Universiteit Leiden

ICT in Business

Measuring data governance:
a structured method to
assess data governance maturity

Name: Lucas Ingmar van der Meer
Student-no: 0811076

Date: 26 June 2015

1st supervisor: Prof. Dr. Aske Plaat
2nd supervisor: Dr. Hans Le Fever

MASTER'S THESIS

Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands

MEASURING DATA GOVERNANCE:

A STRUCTURED METHOD TO ASSESS DATA GOVERNANCE MATURITY

Master's thesis
Lucas Ingmar van der Meer

s0811076
lucasvandermeer@gmail.com

Version:	2.0 (Final)
Date	26 June 2015
First reader:	Prof. Dr. Aske Plaat (Leiden University)
Second reader:	Dr. Hans Le Fever (Leiden University)
Company supervisors:	Prof. Dr. Joost Visser (Software Improvement Group and Radboud University Nijmegen)
	Christiaan Ypma MSc (Software Improvement Group)

Leiden University
Leiden Institute of Advanced Computer Science

SCIENTIFIC ABSTRACT

Data governance

Data is value, and more and more organisations are realising the value that is hidden in their data (Logan 2012). In order to effectively and efficiently use data, it should be properly governed (Weber et al. 2008). We believe that assessing data governance maturity can help organisations that are new to data governance and that consider to implement a data governance program. Also organisations that wish to improve their existing program will benefit from a method to assess their data governance maturity. To the best of our knowledge, no such method exists to date.

Data governance maturity method

Based on a literature study, we derive the following requirements for a method to assess data governance maturity. (1) It should produce a maturity score and capture the full complexity of data governance. (2) It should provide a means to interpret the results and to provide actionable recommendations. (3) The method should be accepted by users and provide objective results. We adopted a qualitative, design science approach to develop a method that meets most of these requirements. The method consists of a data governance maturity model, and instructions for how to use the model.

The maturity model describes the elements that data governance maturity consists of, how they relate to each other and how maturity is computed. We created the model using Factor-Criteria-Metric (Marinescu 2005). It consists of two dimensions: (1) elements that are present in any governance (with the classes objectives, tasks, roles, and responsibilities), and (2) domains that are specific to data governance (with the classes data assets, data quality, metadata, data access, and data lifecycle). Each class contains two to seven questions.

The method goes beyond a model describing data governance maturity. It provides instructions for interpreting maturity results to derive actionable recommendations. Moreover, the method provides a template for data collection, execution, validation, decision-making, diffusion, and the assignment of roles.

Results

The method was evaluated in two case studies at a large and a small IT-intensive organisation. Due to time constraints, we could evaluate the method at only fifteen persons in these organisations, meaning that the quantitative results will not be significant. However, these quantitative results are supported by remarks that the persons gave us during the case studies.

In the case studies, we used the method to assess the organisations' data governance maturity and to provide actionable recommendations. These recommendations were perceived as useful: 73% of 15 participants in the evaluation sessions found the model valid. 74% of 10 participants reported that they would use the method to assess data governance maturity. 10 participants also used the method to assign a maturity score to their organisation. Although there is variation in participants' answers, they tend to agree. This is reflected in an inter-rater reliability of 0.4.

These results suggest that our method is a useful step towards assessing data governance maturity. Participants found, however, that the method currently provides insufficient support to interpret the results. Therefore, in future work we aim to structure the interpretation step with an interpretation framework that connects organisational data objectives to maturity scores. A methodological limitation is that the method was evaluated at only two organisations with a total of 15 participants; one should therefore not generalise the results until the method has been applied at substantially more organisations. We aim to validate the method at a multitude of divergent organisations.

MANAGEMENT SUMMARY

Data governance

Treating data as a valuable organisational asset increases the possibilities to use data, for instance to ensure compliance and enable better decision-making (Otto 2011). Organisational assets require some form of governance, which is why organisations are adopting data governance programs. Data governance programs can result in a better management of data quality (Cheong & Chang 2007). In turn, high data quality can result in a greater confidence in analytics systems, less time spent on reconciling, data and reduced costs.

Data governance programs help in agreeing about clear roles and responsibilities among data stakeholders (Cheong & Chang 2007), positively affecting the efficiency with which data is processes and the effectiveness with which it can be used. A close collaboration of the business and IT is one of the core aspects of data governance (i.a. Weber et al. 2009; Kooper et al. 2011). IT should be responsible for making sure data is securely stored and is retrievable, whereas the business is responsible for the data's content.

Data governance maturity method

We created a method to assess organisations' data governance maturity. A number of questions is to be answered on two dimensions: (1) elements that are present in any governance (with the classes objectives, tasks, roles, and responsibilities), and (2) domains that are specific to data governance (with the classes data assets, data quality, metadata, data access, and data lifecycle). The answers to the questions determine the organisational maturity score.

Assessing data governance maturity can help organisations that are new to data governance and that consider to implement these a data governance program. Also organisations that wish to improve their existing program will benefit from a method to assess their data governance maturity. To the best of our knowledge, no such method exists to date.

The method makes gives instructions how to interpret the model's outcomes and to derive actionable recommendations. The method also provides a template for data collection, execution, validation, decision-making, diffusion, and the assignment of roles. Assessing data governance maturity can be done in about 50 hours.

Recommendations

Organisations should never strive for 100% on all classes, but find the right match between needs and how much effort it takes to get there. In general, organisations should adhere to the following guidelines. (1) Governing data begins with knowing what data assets there are, and how these are used in the business processes. This means that an architecture of the data assets should be made. Moreover, these data assets should be prioritised according to confidentiality, integrity and availability. (2) All levels of responsibilities in managing data should explicitly be assigned. This includes ownership, responsibility for day-to-day operations, and data stewardship. (3) The organisation's management should continuously stretch the importance of data and ensure that persons have time to properly manage data. (4) Organisations should draft guidelines for data quality and metadata. These are the first steps to monitor data quality and to create a common understanding of data.

Results

We tested the method at a large and a small organisation. 73% of 15 participants in the evaluation sessions find the model valid. People value most its structured explanation of the current state and its actionable recommendations. 74% of 10 participants reported that they would use the method to assess data governance maturity. Participants found, however, that the method currently provides insufficient support to interpret the results. Therefore, in future work we aim to structure the interpretation step with an interpretation framework that connects organisational data objectives to maturity scores.

TABLE OF CONTENTS

Scientific abstract ... **5**
Management summary ... **6**
Table of Contents ... **7**
1 Introduction ... **9**
1.1 Data governance ... 9
1.2 Legitimisation ... 9
1.3 Problem statement ... 10
1.4 Method & research context ... 10
1.5 Contribution ... 10
1.6 Thesis outline ... 11
2 Research design ... **12**
2.1 Method ... 12
2.2 Research process ... 12
2.3 Step 1: Literature study and exploration ... 13
2.4 Step 2: Design of the first version ... 15
2.5 Step 3: First evaluation at Software Improvement Group ... 15
2.6 Step 4: Diffusion at Software Improvement Group ... 15
2.7 Step 5 and 6: Analysis of feedback and improvement ... 16
2.8 Step 7: Second evaluation at Bank and ANWB ... 16
2.9 Step 8: Diffusion at Bank, ANWB, and the community ... 19
2.10 Participating organisations ... 20
3 Data governance ... **23**
3.1 Data as a valuable organisational asset ... 23
3.2 Governance over data assets ... 25
3.3 Governance vs. management ... 27
3.4 Data vs. information ... 28
3.5 Data quality ... 28
3.6 Enterprise architecture ... 30
3.7 Relationship with IT and corporate governance ... 31
4 Assessing data governance maturity ... **33**
4.1 Maturity assessments ... 33
4.2 Requirements of a data governance maturity assessment ... 33
4.3 Models ... 36
4.4 Comparison of the models ... 38
5 Data governance maturity method ... **40**
5.1 Maturity model ... 40
5.2 Structure of the model ... 41
5.3 Metrics of the model ... 49
5.4 Questions of the model ... 52
5.5 Instructions ... 56
6 Results ... **63**
6.1 Collection of the results ... 63
6.2 User groups and organisations ... 63
6.3 Construct validity ... 64
6.4 Acceptance ... 68
6.5 Inter-rater reliability ... 71
6.6 Requirements of a data governance maturity method ... 72
6.7 Governance elements ... 73
6.8 Data governance domains ... 74
6.9 Data governance at SIG, Bank and ANWB ... 75
6.10 Data governance & organisational change ... 76
6.11 Data governance organisation ... 76
6.12 Data for innovation ... 77
6.13 Research method results ... 77

7 **Conclusions** ... **79**
7.1 Research question 1 – Data governance ... 79
7.2 Research question 2 – Requirements to assess data governance maturity ... 80
7.3 Research question 3 – Data governance maturity method ... 81
7.4 Other conclusions ... 82
7.5 Discussion ... 83
7.6 Future research ... 85
7.7 Final conclusion and considerations ... 86
8 **References** ... **87**
Appendices ... **90**
A **Publication** ... **90**
A.1 Abstract ... 90
A.2 Introduction ... 90
A.3 Research method ... 91
A.4 Data governance ... 92
A.5 Assessing data governance maturity ... 93
A.6 Results ... 96
A.7 Discussion ... 100
B **Literature review** ... **102**
B.1 Collected sources ... 102
B.2 Test search ... 102
C **First version of the method** ... **105**
C.1 Results of the first version of the method ... 105
D **Evaluation questionnaire** ... **108**
D.1 Validity questions ... 108
D.2 Acceptance questions ... 108
E **Case studies** ... **109**
E.1 Functions we requested to talk to ... 109
E.2 Questions asked at the case studies ... 109
E.3 Participants' functions ... 111
F **Confidential appendices** ... **113**

1 INTRODUCTION

1.1 Data governance

A high percentage of organisations believe that data can be a valuable asset for them (Economist Intelligence Unit 2008; Logan 2012). The reason is that data can be used for instance to ensure compliance and enable better decision-making (Otto 2011). Organisational assets require some form of governance, which is why organisations are adopting data governance programs. Data governance is involved with setting a decision-making framework that is appropriate for meeting an organisation's data objectives (Khatri & Brown 2010; Weber et al. 2009). Research shows that formal data governance programs help in increasing data quality (Cheong & Chang 2007) and that there is a positive correlation between a company's commitment to governing data, and its capacity get value out of data assets (Economist Intelligence Unit 2008).

