II)	Within this category, it is important to pay attention to why data is not meetings the requirements of the users.
d. Data quality characteristics.	According to the literature some data quality characteristics are defined, this category aims to find out if these are known within organizations.	"I have known these characteristics, but not specifically used them." However, these characteristics are kind of self-explanatory. For example, that data should be Free-Of-Error, garbage-in and garbage-out should be prevented. (I2)	Are firms using these characteristics and are they known with their importance? Analysis of the firm's point of view is important there.

Theme 4: data collection and visualization			
a. How do you collect data within your firm and do you experience any issues?	This category aims to find out what procedures or standards firms follow to collect meaningful data within the firm. Furthermore, attention is paid to any issues individuals experience.	We collect our data by making direct linkage with the Excel files. "We mainly have difficulties in combining multiple information sources, because the data within the different systems is not written down the same way." For example, within software package A, credit numbers are saved as 103 etc., but within software package B, credit numbers are saved as 103-1. (I5)	The answers should be analyzed and further in depth questions should be asked in order to find fitting information with regards to the category.
b. Does your firm create visualizations with your data? How?	In order to understand the uses of data within the firm, this question was asked. Making visualizations could improve analysis of data.	Visualizations are made for customers, with the help of a third-party tool provided by a Canadian software developer. (I1)	It is important to take into account the differences each firm have with regards to retrieving data from their systems.
c. Are you satisfied with the current methods for collecting data?	This category is added so that difficulties/problems with regards to data collection can be found, which in turn could provide some overview in how reports are created.	If I am honest, not! Customers should plan their data notation methods, so that the same data is not noted differently. *'Often we need to figure out how the data should be collected or created and what each piece of data exactly means.'* (I5)	Attention should be given to any remarks or opinions the interviewee gives. Additionally, ask in further questions.

Theme 5: maturity model to improve data quality management			
a. What are your first thoughts and remarks about this maturity model?	In order to understand the interviewees' interpretation of the maturity model and gain additional feedback for improvements it is fitting to ask this question.	When I look at the maturity model I can put ourselves within the optimizations part of the model, because we are mostly focusing on optimizing. Furthermore, the maturity model provides a good overview of the different maturity levels that are shown within the model and provides sufficient information. (I4)	What does the interviewee think of the model? Look at statements with regards to the maturity model.
b. Have your ever used a similar model?	The goal of this category is to find similar models and get inspiration from these models, in making additions to the maturity model.	I have seen a similar model, and helped in the development of it myself. This model was called the Prince2 Maturity Model, how much the firm followed the requirements, the better the firm performed. This model was made digitally, filling in the scorecard exported a report with the tasks to do. (I3)	How are these models applied and in what context are they used?
c. Is your firm using any methods, techniques or software for data management?	This category is to understand the experiences and principles of data management applied within the firms. Furthermore, it is to test whether the firms experienced any issues for these.	I have said that we have some programs within the firm to report data, by uploading these to the systems. We use three main programs, I can't give these names because of the confidentiality. (I7)	What are the issues connected to these methods, techniques or software and how are these affecting the firm?

d. After using the scorecard, what are your thoughts and opinion?	The aim of this category is to acquire information to make improvements, changes and additions on the scorecard in order to make it suitable for usage in firms.	The scorecard acts as an easy tool for determining which maturity level is fitting for the firm. However, some points are not that clear, for example the scores of the first factor are not really clear. Because you cannot directly know how much people are within a Team, Division or Enterprise. Furthermore, some factors could be formulated differently so that everyone can understand them. For example, metadata is not a term known by everyone, technical persons would know it, but non-technical persons immediately wouldn't. (I1)	Focus on the answers to the responding questions in detail and acquire in depth information in order to make the correct adjustments to the scorecard.
e. Would you use this model and in what situation would you apply it?	This category is defined so that additional information could be gathered about the usage of the maturity model within practice.	Yes, I would use this model. With the help of your model and scorecard, we can discuss certain points and you are really talking about some points about data quality management and that gives you the possibility to give wise remarks, that is the use of it. (I8)	If negative answers are acquired, ask for deeper understanding so that the maturity model can be updated for better fit.

Appendix VI: Category system

For each theme, the different categories will be analyzed in more depth and provided with the answers and statements of each interviewee. The text represents the expressions, opinions and actions of each interview and are not always directly quoted from the interviews. Some questions are not asked in the same context as written in the categories, because some questions flowed through previous answers.

Interviews I1 - I3

Category	**I1 - Company A**	**I2 - Company B**	**I3 - Company C**
Theme 1: different types of data			
a. What types/kinds of data is the interviewee dealing with?	"It was mainly about employees' performance within the front-office, f.e. amount of times called, time spend on each call, average time of each call"	"We are working with occupational safety numbers, capturing personal data, welfare of employees in the workplace and reports for higher management"	"We are dealing sensitive data, mainly personal data. For example, incident data: what went wrong? How many times?"
b. How do you store data within your firm?	"The data was mainly written on paper by employees and afterwards stored in Excel by the manager"	Before we introduced our system to the customer, they were storing all their data within documents. By introducing our system, all the data stored by the employees went directly into the systems' databases and therefore it was possible to monitor the data real-time and make the right and fitting decisions.	"Our customers are using SAP and the databases attached to it. Furthermore, Excel is heavily used which ensures problems for automatization."