From an IT perspective, data needs to be stored appropriately and several data sources need to be connected to collect relevant data. From a business perspective, the data needs to be interpreted and given context. Therefore, a close collaboration of the business and IT is one of the core aspects of data governance (Cheong & Chang 2007).

1.2 Legitimisation

Gartner stated that in 2012 businesses realised that data can create business value (Logan 2012). In that same year, data governance[1] first appeared on Gartner's hype cycle and has been on the *peak of inflated expectations* ever since (Logan 2012). This indicates that data governance is popular yet immature.

1.2.1 Science

Whereas an area such as IT governance is widely covered in scientific literature, data governance is a relatively new discipline. A search on the large scientific collection Google Scholar on the keyword *data governance* yields 4,400 results, compared to 34,500 for *IT governance*. The outlines of data governance seem to have been drawn, but are largely based on IT governance knowledge. Data governance-specific knowledge comes from a handful of case studies and framework analyses (Tallon et al. 2013). Knowledge that is missing includes an elaborate analysis of the interaction of roles and responsibilities (Weber et al. 2009) and a scientifically evaluated method to assess data governance maturity (see chapter 4).

1.2.2 Society

In a 2005, The Data Warehousing Institute (TDWI; Russom 2006) conducted a survey amongst professionals involved with data quality. Respondents were predominantly in the financial services (26%) and IT consultancy (10%) and predominantly located in the United States (62%). TDWI found that only 25% of the organisations have a data governance program[2]. In 2008, the Economist Intelligence Unit (2008) asked the same question to corporate executives spread evenly around the world and found that 38% of the organisations have such a program[3]. Although both surveys have different demographics and date from different years, it is clear that data governance is far from implemented in all organisations.

Given the promises of a formal data governance program (section 1.1), it is not surprising that the number of organisations that have a program or are considering one is increasing (Economist Intelligence Unit 2008).

---

[1] Gartner uses *information governance*, but uses a definition similar to our definition of *data governance*. We consider the two concepts to be the same (see section 3.4).

[2] 49% of the studied organisations have an initiative to improve the data quality. Of these initiatives, 52% is a data governance program. Assuming there are no data governance programs excluding data quality (see section 3.5), we can conclude that 25% of the organisations have a data governance program.

[3] The EIU uses *information governance*, which we consider the same (see section 3.4).

However, organisations find implementing a program difficult: 34% of the organisations that would like to start a data governance program do not know where to begin (Economist Intelligence Unit 2008). In any organisation, regardless of a formal program, ungoverned data practices will be in place. Assessing these ungoverned data practices will help organisations in determining areas for improvement. Other big challenges for implementing a formal program are determining the costs, risks and returns of managing information company-wide (40%) and enforcing policies company-wide (39%). Because of these difficulties, 43% of the organisations considers seeking assistance from an outside organisation, or had already sought such assistance (Economist Intelligence Unit 2008).

As the world is moving to a more data-intensive world, a good method to assess data governance maturity is highly important.

1.3 Problem statement

We want to assess the current state of data governance in an organisation and suggest areas for improvements. Maturity models are designed for that purpose: they provide a means to assess the maturity (i.e. competency, capability, level of sophistication) of a selected domain. Most maturity models also have prescriptive aspects, identifying actions to increase maturity (De Bruin et al. 2006). We focus on organisations for which data is most relevant: IT-intensive organisations. Our main research goals is finding out *how to assess data governance maturity of IT-intensive organisations*.

The research is split in three parts. First, we will explore data governance in the context of related research areas such as IT governance. Next, we will list requirements for a data governance maturity method, after which we explain design of the method and how we tested what end users think of it. The three research questions are:

Research question 1. How is data governance placed in the context of data management, information, data quality, enterprise architecture, IT governance and corporate governance?

Research question 2. What are requirements for a method that assesses data governance maturity, and do existing data governance models meet these requirements?

Research question 3. Can a new method help intended end users assess their data governance maturity, and how do they value that new method?

The first question is answered in chapter 3, the second in chapter 4 and the third in chapter 5 and 6.

1.4 Method & research context

We adopted a qualitative, design science approach. We designed a first version of our model, which we pre-tested using a case study. Based on the pre-test, we updated the model and tested it more fully in two more extensive case studies. More explanation of the research method is given in chapter 2.

The research was conducted by Leiden Centre of Data Science (Leiden Institute of Advanced Computer Science), Leiden University, and facilitated by Software Improvement Group.

1.5 Contribution

To the best of our knowledge, this thesis presents the first scientifically evaluated method to measure data governance by assessing its maturity[4]. In two case studies, we found that most people (n=15) find this new

[4] Although another maturity method exists, this method is specific to data quality management and does not cover all of data governance (Hüner et al. 2009).

method to assess data governance maturity valid (73%) and that most people (n=10) would use it for that purpose (74%).

1.6 Thesis outline

The thesis is structured as follows. Chapter 2 discusses the research design, chapter 3 provides a literature overview of data governance, and chapter 4 lists requirements for a data governance maturity method. In chapter 5, we discuss the method we developed; the results of testing the method are indicated in chapter 6. Finally, chapter 7 concludes the research and gives suggestions for future research.

2 RESEARCH DESIGN

Chapter 1 describes the research design, including the methodology, process, and participating organisations.

2.1 Method

We adopted a qualitative, design science approach using case studies. This is an appropriate research method in areas that had few previous researches, such as data governance (Cheong & Chang 2007). Case studies are focussed on understanding the nature and complexity of an attribute (Cheong & Chang 2007), in our case data governance maturity. Case studies are qualitative in nature (Baxter et al. 2008), but we will enrich them with quantitative data.

2.1.1 Design science paradigm

Design science is a research paradigm that involves designing innovative artefacts related to information systems, and consequently analysing how these artefacts are used (Kuechler & Vaishnavi 2008). Artefacts include terminologies, models and methods (Österle et al. 2010).

Design science research is executed in close collaboration with organisations that are using the information systems that are being researched (Österle et al. 2010). It is usually focussed on explaining and improving the current situation, whereas classical explanatory research tends to stop at explaining the issues. Design science can therefore be considered more pragmatic than classical explanatory research, without losing scientific rigidness (Van Aken 2005).

Improving any situation tends to take many iterations, making design science iterative (Kuechler & Vaishnavi 2008). Our research consists of two iterations of the four basic design science phases (Österle et al. 2010): analysis, design, evaluation, and diffusion.

- ***Analysis*** includes understanding the scientific as well as business relevance, therefore, it is common to include both scientific literature and material written by practitioners ("grey literature"; Denyer et al. 2008);
- The ***design*** phase involves designing the innovative artefact, in our case a method to assess data governance maturity;
- In design science, much effort is put into ***evaluating*** how intended end users respond to the created artefact. In our case, we conducted case studies and subsequent evaluation sessions;
- ***Diffusion*** involves spreading the results to relevant stakeholders.

Design science has been applied to research areas that are related to data governance, such as data warehousing (Sen et al. 2012) and data quality management (Hüner et al. 2009). However, to the best of our knowledge, it has not been applied to data governance research specifically.

2.2 Research process

Our research followed two full iterations of the four design science phases (Figure 1). We first analysed the field using a literature study, after which we built a first version of our data governance maturity method. We then evaluated the method using a case study and evaluation session, after which we diffused our results. In the second iteration, we analysed feedback we gathered and built a second version of the method. The method was tested more fully in two more extensive case studies. The results were again diffused.

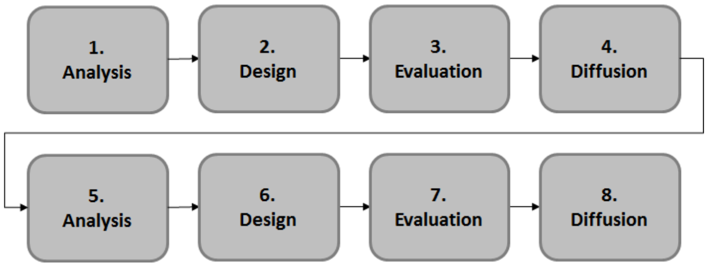


Figure 1 – The research follows two cycles of the design science phases that are proposed by Österle et al. (2010)

The process is further elaborated on in the following sections; the results are discussed in other chapters (Table 1).