Theme 2: the terms ‘data quality’ and ‘data management’			
a. What does data quality and data management mean for the interviewee?	“For me the quality of data is, how good the data is. For example, what is stored within the files? is the stored data correct? etc.” “Data management is more how you monitor your data, is data managed manually or are there automated processes for management?”	“Data quality is known as correct registration of data. For example, making analysis on incidents within a certain city is not possible when employees enter the name differently each time.” With this data quality, also means storing the data in the right way. “Data management means how you work with your data and how do you store it?” Additionally, a firm could have some requirements, for example ISO 27001. Furthermore, the data should be accessible to every user, so that the data is up-to-date.	“In my opinion data quality means the reliability, availability and completeness of data.” “Data management means the maintenance of data, availability, capturing the data, retrieving and securing the data?”
b. How does your firm achieve data quality?	The firm is not directly focusing on increasing data quality, the interviewee mentioned that “management wanted to improve data quality, however the employees were not really fond of it, because they could be monitored more easily.”	We did not focus on enhancing data quality within the whole firm of our customer, because she had many complex projects. Our system was not part of the customers’, we made connections with their databases and ensured that our system had	“Within our own organization, we managed our data accordingly and therefore achieved qualitative data”. “Our customers were not paying that much attention to managing data quality. This proved when we did analysis on their data and

		the right corresponding data and could be managed and maintained by us.	found the same city registered with three different names. For example, Amsterdam centraal, Amsterdam ctr., Amsterdam center. This showed us that open forms were causing bad qualitative data, and preset entries should be used."
c. Is your firm applying any principles or standards for the management of data quality?	No, not really defined within the firm. The employees use their common sense to perform their daily work processes.	"We don't really use any procedures or standards; mostly common sense is applied." "Our customers mostly have their own advisors that determine the data needed and predefine certain KPI's that need to be monitored."	"We follow BPMN for process optimizations, these ensure standardized forms. The forms make it possible that user do not make simple entry errors, for example writing a city name different."
Theme 3: data quality management			
a. What are your thoughts on data quality management?	Data is very important within the firm, especially within this time. It starts with data and this means that firms need to thinks about optimization and automation. Further, what is the best way to store data and how can firms ensure low margins of error?	"Small firms do not pay additional attention to data quality management, because it is a time-consuming task" However, I think that large firms are putting some effort in it, they probably have some defined procedures to store data correctly.	"With regards to data quality management, I think that **there is still a lot to do**." On average firms perform below the quality standards. From the outside, you think the firm is doing good, but from the inside it is a different story.

b. Do you think that data quality management is an issue within firms?	"The problem is often that firms do have the data, but they do not know if the data is usable." Furthermore, insights in data are important and many firms do not really understand this.	"I think that in some cases data quality management is an issue. For example, a customer of ours was making backups of her data on tapes and once these tapes broke and the original data had issues and this had some major consequences and work loss."	"In my opinion, it is a huge problem at the moment."
c. What happens when data does not meet your requirements?	We arranged dashboards, but these sometimes gave errors. This was mainly due to the data not been up-to-date, in one case the data was not updated in 2 hours and this ensured that management was working with the wrong data.	I did not directly come across this, within the work I perform. I mainly install our software within our clients' firms and do not look deeper at their operations.	"I have experienced this a lot. For example, one time I needed to make some calculations with the hours worked of employees. However, the data was missing and a new employee had to be hired in order for this data to be managed accordingly. It took about 1,5 years for the data to be arranged correctly, in order for me to extract the correct data.
d. Data quality characteristics.	I am not very technical; therefore, I use these characteristics in a form of common sense. For example, correct data is important for creating our visualizations and it should be applicable so that every user understands the data.	"I have known these characteristics, but not specifically used them." However, these characteristics are kind of self-explanatory. For example, that data should be Free-Of-Error, garbage-in and garbage-out should be prevented.	On default firms are working with these characteristics. "Data should always be correct, this could be tested by sampling and doing checkups of the data. Many times, firms are using Excel files, and these are hard to automate."

Theme 4: data collection and visualization			
a. How do you collect data within your firm and do you experience any issues?	"During the process of data collection, we combined several data sources and gathered our data with the help of API's." However, data was not maintained correctly within the applications because of the manual activities needed with the help of Excel.	"We collect our data from our own MYSQL-databases and make connections by performing queries. Further, we do not experience any issues, because we maintain our data ourselves."	"We collect our data through employees, we ask which employee can guide us? Further, we gain access to their databases, excel files and files on certain disks"
b. Does your firm create visualizations with your data? How?	Visualizations are made for customers, with the help of a third-party tool provided by a Canadian software developer.	"We create visualizations for our customers with the help of our own software."	Previously, we used Qlikview. Currently, we are using PowerBI by Microsoft, because of easier implementation. If specific requirements are set, then we make the software ourselves.
c. Are you satisfied with the current methods for collecting data?	The customer was satisfied with the methods for data collection, a combination of databases and excel files. However, personally "I am not satisfied with manual work, some procedures should be automated so that time can be spared"	"I am satisfied with our own methods for data collection and do not experience any problems with them."	"Totally not, firms are working too much with Excel. There are many lists that need to be imported to systems, and additional columns needed to be added to databases." "Firms do not have a specific format, it is always different, we have questionnaires who provide solutions."