Design science phase	*Itera-tion*	*Short description*	*Discussion process*	*Discussion results*
1. Analysis	1	Literature study and exploration	Section 2.3	Chapter 3 and 4
2. Design	1	First version using a meta literature study	Section 2.4	Appendix C
3. Evaluation	1	Case study at SIG	Section 2.5	Appendix F[5]
4. Diffusion	1	At SIG: Report, presentation, workshop	Section 2.6	Appendix F[5]
5. Analysis	2	Study the SIG case study results	Section 2.7	Appendix F[5]
6. Design	2	Second version using results of the analysis	Section 2.7	Chapter 5
7. Evaluation	2	Two case studies at Bank and ANWB	Section 2.8	Chapter 6
8. Diffusion	2	Report, presentation, workshop, institute website	Section 2.9	Appendix F[5]

Table 1 – Sections in which the process and results of the design science phases are discussed

2.3 Step 1: Literature study and exploration

We first explored data governance and conducted a literature review. Since this yielded insufficient literature, we contacted authors in the field. They verified that few papers specific to data governance are available, but could send us some related or unpublished work. After this, we believed to have gathered sufficient literature to start exploring existing models and construct our own.

2.3.1 Exploring data governance and related fields

To get an overview of data governance, we discussed it with employees of the organisation that facilitated the study, *Software Improvement Group (SIG)*. Software Improvement Group is an IT advisory organisation that helps large enterprises improving their software. Since people at Software Improvement Group are knowledgeable in pitfalls of IT, understand governance and know the value of data, they were believed to be valuable sources.

We also explored available literature to obtain an initial overview of data governance and related areas. The study spanned *governance, data governance, information governance, IT governance, corporate governance, data governance maturity, enterprise information management, data management*, *data quality,* and *enterprise architecture*. Two types of sources were used: grey and scientific literature.

Grey literature – Grey literature includes sources from practitioners, for example survey results from The Data Warehouse Institute (2006) and a handful of data governance (maturity) models that are created by commercial organisations (see chapter 4). Although these sources are usually not scientific, including them is valuable in design science as it helps in understanding the importance of data governance and its usage and challenges in organisations (Denyer et al. 2008).

[5] Confidential appendix.

Scientific literature – In this exploratory literature study, we searched scientific literature primarily using Google Scholar using the aforementioned keywords. We conducted the searches twice: once without restrictions, and once using a filter on work that has been published in the last two years. We sorted papers on relevance and selected by the presence of the keywords in the title. In papers that we found, references relevant to our research were further explored. In total, we collected about a dozen of relevant scientific literature.

We concluded that we found sufficient literature about fields related to data governance, such as data quality and IT governance, but that we needed more specific information about data governance and available data governance maturity models.

2.3.2 Scientific literature review for data governance maturity methods

After exploring the field, we now focused specifically on data governance maturity models using structured literature review practices. We followed parts of the process as suggested by Kitchenham (2004). She suggests to start conducting a systematic literature study by finding sources and executing a test search. After that, one should select studies to explore and perform a quality assessment on these studies. An extensive quality assessment consists of looking for frequently occurring biases, checking internal and external validity, and determining quality thresholds. Finally, one should extract and synthesis data.

Our goal is not to perform a systematic literature review since is out of the scope of this master's research, but to gain sufficient knowledge from data governance and related fields to design and evaluate our method. Therefore, we selected a limited number of studies to explore and performed a basic quality assessment.

2.3.2.1 Finding sources

We looked where the most highly ranked articles in Google Scholar were published when searching for "data governance". Quotes were included to exclude papers that contain the two words but do not discuss the field of governing data[6]. We also listed all journals in which scientific literature we already found were published. Lastly, we did regular internet searches for "data governance" journals, "information governance" journals, "data management" journals, "information management" journals, "data quality" journals and "data governance" organisations. A list of the collected sources is attached in Appendix B.1.

2.3.2.2 Test search

In a structured literature study, it is common to start with a test search, of which the results are used to determine the next steps. The raw results of our test search are attached in Appendix B.2. We conducted two test searches using the following queries on the sources we found in the previous step:

1. ((data || information) + (governance))
2. ((data || information) + (governance)) + {audit, assessment, evaluation, scoring, scorecard, maturity}.

The first test search resulted in one more peer-reviewed paper (Otto 2011). The second test search resulted in zero papers new papers.

Since only one new papers was found, we decided to terminate our literature review and contact authors in the field.

2.3.3 Contacting authors in the field

We contacted the Competence Center Corporate Data Quality (CC CDQ) at the University of St. Gallen, Switzerland, an institute that is responsible for three scientific papers we collected (Weber et al. 2009; Otto

[6] For instance the paper: "Corporate governance: decades of dialogue and data."

2011; Hüner et al. 2009). A research assistant sent us the dissertation of Weber (2009) about a data governance reference model. The dissertation is accepted, but not online available. The dissertation provides fundamental data governance knowledge and references to a significant number of papers, of which about half was found already. Considering these new papers, we believed to have sufficient literature to start designing the first version of our method.

2.4 Step 2: Design of the first version

In our exploratory phase, we found that no satisfactory method to establish data governance maturity is available (see chapter 4). Based on our literature study and input gathered from Software Improvement Group consultants, we constructed a first version of our data governance maturity method. It is attached in Appendix C.

2.5 Step 3: First evaluation at Software Improvement Group

Measuring what end users think of a method by using it and consequently improving that method is a crucial part of design science. To experience using the method and to get an initial idea about its validity, we pre-tested the method with a case study including an evaluation session.

Both were conducted at the organisation that facilitated the research, Software Improvement Group. Persons at Software Improvement Group are generally highly educated, knowledgeable in IT, understand governance and know the value of data. Although this creates a bias, we believe it is a valid approach for a *pre*-test. Later case studies are conducted using a more representative group of people. More information about Software Improvement Group can be found in sub-section 2.10.1.

2.5.1 Case study

To experience using the method and to gather data for the evaluation sessions, we conducted a case study. We started with studying the existing data asset overview, which was made to be compliant with the ISO 27000 standard for information security (ISO 2012). Then we talked to persons that are directly or indirectly involved with the organisation's most valuable data asset. The participants' functions are attached in Appendix E.3.1.

2.5.2 Evaluation sessions

We conducted evaluation sessions as part the case study. Participants were nine Software Improvement Group employees of which most contributed to the prior case study. In individual sessions that lasted about an hour, we briefly explained the participants some terminology and the developed method.

Because a group of persons that want to assess data governance maturity was not readily available, we asked the participants to emphasise that they were asked by an executive to assess data governance maturity for their own organisation. Consequently, they experienced the method by using it to rate maturity for their own organisation. Then, we showed them their results based on the filled in scorecard, and compared these with the authors' case study results. The participants were encouraged to ask questions and to provide feedback.

2.6 Step 4: Diffusion at Software Improvement Group

To stimulate actual usage of results, diffusion of these results is important (Österle et al. 2010). We presented the results to the organisation, wrote a report on our findings and organised an improvement workshop.

Presentation of the case study results – Preliminary results were presented to Software Improvement Group, which resulted in valuable feedback.

Interpretation report – After concluding the case study, we wrote a report of about fourteen pages on our findings and recommendations. More information on the report can be found in sub-section 5.5.4. The report is attached as a confidential attachment (Appendix F).

Improvement workshop **–** To stimulate usage of the recommendations, we organised a two-hour improvement workshop. Five persons, including two directors, attended the workshop. More information can be found in sub-section 5.5.7. Slides of the workshop are attached as confidential attachment (Appendix 5.5.7).

2.7 Step 5 and 6: Analysis of feedback and improvement

We analysed the results of the evaluation session and the feedback we got about the case study, presentation of the results, the report and the improvement workshop. Based on these, we updated the method.

2.8 Step 7: Second evaluation at Bank and ANWB

We conducted a second evaluation round, consisting of two case studies that included evaluation sessions. The case studies followed a similar process as the first case study, but were conducted at two other organisations and using a more representative group of people. The evaluation session were extended with quantitative measurements.

2.8.1 Preparation process

The two organisations are Bank[7] and ANWB. At each, we were assigned an employee as our primary contact. We handed these persons a list with function we would like to talk to and in which order (Appendix E.1), after which the contact and us jointly made a list of appropriate persons to interview. The list contained seven persons at Bank and eight at ANWB.

The list with people was sent to the secretariat that planned the case study and subsequent evaluation interviews. Between the case study and evaluation interviews were about two weeks for us to analyse the case study results and write a report with recommendations.

2.8.2 Case study process at both Bank and ANWB

Using experiences from the previous case study, we structured the process by talking to functions in a specific order. We first talked to executive management about the company's direction and its primary business processes. Next, we investigated how important data to the organisation is and what the most important data assets are. We talked to the data assets' owners and connected these data assets to applications and business processes. Information about the data assets is mainly gathered by talking to Information Architects, Business Architects and Database Administrators. We also talked to persons that are, either explicitly or implicitly, responsible for the specific functional areas of data governance: data quality, metadata, data access, and data lifecycle (section 3.2.3).