Theme 5: maturity model to improve data quality management			
a. What are your first thoughts and remarks about this maturity model?	The maturity model provides a good overview of the different levels. Furthermore, each firm can think about their selves and put themselves within a certain maturity level. However, it is difficult to put yourself in a level.	The maturity model is interesting and provides a good overview of steps to improve data quality management.	Do all the factors weigh the same? Maybe instead of making columns next to each other, you could make it like stairs, so with each 'step' a newer maturity level is reached. You could maybe also add hidden scores behind each score, so that each factor can be scored and a final score can be given.
b. Have your ever used a similar model?	"I did not use a similar model; however, I have always had these steps within my head. With this I mean, that I always create an overview of the situation for the client and do analysis based on these."	We did not really use any maturity model. However, we worked with a different model named the BSC Matrix, which also provided steps for improvement.	I have seen a similar model, and helped in the development of it myself. This model was called the Prince2 Maturity Model, how much the firm followed the requirements, the better the firm performed. This model was made digitally, filling in the scorecard exported a report with the tasks to do.
c. Is your firm using any methods, techniques or software for data management?	We use third-party software to create visualizations with the data of our clients.	No, we don't use any special methods, techniques or software.	We are not using any models to manage our data, it is mostly done with common sense.
d. After using the scorecard, what are your thoughts and opinion?	The scorecard acts as an easy tool for determining which maturity level is fitting for the firm. However, some points are	The scorecard is clear; however, some factors could be written with some simpler terms so that everyone understands them.	Some factors within the scorecard are difficult to interpret, because they are not directly applicable for small firms. For example, the

	not that clear, for example the scores of the first factor are not really clear. Because you cannot directly know how much people are within a Team, Division or Enterprise. Furthermore, some factors could be formulated differently so that everyone can understand them. For example, metadata is not a term known by everyone, technical persons would know it, but non-technical persons immediately wouldn't.	Furthermore, I find it difficult to mention how many strengths and weaknesses we have defined, maybe the scores to this question could be formulated differently. For the last factor, it is difficult to know where you stand with regards to waste management, because employees are unaware of which data is where.	fourth factor mentions procedures and standards defined within the firm, however small firms often do not tend to define these, because they are time consuming to develop.
e. Would you use this model and in what situation would you apply it?	The maturity model and scorecard are applicable after some explanations, however after simplifying some factors and optimizing the model it could be more applicable within practice. These changes could also make it possible for non-technical persons to use it.	This model of yours is more applicable for large firms instead of small firms, because some factors do not directly apply to small firms. Both types of firms work differently, and therefore probably have different maturity levels. Maybe you could make a final calculation based on single scores given to each maturity level. Furthermore, how many times do firms need to score themselves?	Currently, the model has some points that can be improved so that the understandability is enhanced. However, after improving the maturity model and making some additions it would be applicable within all kinds of firms. Like I said, maybe you could add some scores so that an overall score is given and each firm can score themselves based on the score.

Interviews I4 - I6

Category	I4 - Company D	I5 - Company E	I6 - Company F
Theme 1: different types of data			
a. What types/kinds of data is the interviewee dealing with?	We are working with different kinds of data, divided between different categories. (1) Sales figures; revenues, profits etc. (2) CRM data; time called, information send. (3) HRM data; hours worked, time planned.	We are mostly working with Financial data (turnover, finance figures from 'Accountview'. We visualize financial statements for our clients. Furthermore, we link ERP software 'Reflex', wherein production, inventory, material costs and production costs are registered.	We have data for analytics, for example users that visit a website. We also work a lot with data regarding product information on websites/web shops that are stored within databases. Internal data is mostly about financials, projects, CRM and HRM which we store within software and their databases.
b. How do you store data within your firm?	We mainly store our data within databases, which speaks for itself since we are an IT company. "Further, we do not use any Excel files, because arranging these is not simple and brings a lot of work with their selves." All the documents/agreements we use are stored within the software that we use.	The data is mostly stored within ERP systems. Excel is also used by our customers. Excel files are directly loaded into Qlikview. "However, Excel files are very prone to error, because these are adjusted many times during the day and therefore some data is mostly missing." "Firms make errors when duplicating data from previous years and therefore cause problems when making comparisons with this data."	We use a combination of databases and documents. Most documents are project related, for example regulations. We have these stored within the network of the firm and these are publicized through our intranet where employees with the right roles can get to the files.

Theme 2: the terms ‘data quality’ and ‘data management’			
a. What does data quality and data management mean for the interviewee?	“Data quality is converting your data to valuable information, and this information is also usable and that you also store the necessary data in the right and correct way.” “This links back to the term data management, where you need to maintain your data correctly and don’t use fragmented data, where eventually this data does not provide any value. Also, that everything is automated and clear, and that you store the correct data.”	“Data management is how you arrange your data and the structure of the data within different software of the firm.” “Data quality mentions reliability, correctness and completeness of the data.”	“Data quality is about data integrity and the things around that.” “Data management is how you store the data and how you manage these.”
b. How does your firm achieve data quality?	We have automated everything within our firm and try to prevent many manual proceedings. Furthermore, we try to perform updates and check our data regularly.	Our firm is not internally giving additional attention to managing data quality, we are not that big and provide our services externally. Externally, we work with the data that is given to us and compare the given data with the data stored in the systems. We do not specifically search for errors within the systems.	We are focusing on version management, for example Backups. It is more data management, we have strict procedures for making these backups.