Questions that we asked are attached in Appendix E.2. Typical questions are how happy the interviewee is with the current situation of the discussed subject and how he or she would like to improve it. Next, we asked more specific questions, such as if there is a data asset overview. Some important questions, such as who the owner is of data assets, are asked to all persons as control questions.

The interviews were audio recorded and later transcribed and analysed.

[7] This is not the organisation's real name. The organisation does not seek publicity in areas that are not its core business, hence requires to remain anonymous.

2.8.3 Evaluation at both Bank and ANWB

Like the first evaluation sessions, we asked the participants to emphasise that they were asked by an executive to assess data governance maturity for their own organisation. Consequently, they used the method to rate maturity for their own organisation.

In the earlier evaluations, we only showed the participants a chart with the numeric outcomes of the method. We noticed that this gives few context and interpretation, which makes it hard to understand the implications and to derive actionable recommendations. To better prepare the persons for the session, we drafted a version of our findings, including an explanation of the method, an elaboration of the results and prioritised recommendations. We sent the report at least two working days in advance.

We also added quantitative measurements for ***construct validity***, ***acceptance***, and ***inter-rater reliability***. Since the total number of participants is only fifteen or ten, depending on the test, we will not be able to draw statistically significant conclusions from the results. Therefore, the results are only used to enrich the qualitative results.

2.8.3.1 Quantitative measurements

An overview of how each experiment is measured is given in Table 2. Our observations while the participants were assessing data governance maturity for their own organisation, and the discussion we had about the results, form the basis for the qualitative results.

Subsequently, we asked the participants to answer some questions about acceptance and construct validity, which form these quantitative results. These questions are combined in an online questionnaire (Appendix D).

Quantitative results for inter-rater reliability are calculated afterwards using filled-in scorecards.

Test	*Quantitative measurement*	*Qualitative measurement*
Acceptance	Interview and discussion	Questionnaire (Likert)
Construct validity	Interview and discussion	Questionnaire (Likert)
Inter-rater reliability	-	Intraclass correlation (ICC)

Table 2 – The three tests we conducted on the method

Acceptance and construct validity indicate different phases of using the method (Figure 2). First, an assessor will assess an organisation's data governance maturity using the method's scorecard; acceptance indicates whether persons would accept using that scorecard (Riemenschneider et al. 2002). Next, the assessment's results should be used to improve the organisation's maturity*;* construct validity indicates if these results reflect reality.

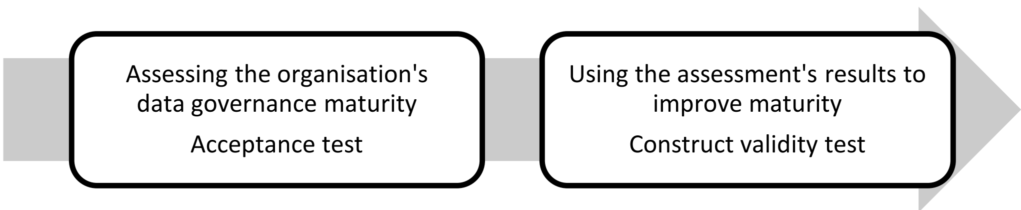


Figure 2 – Different phases of the data governance maturity method in which the acceptance and construct validity test are used

2.8.3.2 Acceptance

We based our acceptance measurement on Riemenschneider et al. (2002), who found that the five most important determinants of intention to use a new software development methodology are usefulness, ease of use, voluntariness, compatibility, and subjective norm (Table 3). Riemenschneider et al. also provide questions to measure these five determinants.

Given the close strong connection between data and software (section 3.6 and 3.7) and the lack of an acceptance method specific to data methodologies, we believe these determinants and questions are useful for our acceptance measurement.

Acceptance determinant	*Description (Riemenschneider et al. 2002)*
Usefulness	The extent to which the person thinks using the system will enhance his or her job performance.
Ease of use	The extent to which the person perceives using the system will be free of effort.
Voluntariness	The extent to which potential adopters perceive the adoption decision to be non-mandatory.
Compatibility	The degree to which an innovation is perceived as being consistent with the existing values, needs, and past experiences of potential adopters.
Subjective norm	The degree to which people think that others who are important to them think they should perform the behaviour.

Table 3 – Overview of the acceptance criteria

Each determinant was measured using three to six questions on a five-point Likert scale (Strongly disagree – disagree – neutral – agree – strongly agree). The answers from all participants within a determinant are aggregated using a proportion table (Table 4 and Table 5) and illustrated using a vertical stacked bar chart.

	Usefulness question 1	*Usefulness question 2*	*Usefulness question 3*	*Ease of use question 1*	*Ease of use question 2*	*Ease of use question 3*
Participant A	4	5	1	2	3	5
Participant B	1	3	3	1	1	1
Participant C	2	4	4	4	4	5

Table 4 – Example answers to the acceptance questions

Value	*Usefulness frequency*	*Usefulness proportion*	*Usefulness frequency*	*Usefulness proportion*
1	2	0.22	3	0.33
2	1	0.11	1	0.11
3	2	0.22	1	0.11
4	3	0.33	2	0.22
5	1	0.11	2	0.22
Sum	***9***	***1.00***[8]	***9***	***1.00***[8]

Table 5 – Aggregation of the example acceptance answers

[8] Rounded.

2.8.3.3 Validity

The participants answered questions designed to indicate how well the outcomes of the method reflect data governance maturity within their organisation, if they agree with the authors' score and if they prefer this method over their current method. We constructed the questions ourselves.

All questions were to be answered on a five-point Likert scale (Strongly disagree – disagree – neutral – agree – strongly agree). Since there are only three questions, the answers of all participants were aggregated per question using a proportion table and illustrated using a vertical stacked bar chart.

2.8.3.4 Inter-rater reliability

Inter-rater reliability (IRR) measures the degree of agreement between raters. After the interview sessions, we used data of the filled in scorecards to measure IRR. For such situations, the intraclass correlation (ICC) for IRR is preferred over for instance regular correlation. The reason is that if items linearly correspond, such as in Table 6, their correlation is 1.00 whereas there is no agreement on any question (King's College London n.d.). This is not the case with ICC, where the score is 0.67.

	Rater A	*Rater B*
Question 1	2	3
Question 2	3	4
Question 3	4	5
Correlation(A,B)[9]	1.00	
ICC(A,B)[10]	0.67	

Table 6 – Example intraclass correlation, compared to a 'regular' correlation

According to Shrout and Fleiss (1979), one should decide for the model, type and unit of the ICC calculation.

Model – If each subject is assessed by a different set of randomly selected raters, model 1 *(one-way random)* is chosen. If each subject is assessed by each rater, and raters have been randomly selected, one chooses model 2 *(two-way random)*. Finally, if each subject is assessed by each rater, but the raters are the only raters of interest, model 3 is the right choice (*two-way mixed*).

Type – If differences in judges' mean ratings are of interest, one chooses agreement (type 1). Otherwise, consistency is chosen *(type 2)*.

Unit – If reliability is calculated based on one measurement, the unit is *single*. If it is based on multiple measurements, one selects *average*.

Since each subject is assessed by each rater and these raters are the only raters, our model is *two-way mixed (model 3)*. We want to measure agreement between raters, so our type is *agreement (type 1)*. Finally, raters score only once, which is why our unit is *single*. This correspond to the ICC name ICC(3,1) (Shrout & Fleiss 1979). An ICC of *1* indicates a perfect correlation between the raters, *0* indicates no correlation. We calculated ICC with the package *irr* in the statistics tool *R*.

2.9 Step 8: Diffusion at Bank, ANWB, and the community

To stimulate actual usage of results, we actively diffused them.

[9] Using Pearson correlation

[10] Using ICC(3,1), as discussed below the table.

Publication – We want to spread the results of our research with a congress or journal publication. We drafted a publication according to guidelines of the Association for Information Systems (AIS 2015), which is attached in Appendix A.

Interpretation report – At both Bank and ANWB, we wrote a report with our findings and recommendations, just like at Software Improvement Group. The structure of the report is outlined in sub-section 5.5.4; the reports are attached in Appendix F (confidential).

Improvement workshop – We organised a three-hour improvement workshop at Bank at which nine people attended. An improvement workshop at ANWB could not take place during the research, but is scheduled. The workshop is explained in sub-section 5.5.7; the sheets of the improvement workshops are attached in Appendix F (confidential).

Management team presentation – At Bank, we also presented our results to the management team. Slides are attached in Appendix F (confidential).

Further study – We aim to further study the subject.

Institute website – The thesis, excluding confidential appendices, will be published on the website of the university's institute.

2.10 Participating organisations

This section briefly describes the participating organisations.

2.10.1 Software Improvement Group

Software Improvement Group facilitated the research. An overview of the organisation can be found in Table 7.

Software Improvement Group is a Dutch IT advisory organisation that is specialised in software and has about a hundred employees. The organisation is good at measuring several software dimensions, including quality. Software Improvement Group develops its own IT tools and uses agile methodologies. In agile, development evolves in self-organising, cross-functional teams (Beck et al. 2001). With self-organisation teams comes a lot of responsibility lower in the organisation, having influence on the governance structures. Since Software Improvement Group develops its own tools and provides IT advice for its clients, we label the organisation as IT-intensive.

Software Improvement Group employees generally are experienced in the IT industry, familiar with governance aspects, know the importance of data, and understand practical implications of scientific methods.