c. Is your firm applying any principles or standards for the management of data quality?	We have some guidelines, that we follow. For example, we have software for handling invoices. These invoices are audited by big accounting firms. Furthermore, we receive feedback for our software by our users and update the security and characteristics of the software based on these.	"We don't have any quality procedures, only responsibilities of what needs to be done." Quality is more a form of art for consultancy, for some the quality of dashboards is good for some it isn't.	We use oWasp checklists, which is continuously improved by the community. Within this list, you are talking about data encryption till the security of a server. Furthermore, we started using a new ERP software, which we also have chosen to increase the quality of our data and our reports.
Theme 3: data quality management			
a. What are your thoughts on data quality management?	"In my opinion this is very important. You need to adapt to the developments within the organization on each level, not only software development, but also other things round it." A firm should know their points of improvement, so that the procedures can be followed correctly.	"The quality of data is important, because our firm is based on qualitative data and what we can do for our clients with this data." For example, the visualizations we make are nothing without good qualitative data.	"For me, it is important to separate the different kinds of data. The ones that are critical and may not contain any mistakes."
b. Do you think that data quality management is an issue within firms?	We do not experience any issues with regards to data quality management. "Our processes are defined correctly, with regards to communication and quality. You want to be on the same line as the customer and prevent any forms of miscommunication"	"The main issue currently is that firms are still using Excel files to maintain their data, this is a huge issue. Firms should make the switch to software packages and stop using Excel files or other kind of documents to store their data."	Not directly, I am amazed at how some customers have arranged their procedures. We encountered a customer which had 10.000 products on their website, which she had written down in documents and not in databases.

c. What happens when data does not meet your requirements?	"Yes of course, within our open source CRM software I saw directly what I needed and what I really didn't need and rearranged the software so that it would meet my requirements." Sometimes customer data is outdated and we need to find out what the newest data is. However, finding the new data is not always that hard and depends on every person ways of working.	"In the case of Excel files, I noticed that the data is not copied over right to new files, which ensures that the data loaded in Qlikview shows some problems." We take actions by contacting the firm and mentioning that the data is not correct. We don't have the authority to change the files ourselves and therefore it takes a lot of time for adjustments to come through and match every year format.	"**The most common reason for missing data is that employees don't follow procedures or standards."** For example, employees create documents which they in turn do not store on the disks of the firm, but keep it on their personal computer. This ensures that it is not accessible by others and sometimes causes delays in our work.
d. Data quality characteristics.	These are known to me, we have our own tool for finding bugs within our software. By finding these bugs we solve problems with regards to these characteristics.	I have heard of them before, but not specifically used them. We ensure within the software package of Qlikview that the data is complete, reliable, correct and up-to-date. Manual entry within systems are delegated to other employees.	We do not directly focus on these characteristics, but keep these in mind during our daily tasks. **"The data within out firm is not always correct, but this is mainly because employees forget certain tasks by not following the defined procedures."**

Theme 4: data collection and visualization			
a. How do you collect data within your firm and do you experience any issues?	"We collect our data from forms that users can fill in within the front-end of our software. When the data entered by the customer is not clear, we contact the customer and try to improve the data."	We collect our data by making direct linkage with the Excel files. "We mainly have difficulties in combining multiple information sources, because the data within the different systems is not written down the same way." For example, within software package A, credit numbers are saved as 103 etc., but within software package B, credit numbers are saved as 103-1.	We collect our data from systems, and some manual work is still needed. For example, an employee needs to collect data from different systems and create a report within Excel. Furthermore, some tasks could be automated but are not done, because employees don't take actions for making improvements they just repeat the same tasks over and over again.
b. Does your firm create visualizations with your data? How?	We use our own software to create visualizations with regards to our projects. For example, we track project status, employee performance, deadlines etc.	We create visualizations on the location of our clients by combining software with Qlikview. The dashboards within Qlikview are arranged based on the requests of our customers.	Within the new ERP system that we use, the calculations for the visualizations are done within one central Excel file, which is maintained by one individual. "The system is well organized, the reports are automatically created, the data sets, queries and filters are done within Excel."

c. Are you satisfied with the current methods for collecting data?	"Till this time, we are satisfied, what is important for us it that we can see if everything is going as scheduled but also if the customers has everything they need."	If I am honest, not! Customers should plan their data notation methods, so that the same data is not noted differently. *'Often we need to figure out how the data should be collected or created and what each piece of data exactly means.'*	It could be done better, but I am not dissatisfied with the current methods. However, we could make improvement within our file management system, by for example switching to a SharePoint environment.
Theme 5: maturity model to improve data quality management			
a. What are your first thoughts and remarks about this maturity model?	When I look at the maturity model I can put ourselves within the optimizations part of the model, because we are mostly focusing on optimizing. Furthermore, the maturity model provides a good overview of the different maturity levels that are shown within the model and provides sufficient information.	The maturity model is different with the models I have used during my study. I have used some similar models with stages for improving some certain aspects of a firm. Your maturity model looks like a process diagram, which firms can follow to reach a certain fitting maturity level.	The maturity model provides a good overview of the different maturity levels. However, without further explanations, you do not directly know what the maturity model represents. Therefore, maybe the maturity model could be arranged differently, so that directly you can see which maturity level is the highest and how it can be reached.
b. Have your ever used a similar model?	I do not know if our firm has use a model similar to your model before, because I did not participate on developing the procedures and standards within our firm.	I did not use a similar model in the past, and not seen one to date. Maybe top management has used one before, but if they did I do not know.	"We have used some similar models, but not one for data management. We used some maturity levels, but yeah not specific with regards to data quality management."

c. Is your firm using any methods, techniques or software for data management?	"We don't use specific methods, but rely on feedback received from employees or customers. We developed the first version of our software and with the feedback of our customers made adjustments and came to the current version."	We don't use any methods or techniques for managing data. We only use the software of Qlikview to provide good overviews of the data, but do not directly manage the storage of data.	"We have several procedures, it is not like that we can say we have a standard ITIL procedures and that kinds of things."
d. After using the scorecard, what are your thoughts and opinion?	The scorecard is clear; however, some points can be changed within context. For example, what is a division, team or enterprise? And some terms like waste, standardized should be simplified so that a normal person can understand this.	Your scorecard is different with regards to other scorecards I have used, normally you score the given factors and reach a final score. Maybe this could be done as improvement to yours. Furthermore, some factors are difficult to comprehend and should be written simpler so that everyone can understand it. Additionally, numbers could be added before each maturity level, so that it is easy to address a certain factor.	The scorecard is filled in step by step, and each factor is discussed in depth so that everything is clear. Terms like 'metadata' should be worded differently, so that everyone can understand it, or maybe provided with additional explanation. The scores for the first factor, should be worded differently, because it is not applicable for small firms. For the ninth factor, the strength and weaknesses are weird with the given numbers in the scores, it is not directly applicable, maybe percentages would be better here. "We always mention waste management, this needs to happen but it just doesn't. People just don't see the added value."

e. Would you use this model and in what situation would you apply it?	If I am experiencing issues with regards to data quality management, then I would use this model. However, maybe it is interesting to see which other models are of play within this field of study, are there any other models that do something similar?	This model is applicable within practice. However, maybe you could use open questions in order to gather more information from the company. Additionally, you should explain what happens when a firm reaches a certain maturity level.	"Models like these can help firms to see in what situation they are situated. We for example can see where we are and where we aim to be." So, it could give a good first overview of the situation we are in. There are moments that you feel that you are within a certain maturity level and then you could see that within this aspect a lot of things go wrong and should go better, then it adds more value in my opinion.

Interviews I7 - I9

Category	I7 - Company G	I8 - Company H	I9 - Company I
Theme 1: different types of data			
a. What types/kinds of data is the interviewee dealing with?	We are mainly working with documents in the form of agreements with other firms. For example, we do project management wherein we check if every needed product is delivered and if it meets the requirements of the client. Furthermore, we work with financial budgets, invoices, certificates etc.	We work with a lot of different types of data. We have the basis data, that are legally required. For example, basis registrations of persons, buildings and the connection between these. The registrations about persons, is that every person within the Netherlands is unique, with their BSN number, the information around the BSN number needs to be exactly correct. The responsibility of this data lies with the local governments.	That depends, financial data, employee data also about company resources. But it has mainly to do with the needs of the customers. Some customers want to do market analysis, then we work with for example Open Data. (I9-I) From technical perspective, it is about files that come from systems, or databases of the customers. Online sources, like Google Analytics and Salesforce is also used for data collection. (I9-II)
b. How do you store data within your firm?	We retain hard copies of everything and also soft copies of data. Important documents are stored within the systems, but we have an archive of all the files for a project situated within a different floor, together with	In principle, the data is stored within databases. We use a lot of different company applications, line of business applications. Each of these applications has their own database. We aim to link all these databases together,	Internally, we store the data within the systems that we use, but at our customers it differs every time based on the procedures of the customer. The data sources that we use mainly depends on the demands of the

	thousands of other projects done in the past.	so that all these databases are feed from one source.	customer and resources provided by the customer.
Theme 2: the terms ‘data quality’ and ‘data management’			
a. What does data quality and data management mean for the interviewee?	“Data quality is about the quality of the documents. The quality is about its originality, its understandability, correctness and about its confirmability.” It is possible that we receive falsified documents and therefore all the documents should be original and verifiable. “For data management, all the documents should be uploaded to an electronical system, so that we can prevent the loss of documents and speed the process of finding documents. For example, sometimes we need a document that is stored within the archives and are spending too much time on finding the file.”	For me ‘data quality’ is the care and duty you have to ensure that the data that you own, that these are consistent, accurate and are up-to-date. That is one aspect, and the other is that data that you should not own, should not be present within the firm and other things that are concerning these data. And with this you come to ‘data management’, not only ensure that the data is accurate, but we are also within the social services, within this service you are forced to store certain amounts of data, and wherein also it is mandatory to delete this data after a certain amount of time.	Data quality is in my opinion, when you think about it, it is completeness, consistency, reliability and more have to with regards to quality. (I9-I) Data management, how you arrange this, is also a really broad term. How am I storing my data, how can I guarantee quality and how do I organize this and all the other things that fall under data management. Data quality is something that you have to arrange with the help of data management. (I9-I) With ‘data management’ I am more thinking about metadata. You should define everything so that each kind of data can be found in the right place without any issues. Data quality is more a technical aspect for me, because I am very technical, the number of

			lines should be correct at the end, I should not lose any data. The data quality is the verifiability of my process and this is not only counting data. The outcome of data should always be correct with what your expectations are. (I9-II)
b. How does your firm achieve data quality?	We work together with the third-party developers in order to improve our software and these also maintain our data within the software and ensure that the data can be registered at the right location.	Gradually, we are increasing this attention. There are also more possibilities now, we optimizing the way of storing applications so that errors in for example addresses is can be avoided.	There is a difference between process quality and output quality. For the process quality, we look at the data and check if the measurements are correct. We do not check if the data within the data sources is correct, the visualizations are checked on errors by our customers at release, but not us afterwards. (I9-II)
c. Is your firm applying any principles or standards for the management of data quality?	We have some documents within the firm wherein some principles and standards are defined. Further, when we perform new tasks, we write these on paper and afterwards transcribe these on documents within the systems. Before accepting these principles, we discuss them within the firm and make a final	Procedures are developed, linkages with other applications is realized, scans are done within different databases, are the linkages correct and we search for mistakes. This is a continuous struggle.	Those are two things, creating a dashboard is a creative process, you cannot define some standards for these, we have standards with regards to the way in which we create our dashboards. (I9-I)