The organisation has no formal data governance program and is not considering setting one up. Software Improvement Group has several data sources that are important to the business. In these data sources, there are some data quality issues. Responsibilities for data are somewhat unclear, but this is not causing major problems.

We believe that a company of this size is a good place to start the research as the size and relative informal atmosphere will enhance data collection.

Software Improvement Group	
Business	Software advisory
Customers	Government, financial institutions, utilities
Turnover	€ 12 M
Employees	100
IT-intensive	Yes, using variety of IT tools; providing IT advise
Data governance	No, ungoverned data practices

Table 7 – Overview of Software Improvement Group (SIG)

2.10.2 Bank

Bank is a Dutch bank with a niche position in international cash management. An overview of Bank is illustrated in Table 8.

Of its two hundred employees, about a hundred are IT. This indicates that the organisation is IT-intensive. The internal software development organisation recently switched to agile methodologies.

Bank recently planned for a data lake, integrating its major data sources into one central place. There are three motivations to do so: (1) increasing information demand from clients, (2) the desire for more management information, and (3) increasing compliance requirements. Bank knows that its current data practices are not adequate for the data lake. To improve its data practices, the organisation drafted a target data architecture, including new tasks, roles and responsibilities.

Bank's primary motivation for participating in the study is to see if the target data architecture will improve data governance to an adequate level.

There is no explicit data governance program at Bank. Data ownership and responsibility are arranged via regular business functions and there are no specific data governance roles such as data stewards and a data council.

Bank	
Business	Cash management
Customers	Large multinationals
Turnover	€ 100 M
Employees	200
IT-intensive	Yes, about half of the employees works for IT
Data governance	No, ungoverned data practices

Table 8 – Overview of Bank

2.10.3 ANWB

ANWB is a large Dutch association involved with travel and tourism. An overview of ANWB is given in Table 9.

The organisation realised years ago that IT and data are necessary for the organisation's future, making the organisation IT-intensive. One example of that are ANWB's aspirations for substantially transforming the organisation using big data. Highest management listed data quality as one of its top priorities and to emphasis this, the CIO recently entered the executive board.

To bring the business and IT closer together, ANWB recently adopted the agile way of working. The organisation is organised strongly bottom-up and every employee has full responsibility for its work. With that comes an organisational aversion for large, top-down governance projects.

Since ANWB is an association, making its members happy has higher priority than making profit. This is visible in the organisation, where a lot of attention is being paid to for instance the correctness of member information.

ANWB is a diverse organisation spanning magazines, emergency services and insurance. With that diversity comes a large variety of IT systems. The size of the organisation allows for specialised departments such as *systems integration* and *customer intelligence*.

ANWB's main motivation to participate in the study is to learn the latest science on data governance practices.

Data governance at ANWB is similar to that of Bank: there is no formal program, data ownership is arranged via regular business functions and there are no such functions as data stewards and a data council. The need for governance is the highest in large organisations (Weber 2009) such as ANWB and we are curious if the absence of a formal data governance program is causing significant problems.

ANWB	
Business	Travel, tourism
Customers	Everyone who travels
Turnover	€ 995 M
Employees	5000
IT-intensive	Yes, IT and data are necessary for the organisation's future
Data governance	No, ungoverned data practices

Table 9 – Overview of ANWB

3 DATA GOVERNANCE

In this chapter we will discuss data governance and its relation with information governance, data management, data quality, enterprise architecture, IT governance and corporate governance.

3.1 Data as a valuable organisational asset

Until recently, the primary attitude of organisations towards data was that it is simply ‘present’, ‘has something to do with IT’, or ‘is primarily costing money’. Around the year 2012, the attitude globally tilted to ‘data can create business value’. Nowadays, a high percentage of organisations across the world are believing that data can be a valuable asset to them (Logan 2012; Economist Intelligence Unit 2008). We define ***data assets*** as data sources that are of special value to the organisation (ISO 2012; Weber 2009). Just like with other organisational assets, such as people and intellectual property, this will require a certain form of governance.

Five primary motivations to treat data as a valuable organisational asset, hence to adopt data governance practices, can be extracted from the literature (Otto 2011).

3.1.1 Ensure compliance

Ensuring compliance with regulation is the most frequently mentioned reason to start a data governance program. Examples of such regulations are the international Basel II standard and the European directive Solvency II. Basel II and Solvency II are intended to control how much capital banks and insurers need to hold guard to manage the financial and operational risks they face (Basel Committee on Banking Supervision 2006). Regulatory pressure increases, especially for financial institutions (Ernst & Young 2011). With this comes an increased need to improve data management practices. Data governance can greatly help organisations in improving these data practices, which is why organisations are implementing data governance programs. An example is illustrated in Example 1.

Example: data needed to prove compliance.

Bank recently needed to show a supervisory body whom of its commercial clients have over EU 100k combined on all their bank accounts. The clients should be displayed as the legal owner. Bank’s clients are large multinationals that typically have dozens of bank accounts across many different countries and currencies. Each bank account has one or more representatives, of which some are the legal representatives.

In order to gather the required information, Bank needs to (1) collect bank account information from various countries, (2) convert all these to the same currency, and (3) connect these to the legal representative of that client.

All data will come from various sources, such as different banking systems for different currencies, exchange rates, what bank account is from which client, and who the legal representative of a client is. Most data is present in Bank, but previously it was not relevant to connect these sources. It has taken significant time and effort to deliver the desired information to the supervisory body.

Data governance can help in for instance setting company-wide standards for metadata, making it easier to combine datasets.

Example 1 – Illustration of Bank needing data to prove that it is compliant

3.1.2 Enable better decision-making

A second reason to implement a data governance program is to enable better decision-making, commonly referred to as business intelligence. For business intelligence, several systems need to be connected. The quality of data is imperative for the confidence that a decision-maker has in the decision (Cheong & Chang 2007). A data governance program can support in that, as illustrated in Example 2.

> ***Example: data needed to enable better decision-making***
> ANWB wants to know how many of its members are between 20 and 30 years old. If management finds this number too low, they may start an advertising campaign targeted at that group. Good insight into the member database helps management in determining what campaigns are needed, and if these are successful.
>
> Given the diverse nature of ANWB, client information is stored in various systems. Each system is operated and maintained by different persons and different definitions for data quality may be in place. A member may be stored in multiple systems, but with different notations; when combining these systems, the client is counted multiple times which results in an inaccurate client count.
>
> Data governance can help to increase data quality, which enables better decision-making

Example 2 – Example of ANWB needing data to enable better decision-making

3.1.3 Improve client satisfaction

Another common reason for data governance programs is improving client satisfaction. Examples are by offering client new insights using information that previously was not available in that form, or by offering services to specific clients based on client preferences. Data governance can help to improve client satisfaction, as illustrated in Example 3.

> ***Example: data to improve client satisfaction***
> To reward loyal members (clients) and to stimulate members to visit stores, ANWB wants to send members personalised offers. These offers are based on the members' shopping history. To generate offers for all these members, an automated computer system needs to analyse large amounts of data, for instance recent purchases. The quality of that data needs to be sufficient for the system to do its work.
>
> Data governance can help by giving handles to draft a definition for data quality in these systems, which can be used to monitor data quality and to identify what functions should execute the monitoring.

Example 3 – Example of ANWB needing data to improve customer satisfaction

3.1.4 Increase operational efficiency

A fourth reason for data governance mentioned in literature is increasing operational efficiency. Examples are by better registering nonconformities in a factory resulting in more insight, or by forecasting required occupation. The latter example is illustrated in Example 4.

Example: data to increase operational efficiency

ANWB's road services includes cars that are driving around the country to help members when their car breaks down. ANWB can optimise their road services by predicting where the highest number of breakdowns will be, or how high the total number of breakdowns will be.

To predict that, ANWB needs historical data of the road services, which may need to be combined with traffic, road and weather conditions. All these datasets need to be of sufficient quality and definitions, such as road number, should match.

Data governance programs can help, for instance by assigning an appropriate person to manage the different format in road numbers.

Example 4 – Example of ANWB needing data to increase operational efficiency

3.1.5 Support business integration

A final goal may be to support the integration of business, for instance after a merger or acquisition. Several systems may be combined into one, which requires that data will be combined. As illustrated in Example 5, data governance can help in business integration.

Example: data to support business integration

Bank recently merged with another organisations. In the new, joint organisation, there is an overlap in information systems. Simply shutting down half of them would cause significant problems since not everybody is trained in the new systems and the old systems still contain a lot of valuable data.

A data governance program can help, for instance in mapping what data is present in which system, who is responsible for that data, how that will change when some of the systems are shutdown, and by merging two data quality definitions of similar systems.

Example 5 – Example of Bank needing data to support business integration

3.2 Governance over data assets

We just identified five primary motivations for organisations to use data as a valuable data asset. Just like any organisational asset, this requires some form of governance. For data, this is called ***data governance*** and is promoted by both researchers and practitioners.

3.2.1 Positive effects of data governance

Cheong and Chang (2007) found that managing the data quality of organisational data is not effective without a formal data governance program. The reason is a lack of clear roles and responsibilities among data stakeholders. In a large survey under corporate executives around the world, the Economist Intelligence Unit (2008) found that 85% of the companies that have implemented a data governance strategy positively rate their company's ability to protect sensitive data. Only 51% of those without a formal program in place positively rate that ability. This suggest a correlation between a company's commitment to governing data, and its capacity to mitigate risk and reduce cost, as well as getting more value out of its data assets.