	decision if agreed on by everyone.		
Theme 3: data quality management			
a. What are your thoughts on data quality management?	"The environment of the firm should be arranged in a way that every employee can give their feedback or points for improvement. The firm should benefit from everyone's knowledge and experience. An environment like this should be created, we need to have a good electronical environment for managing files accordingly."	There are two things. The one are the simple registrations that should easily be solved, but are difficult enough. We have audits for our basis registration persons and building, at the audit we look how the quality is and a small margin of error is allowed. The very difficult point is with regards to making managerial decisions, because documents with regards to these decisions are stored on three to five different locations, which ensures difficulties in retrieving the right document.	"I think that many firms are neglected data quality management." They mention that they have systems and that these are correct and start to work with these. "From business perspective firms are giving much too little attention to the quality of data." (I9-I)
b. Do you think that data quality management is an issue within firms?	Within our firm, the data quality management is not mature enough. We are an average firm when it comes to this topic. Because we cannot get all the data from electronical environments and because many documents only have hard copies, we experience issues when verifying documents and	I think that it is not too bad for our organization, if it is mainly about the quality of our data, then I think that we really think through with regards of the company data we possess, data with regards to privacy, then I think that we are not doing too bad. We are really focusing on	Sometimes the customer mentions that they can find the data for you and then you come across different principles of how the table has to be visualized. In that way, YES, we are experiencing many situations where data is not meeting our requirements the first time, for what we want to do with it. (I9-I)

	finding documents. Because of these issues, we lose a lot of time.	the quality of this data and are doing really good.	
c. What happens when data does not meet your requirements?	When data is not meeting our requirements, the projects of our clients may suffer because of delays with regards to budgeting.	I call the person responsible telling him: 'I need this document, find it for me.' And then I hope that the person understands what I need and that I eventually will receive the correct document.	Sometimes firms want to know something with regards to their data. However, they do not have this data and therefore making visualizations etc. is not possible. Sometimes relations between two datasets are just not possible and therefore the needed data cannot be gathered. (I9-II)
d. Data quality characteristics.	I am not known with these characteristics, but now you have mentioned them, data should be up-to-date, data should be comprehensible, the needed data should be available. More importantly, the data should be verifiable, because in our line of work we cannot afford to process falsified information.	We aim to use these characteristics within our processes and guide based on these. Especially, with regards to basis registrations and registrations based on managerial decisions or permits.	The data you are working with should be available, you just focus on the quality, is it accessible, usable, is the data correct. (I9-I) You are performing a scan of the data, and you look at the completeness of this data and look at what you could expect. (I9-II)
Theme 4: data collection and visualization			
a. How do you collect data within your firm and do you experience any issues?	We have electronical data and hard copies, for the electronical data our most important data set are e-mails, and the files that are	"Management reports are a constant headache." Making these reports within IT is very difficult and the reports	We are collecting data from the sources that our customers provide to us. This differs for every customer, sometimes there

	attached within e-mails. The e-mails that we send are formal, in some cases an e-mail can act as a signature. Our main item are documents that are signed. Within the hard copy environment, we register every document.	concerning governmental activities, are mostly concerning social domains within the city and the interrelations that are present.	are systems within the firm, sometimes these are Excel files etc. It just solely depends on the needs and demands of the customer. We do not directly experience any issues, but sometimes struggle in gathering the data. (I9-I) (I9-II)
b. Does your firm create visualizations with your data? How?	A reporting unit exists within our firm, these do all the reporting. I don't know a lot of these systems. Statistical issues are always uploaded to the systems. Forecasts for the payments are made, they mention when some certain payments will be executed, because there are deadlines.	In general, we use software like Cognos, data warehouse tool. We do little within Excel.	Our firm is focused on creating visualizations for our customers with the help of third-party applications. We listen to our customers and arrange their data based on their demands and wishes. (I9-I)
c. Are you satisfied with the current methods for collecting data?	I am not satisfied with it. In my opinion it is not very beneficial, because the reports within our firm always depend on forecasts. And within these forecasts the margin of error is present, some deviations exist. For example, sometimes you mention you will pay for a task in February and because of some issues this gets extended by two months.	No one is ever satisfied, but we are doing it very well. Just like I mentioned because the requirements change, we are always looking at ways to collect data better.	I am satisfied, personally. We are looking very good at a combination of both collecting and visualizing the data. Sometimes this causes errors, but that is ok. The structural and traditional approach and reach fast conclusions, we have both. However, the question is what the customer wants and what we want. (I9-I)