For data governance, no studies are conducted on the relationship between governance maturity and firm performance. However, these have been done for IT governance. Luftman and Kempaiah (2007) found that organisations with a higher IT governance maturity perform significantly better. Since IT governance and data governance are closely related (see section 3.7), we expect that the same positive correlation holds for data governance.

3.2.2 Specification of data governance

In its report *Principles of Corporate Governance*, the Organisation for Economic Cooperation and Development defines governance as the structure through which the ***objectives*** of an organisation are set, and the ***means*** of attaining those objectives (emphasis added) (Demise 2006). Objectives are what the organisations wants to achieve on the long run. Means include the allocation of decision rights and voting powers, the acceptance of responsibility, coordination of stakeholders, communication and conflict resolution and the definition and enforcement of guidelines (Weber 2009). In short, for means the organisation's ***decision-making framework*** should be specified.

Weill & Ross (2004) explain that such a decision-making framework includes ***tasks*** in decision-making (for instance: "maintaining what data assets there are"), ***roles*** that fulfil certain tasks ("the information architect should maintain what data assets there are") and ***responsibilities*** of these roles ("the information architect is responsible, but the CIO is accountable").

These objectives, tasks, roles and responsibilities can play over several data governance domains (Khatri & Brown 2010). The first domain is ***data principles*** from which guidelines how data should be used in the organisation are derived. Based on these data principles, there should be guidelines for the other domains: ***data quality***, ***metadata***, ***data access*** and ***data lifecycle***.

In the next sub-section we will use the named aspects to define data governance. The aspects are explained in detail in chapter 5.

3.2.3 Definition

In line with the definition of governance and with popular definitions in IT governance (Weill & Ross 2004)[11] and data governance research (Khatri & Brown 2010; Weber et al. 2009), we define data governance as:

> Data governance is the set of processes within an organisation that serves two purposes: specifying an organisation's data objectives, and specifying a decision-making framework that is appropriate for meeting these objectives. The framework consists of tasks, roles and responsibilities that all play on the domains data assets, data quality, metadata, data access and data lifecycle.

Other aspects that are present in some governance definitions, such as risk management (IT Governance Institute 2008) or organisational training programs (DataFlux 2010) are relevant but are out of the scope for this research due to time constraints.

3.2.4 Organisational placement

Organisations should decide where to place responsibility for the data governance program. There are three independent but interconnected decisions to be made (Otto 2011): whether data governance should be the responsibility of the business or IT (*functional positioning*), whether executive or middle management should be responsible (*hierarchical positioning*), and whether it should be arranged centralised or decentralised (*organisational form*).

3.2.4.1 Functional positioning

Responsibility for data management is often implicitly placed at the IT department (Weber et al. 2009). At first glance, this may seem reasonable since data needs to be stored appropriately and several data sources

[11] IT governance and data governance show strong similarities; see sub-section 3.7.1.

need to be connected to collect relevant data. However, is it fair to make the IT department responsible for a wrong data entry? Is it clever to make IT responsible for interpreting data, while their core task is making sure all required software and hardware is up and running?

IT plays a large role in managing data, but the business should be responsible for the content and interpretation of data (Kooper et al. 2011). Therefore, a ***close collaboration of the business and IT*** is one of the core aspects of data governance (i.a. Weber et al. 2009; Kooper et al. 2011). IT should be responsible for making sure data is securely stored and is retrievable, whereas the business is responsible for the data's content. Or, to use the Cheong & Chang's analogy, "the infrastructure (the pipe) should be the responsibility of IT and the data (the information that flows through the pipe) should be the business' responsibility." Decision-making between business and IT may occur in a data board (see sub-section 5.2.1).

3.2.4.2 Hierarchical positioning

There is no clear trend about the hierarchical positioning for the end responsibility of a data governance program. Most scientists suggest that it should be located both on the executive management level as well as on the middle management level (Otto 2011).

Regardless of where it should be located, research has been done on where it often *starts*. In a 2013 survey among largely information professionals from a large range of industries and from both small and large organisations, it was found that about 47% of the programs started top-down, 28% bottom-up and 25% middle-out (Dataversity 2013). Lucas found that both top-down and bottom-up programs have proven successful (2010).

3.2.4.3 Organisational form

The third question regarding organisational placement of data governance involves whether it should be centralised or decentralised. In extremely centralised companies (*monarchy*), all decision-making powers are at top-level management. In extremely decentralised companies, the management delegates all decision-making powers to lower power levels (*feudalism*). Most organisations place authority for decision-making somewhere in between these two, for instance in a *federalist* or *duopolic* structure (Weill & Ross 2004; Khatri & Brown 2010; Weber 2009). In a federalist structure, only important strategic decisions are made by highest management and departments operate largely independently. In a duopoly, decisions represent a bilateral agreement between executives of two groups, for instance IT and Sales.

To the best of our knowledge, no studies are conducted on the relationship between the organisational form of data governance decision-making, and firm performance. However, they have been done for *IT* governance, which is related to data governance (section 3.7). These studies show that top-performing organisations tend to have federalist (Luftman & Kempaiah 2007) or duopolic (Weill & Ross 2004) decision-making structures.

3.3 Governance vs. management

There are important differences between the concepts governance and management. Peterson et al. (2000) illustrate that whereas IT governance deals with the future state of the IT in an organisation and the road on how to get there, IT management focuses on the management of IT operations (Figure 3). We conclude that governance means taking a step back from day-to-day operations and asking yourself what the organisation wants to achieve on the long run.

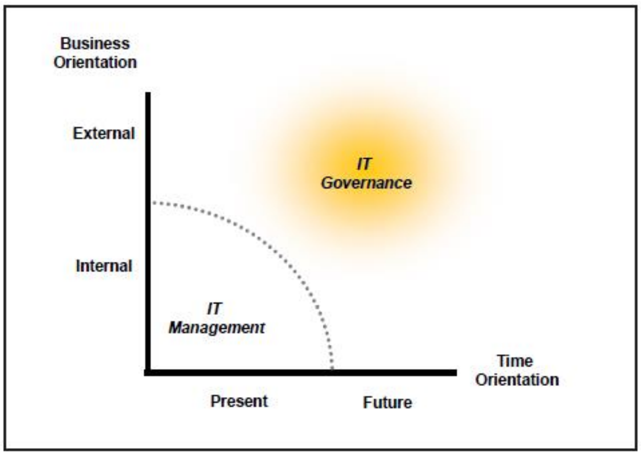


Figure 3 – The difference between IT governance and IT management (Peterson et al. 2000). The same distinction is made between data governance and data management.

Weill & Ross (2004) further elaborate on this distinction by stating that IT governance refers to what decisions must be made to ensure effective management and use of IT, whereas management involves the actual making and implementing of decisions.

In scientific data governance literature, the same reasoning is applied to separate data governance from data management (Khatri & Brown 2010; Weber 2009). Data governance is usually executed on the strategic level of an organisation; data management on the tactical or operational level (Lucas 2010).

3.4 Data vs. information

Business require information, rather than data. However, data can be processed into information.

Data is stored in for instance a database and technical in nature. Data is raw and can be defined as a "stored representations of objects and events that have meaning and importance in the user's environment" (Lucas 2010). When data is processed such that it increases knowledge for the person who is using the data, it can be considered information (Bellinger et al. 2004; Lucas 2010). Processing data into information should be done by humans since it requires interpretation, interaction and context (Kooper et al. 2011). Information is non-technical in nature, but needs to be derived from data.

We can state that organisations are treating *data* as an organisational asset since that is the way to unlock the *information* they desire. Many researchers do not distinguish data governance from information governance or consider it to be the same (Lucas 2010; Khatri & Brown 2010). However, Kooper et al. (2011) distinguish them by saying that data governance focuses on data assets, whereas information governance focuses on interactions. We consider both data assets as well as interactions part of data governance, hence do not distinguish between data governance and information governance. Since most of the effort is in data rather than information, we prefer the term data governance.

3.5 Data quality

Data quality is imperative in exploiting data to its fullest potential, as illustrated in Example 6. Both researchers and practitioners claim that achieving high quality data is a very important reason (Cheong & Chang 2007; Khatri & Brown 2010), or the only reason, (Hüner et al. 2009; Weber et al. 2009; Russom 2006) to adopt data governance practices. Several researches about data quality management touch upon data governance aspects, such as the data ownership and importance of assets (Lucas 2010; Hüner et al. 2009).

Example: high data quality to unlock information in data
To understand data quality, think of a house: if the roof leaks above the living room, one cannot sit in the living room. I.e.: since the quality of the house is low, one cannot unlock all of the house's potential. An example of low data quality is that in one system suffixes are stored as 'Van der', whereas in the other systems as 'V.d.'. Although this problem is not difficult to solve, it is an extra barrier in comparing clients in these two systems.

Example 6 – High data quality to unlock information in data

3.5.1 Defining data quality
One of the most widespread definitions of data quality states that data is of high quality "if it is fit for its intended use in operations, decision-making and planning" (Lucas 2010; Hüner et al. 2009; Wang & Strong 1996). In short, data should be fit for purpose. A lot of data is not fit for purpose: it is estimated that more than 25 percent of critical data in Fortune 1000 companies is inaccurate, incomplete or duplicated (Gartner 2007).