Theme 5: maturity model to improve data quality management			
a. What are your first thoughts and remarks about this maturity model?	The maturity model could be more done as stairs, so that you directly can see which maturity level is the highest and which steps you need to undertake to reach them. Furthermore, what are the data quality characteristics saying below the maturity level, do these have any additional value, is one positive and is one negative characteristics. These could be more elaborative.	I have some knowledge with regards to maturity levels, but not with regards to data quality management. Your maturity model provides a good overview with regards to each maturity level and how they are matched against each other.	When looking at your maturity model the first thing I am thinking about is the applicability of it for firms of any size. For example, small firms and large firms would probably score differently and it would differ in use. Therefore, you should change your factors so that they also apply for small firms and not only large firms. This should be defined accordingly within your research. (I9-I)
b. Have your ever used a similar model?	Not on the topic of data quality management. We are more using surveys within our firm, based on these surveys we get send to some sites. I have not used any models within the firm, only during the time of my study.	With regards to data quality management, not directly. We use the model of King, maturity levels of King quality institute of the Dutch government. That model is over similar issues, but your model is really about data.	No, not to determine the maturity of a firm, we surely have not. Once a firm only asked me to arrange their metadata management, so that they could monitor where their data is. (I9-II)
c. Is your firm using any methods, techniques or software for data management?	I have said that we have some programs within the firm to report data, by uploading these to the systems. We use three main programs, I can't give these names because of the confidentiality.	We store all the data we use, with the help of the main processes of the government. We save this within our document management system. There are processes that we automated, we don't create any	We are not directly using any methods, techniques or software to manage data quality. We have some procedures with regards to visualizing the data, for example that important views should be on the left top side of the screen etc.

		files manually with regards to the citizens.	But not any with regards to managing data quality. (I9-I)
d. After using the scorecard, what are your thoughts and opinion?	Firstly, like I said an overall score would be handy. Additionally, this model can't be easily applied within firms of a small size, for example firms whom have 1 – 5 employees. Furthermore, maybe some additional factors could be added, and within some factors the scores could be changed. For the **ninth factor**, where you define your strengths and weaknesses, maybe these could be exchanged with percentages.	In overall the scorecard is clear and is simple to fill in. However, with the first factors it is not really applicable within practice, because applications are maintained by some individuals and not the whole organization. Maybe this could be changed in a way that it reflects common practices. For the factor with regards to updates, it is hard to say when you do them, because it is really application dependent.	*The scorecard was not filled in for this firm, but rather used as a discussion point about certain factor. For example, for the first factor the discussion started if the scores are defined correctly, because an enterprise and team cannot do the maintenance of applications. For the other factors also discussion took place, these will be analyzed in the data analysis chapter of this research.
e. Would you use this model and in what situation would you apply it?	It could be useful in a firm; the factors are well determined but it could be developed with more factors. Some adjustments could be made on the maturity levels. When we come across, many errors within our data, and have many problems, each new problem and specific problems can be added to the factors we face these problems. So that the model is continuously developing. Within certain data	Yes, I would use this model. However, a problem arises at the **first factor**: *"The maintenance of applications is person dependent"*. Based on the maturity levels, team is a higher maturity level, however the problem is we have 300-400 applications, wherein 50 are line of business applications where you should test this model in. You can't put a team that maintains all the applications	A number of questions could be applicable within practice, when we know that a company is working correctly with their data and they know what data is stored where and how we could gather it, we need to know questions with regards to gathering the correct data. With the use of your model we could tell our customers that they first need to tackle their data management issues, before trying to make

	quality management issues, this model would be really useful.	technically and functionally. There are always 2 or 3 persons responsible, which are situated within a team of functional maintenance. With the help of your model and scorecard, we can discuss certain points and you are really talking about some points about data quality management and that gives you the possibility to give wise remarks, that is the use of it.	analysis out of it. We can point to your model as an attention point so that firms know what they have to do. “Your maturity model also determines the complexity of the projects that we conduct.” (I9-I)

References

Babad, Y. M., & Hoffer, J. A. (1984). Even no data has a value. *Commun. ACM*, *27*(8), 748–756. https://doi.org/10.1145/358198.358204

Babbie, E. (2012). *The Practice of Social Research*. Cengage Learning, Inc.

Batini, C., Cappiello, C., Francalanci, C., & Maurino, A. (2009). Methodologies for data quality assessment and improvement. *ACM*, *41*(3), 16:1–16:52. https://doi.org/10.1145/1541880.1541883

Bryman, A. (2013). *Social research methods*. *Oxford university press*. https://doi.org/10.1017/CBO9781107415324.004

Cai, L., & Zhu, Y. (2015). The Challenges of Data Quality and Data Quality Assessment in the Big Data Era. *Data Science Journal*, *14*(2), 1–10.

Cappiello, C., Francalanci, C., & Pernici, B. (2004). Data quality assessment from the user ' s perspective ASSESSMENT FROM THE USER. *ACM*, 68–73.

Chang, V. (2014). The Business Intelligence as a Service in the Cloud. *Future Generation Computer Systems-the International Journal of Grid Computing and Escience*, *37*, 512–534. https://doi.org/10.1016/j.future.2013.12.028

Chengalur-Smith, I. N., Ballou, D. P., & Pazer, H. L. (1999). The impact of data quality information on decision making: an exploratory analysis'. *IEEE Transactions on Knowledge and Data Engineering, Vol. 11, No.*, *6pp*(6), 853–864.

Cong, G., Fan, W., Geerts, F., Jia, X., Ma, S., Zhang, Z., … Ma, S. (2007). Improving data quality: consistency and accuracy. *ACM*, *Vienna*, *Au*, 315–326.

Creswell, J. W. (2013). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. SAGE Publications.

Crossman, A. (2016). The Purposes Of Research. Retrieved June 30, 2016, from http://sociology.about.com/od/Research/a/Purposes-Of-Research.htm

De Bruin, T., Freeze, R., Kaulkarni, U., & Rosemann, M. (2005). Understanding the main phases of developing a maturity assessment model. *In Proceedings of the Australasian Conference on Infor-Mation Systems (ACIS)*, (December).

Eckerson, W. W. (2002). Data quality and the bottom line. *TDWI Report, The Data Warehouse Institute*, 1–32.

Edwards, R., & Holland, J. (2013). *What is Qualitative Interviewing? "What is?" Research Methods Series*. https://doi.org/10.5040/9781472545244

Fischer, R., Aier, S., & Winter, R. (2007). A Federated Approach to Enterprise Architecture Model Maintenance. *Enterprise Modelling and …*, *2*(2), 14. https://doi.org/10.1145/253260.253294

Graham, D. (2015). How To Make A Bad Data-Driven Decision In Three Easy Steps. Retrieved May 17, 2016, from http://www.forbes.com/sites/teradata/2015/07/08/how-to-make-a-bad-data-driven-decision-in-three-easy-steps/#689519a66920

Haug, A., Zachariassen, F., & van Liempd, D. (2011). The costs of poor data quality. *Journal of Industrial Engineering and Management*, *4*(2), 168–193. https://doi.org/10.3926/jiem.2011.v4n2.p168-193

Hoffer, J. A., Prescott, M.B., Topi, H. (2008). *Modern Database Management* (Ninth). Pearson International Edition.

Karr, A. F., Sanil, A. P., & Banks, D. L. (2006). Data quality : A statistical perspective. *Statistical Methodology*, *3*(2), 137–173. https://doi.org/10.1016/j.stamet.2005.08.005

Lee, Y. W., Strong, D. M., Kahn, B. K., & Wang, R. Y. (2002). AIMQ: A methodology for information quality assessment. *Information and Management*, *40*(2), 133–146. https://doi.org/10.1016/S0378-7206(02)00043-5

Mantha, B. (2014). Five Guiding Principles for Realizing the Promise of Big Data. *Business Intelligence Journal*, *19*(1).

Marsh, R. (2005). Drowning in dirty data? It's time to sink or swim: A four-stage methodology for total data quality management. *Journal of Database Marketing & Customer Strategy Management*, *12*(2), 105–112. https://doi.org/10.1057/palgrave.dbm.3240247

Mayring, P. (2000). Qualitative Content Analysis. *Forum: Qualitative Social Research*, *1*(2).

McAfee, A., & Brynjolfsson, E. (2012). Big Data. The management revolution. *Harvard Buiness Review*, *90*(10), 61–68. https://doi.org/10.1007/s12599-013-0249-5

Merino, J., Caballero, I., Rivas, B., Serrano, M., & Piattini, M. (2015). A Data Quality in Use model for Big Data. *Future Generation Computer Systems*. https://doi.org/10.1016/j.future.2015.11.024

Miles, J., & Gilbert, P. (2005). *A Handbook of Research Methods for Clinical and Health Psychology*. Oxford University Press. Retrieved from https://books.google.nl/books?id=kmZ3Yt5pY0YC&dq=semi-structured+interview&lr=&hl=nl&source=gbs_navlinks_s

PacificWorlds. (2016). Agreement ot Participate in Interview. Retrieved November 9, 2016, from www.pacificworlds.com/homepage/education/essays/consent.doc

Paulk, M. C., Curtis, B., Chrissis, M. B., & Weber, C. V. (1993). Capability maturity model, version 1.1. *IEEE Software*, *10*(4), 18–27.

Pipino, L. L., Lee, Y. W., Wang, R. Y., Lowell Yang Lee, M. W., & Yang, R. Y. (2002). Data Quality Assessment. *Communications of the ACM*, *45*(4), 211. https://doi.org/10.1145/505248.506010

Pöppelbuß, J., & Röglinger, M. (2011). What makes a useful maturity model? a framework of general design principles for maturity models and its demonstration in business process management. *ECIS*, 13.

Redman, T. C. (1996). Improve data quality for competitive advantage. *QUALITY CONTROL AND APPLIED STATISTICS*, *41*, 49–52.

Romero, D., & Vernadat, F. (2016). Enterprise information systems state of the art: Past, present and future trends. *Computers in Industry*, *79*, 3–13. https://doi.org/10.1016/j.compind.2016.03.001

Rouse, M. (2005). Definition: Data Quality. Retrieved August 1, 2016, from http://searchdatamanagement.techtarget.com/definition/data-quality

Ryu, K., Park, J., & Park, J. (2006). A Data Quality Management Maturity Model. *ETRI Journal*, *28*(2), 191–204. https://doi.org/10.4218/etrij.06.0105.0026

Silverman, D. (2013). *Doing qualitative research: A practical handbook.* SAGE Publications Limited.

Strong, D. M., Lee, Y. W., & Wang, R. Y. (1997). Data Quality in Context. *Communications of the ACM*, *40*(5), 103–110. https://doi.org/10.1017/CBO9781107415324.004

Tayi, G. K., & Ballou, D. P. (1998). Examining data quality. *Communications of the ACM*, *41*(2), 54–57. https://doi.org/10.1145/269012.269021

Veritas. (2015). THE DATABERG REPORT: SEE WHAT OTHERS DON'T. Retrieved from https://www.veritas.com/content/dam/Veritas/docs/reports/veritas-strike-global-report_a4-sdc2.pdf

Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: what data qualty means to data consumers. *Journal of Management Information Systems*, *12*(4), 5–33.

Xu, H., Horn Nord, J., Brown, N., & Daryl Nord, G. (2002). Data quality issues in implementing an ERP. *Industrial Management & Data Systems*, *102*(1), 47–58. https://doi.org/10.1108/02635570210414668