3.5.2 Consequences of poor data quality
When data is of low quality, the consequences can be enormous. Gartner research shows that when the anticipated value of business initiatives is never achieved, in 40% of the cases this is due to poor data quality in both the planning and execution phases (2007). Poor data quality is associated with the increase in cost and complexity of several systems, such as customer relationship management (CRM) and enterprise resource planning (ERP) (Cheong & Chang 2007). Process failure and information scrap and rework caused by defective information costs the United States alone $1.5 trillion or more (Gartner 2007). The opposite is also true: high data quality can result in significant benefits, such as a greater confidence in analytics systems, less time spent on reconciling, data and reduced costs (Cheong & Chang 2007; Russom 2006).

3.5.3 Data quality management
Data quality management (DQM) deals with the quality-oriented management of organisational data (Weber 2009). Although DQM is involved with data quality *management* rather than governance, many of the overarching principles of DQM are relevant to data governance. In turn, DQM is complemented by governance since governance clarifies the stakeholders in the data quality management (Weber 2009).

Two of the most popular DQM approaches are Total Quality Management data (TQdM) and Total Data Quality Management (TDQM), developed by English (2006) and Wang et al. (1998). Both promote the information product (IP) approach, aiming to treat data as if it were a tangible product.

3.5.4 Data stewardship
Since data is produced and used throughout the entire organisation and since employees are dependent on the information of others, employees should be trained to feel accountable for data. ***Data stewardship*** is the willingness to be accountable for a set of business information for the well-being of the larger organisation (Kooper et al. 2011; English 2006; Wang et al. 1998). Contrary to stewardship theory is agency theory, where employees are generally believed to be self-interested (Davis et al. 1997). The notion of agency theory is absent in data governance literature, which indicates that responsibility for data governance should lie throughout the entire organisation, rather than at individual agents.

Data stewardship is part of the corporate culture and as with any corporate culture, top managers play a crucial role in institutionalising it (Economist Intelligence Unit 2008; Dataversity 2013; Weber 2009). Without data stewardship it is virtually impossible to get high quality data (English 2006; Lucas 2010). Also, data quality problems become more of an IT problem (Lucas 2010).

3.5.5 Data stewards

No matter how much people will feel responsible for business information, having specific ***data stewards*** will always be needed. Data stewards are responsible for managing the quality of a set of data on a daily basis (English 2006).

3.6 Enterprise architecture

Understanding the relationship between information and data is critical. We derive this knowledge from the field of enterprise architecture.

Enterprise architecture (EA) designs the IT organisation in order to optimally support the business (Spewak & Hill 1993). EA states that the IT organisation is into five layers: business, information, applications, data, and infrastructure (GAO 2001; Figure 5). Data governance deals with all these layers.

At the business layer, data governance it is involved with understanding the business processes and how the organisation makes money. This business layer is connected to the information layer, which deals with the business' information requirements. These information requirements need to be fulfilled using data from the data layer, however, most data sources are tightly connected with specific applications, for which the application layer should also be included. Between the data and application layer tends to be a data model. Finally, data is stored on hard disks in laptops computers and servers – the infrastructure layer. This layer is important for including physical access, backup, retention and deletion of data.

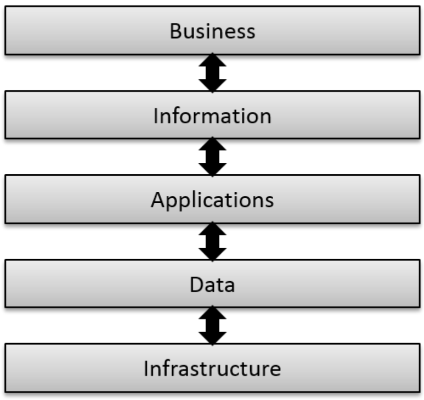


Figure 4 – The five enterprise architecture layers, illustrating how IT supports the business

Knowing an organisation's information requirements and their connection to business processes, data assets and applications that connects these and the infrastructure on which they run, is critical in understanding how data is used in organisations and who should be responsible for certain parts. Remarkably, these advantages are widely discussed in professional literature (Dember 2009; Logan 2012), but hardly in popular scientific data governance literature.

Another version of enterprise architecture consists of four models: business, data, applications, and infrastructure (Spewak & Hill 1993). Since it makes no distinction between data and information, we believe it is not suitable for data governance.

3.7 Relationship with IT and corporate governance

In the previous sections we occasionally extracted data governance knowledge from other types of governance, such as the definition and the difference between governance and management. In this section we will explore the differences between corporate governance, IT governance and data governance.

3.7.1 IT governance

Data governance is a relatively new discipline where much of what we know is extracted form IT governance. This makes sense, as both share many of the same properties. Both support the business, strongly distinguish between governance and management, deal (partly) with intangible assets and given their technological nature require distinct knowledge (Kooper et al. 2011).

However, IT and data governance are inherently different. Whereas IT governance deals with ***IT assets*** (that is: applications and infrastructure), data governance focuses on ***data assets*** in order to transfer it into information (Khatri & Brown 2010). Although IT stands for information technology, the information aspects seems not to be part of at anymore. Some researchers even go so far to contribute the difficulties in aligning business and IT to IT neglecting information (Kooper et al. 2011; Maes 1999).

In line with other researchers, we conclude that IT and data governance are interconnected but independent disciplines (Weber et al. 2009; Kooper et al. 2011; Khatri & Brown 2010; Cheong & Chang 2007). Figure 5 illustrates the different scope of IT and data governance using the five layers of enterprise architecture.

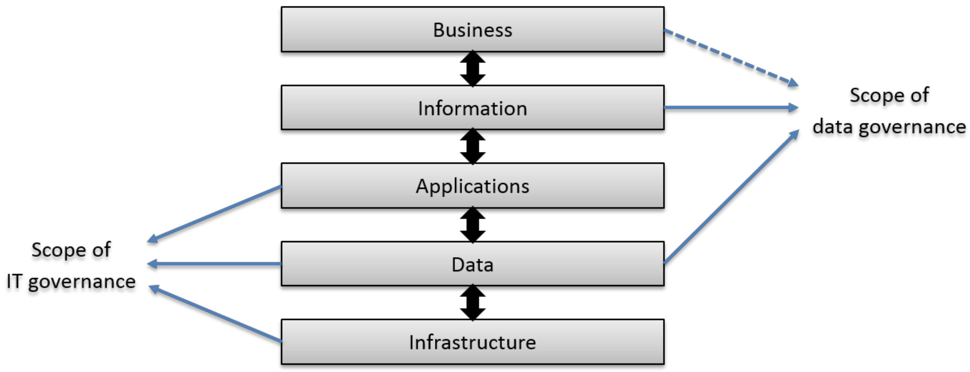


Figure 5 – The scope of IT and data governance, illustrated using the five enterprise architecture layers

3.7.2 Corporate governance

Where the scope of IT and data governance is limited to specific types of assets, corporate governance deals with all assets in an organisation (Figure 6). Both IT and data governance are generally considered subsets of corporate governance (Kooper et al. 2011). However, there is a reason that specific governance fields are needed for IT and data. As mentioned earlier, these deal with technical asset and require distinct knowledge. Therefore, we believe IT and data governance are rather independently operating subsets of corporate governance.

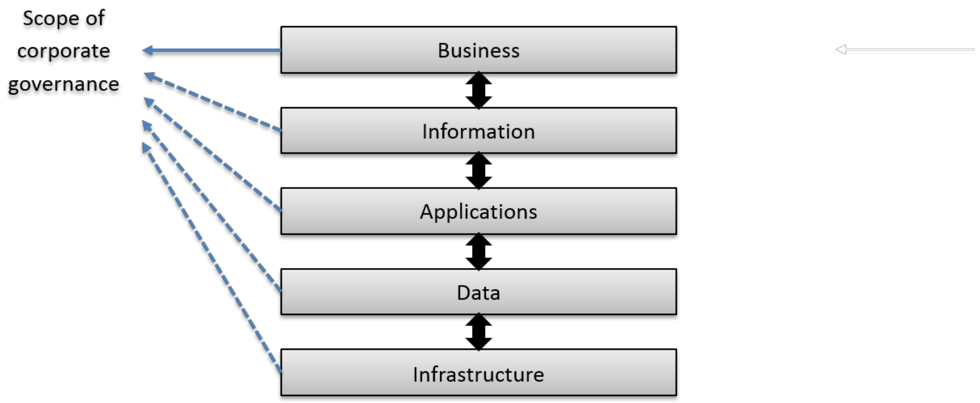


Figure 6 – The scope of corporate governance, illustrated using the five enterprise architecture layers

4 ASSESSING DATA GOVERNANCE MATURITY

In this chapter we will discuss how a maturity assessment can help organisations, what requirements of such an assessment are and if available models for data governance meet these requirements.

4.1 Maturity assessments

In sub-section 1.2.2, we illustrated that many organisations want to implement a data governance program but do not know where to begin. Organisations that already implemented a data governance program may not know if and where improvements are needed.

Maturity assessments are common to determine how mature, competent, capable or sophisticated the processes of a specific area are. Not only can they help in understanding the current situation, they frequently also mark areas for improvement (Hüner et al. 2009; De Bruin et al. 2006)(De Bruin et al. 2006).

It makes sense that a data governance maturity assessment should be capable of assessing the maturity of a formal data governance program. However, also organisations that do *not* have such a formal program in place should be able to identify areas for improvement. Practically every organisation manages data and these organisations will have structures or guidelines in place to do so. We call these ***ungoverned data practices***, of which three examples are illustrated in Example 7. Even organisations without a formal data governance program can assess these ungoverned data practices and will not start from scratch if they implement a formal program.

Example: guidelines
A financial employee daily enters invoices in the accounting software. Likely, there are guidelines how the employee should do his or her work. Although the word 'data' does not occur in the job description, he or she is working with it.

Example: maintaining data sources
Responsibility for maintaining a specific data source is neither explicitly nor implicitly assigned, but it is likely that some employees will find the asset important to the organisations and will take care of it.

Example: customer databases
Practically every organisation will know that its customer database is important, but likely not all will have identified it being a 'data asset'.

Example 7 – Three examples of ungoverned data practices. Organisations that do not have a formal program can assess the maturity of these ungoverned data practices.

4.2 Requirements of a data governance maturity assessment

Having identified the advantages of data governance maturity assessments for organisations that do and do not have a formal data governance program, we will now state what requirements a method to assess data governance maturity should have.

The method should be capable of assessing the maturity of formal data governance programs (requirement 1) and of ungoverned data practices (2). To do so, we naturally need some way of producing a maturity score (requirement 3).

Data governance in its current form can be seen as a collection of concepts without a consensus on what the most important concepts are and how they relate to each other (see sub-section 1.2.1 and section 4.3). Given this lack of structure, we consider all these concepts being separate dimensions, making data governance multidimensional. The method should capture that multidimensionality (requirement 4).

To comprehend this multidimensionality, we should be able to decompose it into meaningful units. Research about the human mind showed that it can comprehend a limited number of items at the same time, hence we want the maximum amount of meaningful units per dimension to be that number. Since results on what exactly that number is vary between five and nine (Miller 1956)[12], we can conclude that a maximum number of five in general will be comprehensible (requirement 5).

A maturity assessment is usually not executed to map the current situation, but to mark areas for improvement (De Bruin et al. 2006). In order to know if improvements are needed, the method should provide a means to interpret the results for a specific organisation (requirement 6). If improvements are needed, it should be possible to derive actionable recommendations (requirement 7).

Also, assessors should accept using the model (requirement 8). To ensure trust in the model, it should provide reliable results between different raters (requirement 9). Finally, to ensure the quality, it should be scientifically evaluated (requirement 10). In summary, the method has the following requirements:

1. Assess maturity of formal data governance programs;
2. Assess maturity of ungoverned data practices;
3. Produce a maturity score;
4. Capture data governance's multidimensionality;
5. Each dimension should be decomposed into a maximum of five units;
6. Provide a means to interpret the results;
7. Provide a means to derive actionable recommendations;
8. Accepted by the assessor;
9. Provide reliable results;
10. Scientifically evaluated.

We tested eleven models we encountered during our literature study, both scientific and professional, on these ten requirements. It is discussed in the following section and summarised in Table 10.

[12] This is known as 'Miller's law'

Origin	*Short description*	*Assess data governance maturity*	*Maturity score*	*Multidimensional*	*Maximum of 5 meaningful units*	*Organisation-specific interpretation*	*Actionable recommendations*	*Accepted by end users*	*Reliable results*	*Scientifically evaluated*
(Weber 2009)	Data governance reference model	No	No	Yes	Yes	Yes	Yes	N/A	N/A	Yes
(Cheong & Chang 2007)	Data governance roles for organisational layers	No	No	Yes	Yes	No	Partly	N/A	N/A	Yes
(Khatri & Brown 2010)	Data governance decision areas	No	No	Yes	Yes	No	Partly	N/A	N/A	Yes
(Tallon et al. 2013)	Data governance antecedents	No	No	Yes	Yes	No	Partly	N/A	N/A	Yes
(Data Governance Institute n.d.)	Data governance components	No	No	Yes	No	No	No	N/A	N/A	No
(Informatica 2006)	Data governance components	No	No	Yes	Yes	No	No	N/A	N/A	No
(Global Data Excellence 2011)	Components and best practices for data governance	No	No	Yes	Yes	No	No	N/A	N/A	No
(Chen 2010)	Generic data governance maturity model (CMMI-based)	Yes	Yes	Partly	Yes	Partly	Partly	N/A	Yes	No
(Salido et al. 2010)	Generic data governance maturity model (CMMI-based)	Yes	Yes	Partly	Yes	Partly	Partly	N/A	Yes	No
(DataFlux 2010)	Generic data governance maturity model (CMMI-based)	Yes	Yes	Partly	Yes	Partly	Partly	N/A	Yes	No
(Hüner et al. 2009)	Self-assessment for data quality	Partly	Yes	Yes	Yes	Partly	Partly	N/A	Partly	Yes

Table 10 – Overview of the studied data governance models and how they meet the stated requirements

4.3 Models

We tested eleven models on how they meet the stated requirements.

4.3.1 Data governance reference model & contingency approaches

Weber (2009) developed an extensive reference model for data governance, which provides a starting point for the design of organisational, corporate-wide data quality management to globally operating companies. It identifies design options, solution proposals and recommendations for company-specific organisational structures. The model defines roles and responsibilities, their assignment into the existing organisational structure, their cooperation and coordination.

Weber extensively studied data governance and her model incorporates the multidimensionality of data governance. By showing contingencies (satellite paper; Weber et al. 2009) in data governance design, the model can help to interpret data governance. Although the model will help organisations in designing their data governance, Weber provides no means to assess maturity.

4.3.2 Data governance structures

Cheong and Chang (2007) identify a specific data governance structure, based on a case study. The structure consists of several organisational bodies and their relation with each other.

On an organisation's strategic level, there should be a data governance council that is responsible for endorsing policies, aligning business and data initiatives, and reviewing budget submissions for data related projects. On the tactical level, data custodians and data stewards play a large role. On the lowest level, user group are involved. User groups consist of key data stakeholders from various divisions.

Cheong & Chang's model helps in understanding what data governance roles should operate on what organisational layer, but it does not provide a way to establish data governance maturity.

It is noteworthy that where Weber promotes a contingency approach to data governance, Cheong and Chang seem to promote that all organisations should adopt the same structures.

4.3.3 Data governance decision areas

As has been discussed in section sub-section 3.2.3, Khatri & Brown (2010) identify five data governance decision areas: data principles, data quality, metadata, data access, and data lifecycle (Table 11). These have been translated from the decision areas that Weill & Ross (2004) designed for IT governance, however, it is not explained how this translation is performed. All decision areas are interrelated but deal with a distinctive set of core issues.

<table>
<tr><td colspan="3">Data principles</td></tr>
<tr><td rowspan="2">Data quality</td><td>Metadata</td><td rowspan="2">Data lifecycle</td></tr>
<tr><td>Data access</td></tr>
</table>

Table 11 – The data governance decision areas, as proposed by Khatri & Brown

The model is multidimensional and structures data governance in an intuitive way, but it is generic and there is no way to measure data governance maturity.

4.3.4 Data governance antecedents

Tallon et al. (2013) studied data governance literature and conducted interviews with data professionals. Their final research model gives an overview of positive and negative antecedents for data governance, the composition of data governance, and positive and negative consequences of data governance on firm performance (Figure 7).

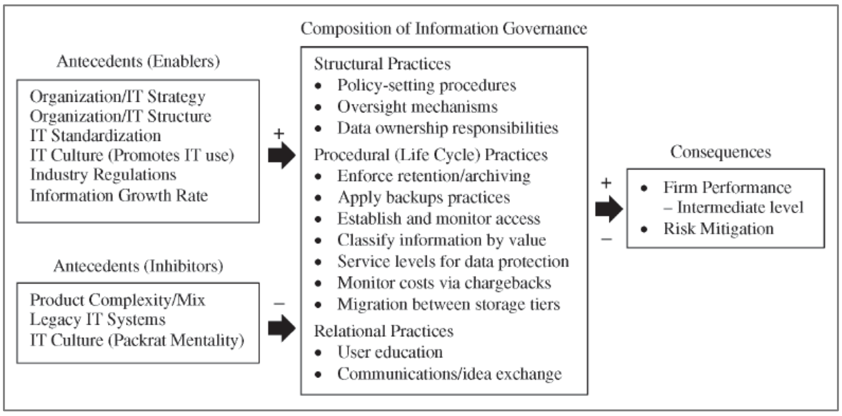


Figure 7 – The research model by Tallon et al.

Whereas their research models gives welcome insight into antecedents for, and composition and consequences of, data governance, again no way is provided to score data governance maturity.

4.3.5 Industry models of data governance composition

There are several data governance frameworks available that are created by commercial industry players.

For example the Data Governance Institute, an independent institute consisting of commercial industry players, names eleven data governance components, grouped in three categories: rules of engagement, peoples and organisational bodies, and processes. It also provides an image illustrating the relationship between these elevens components, but it is unnecessary complex and provides little insight (Data Governance Institute n.d.; Figure 9).

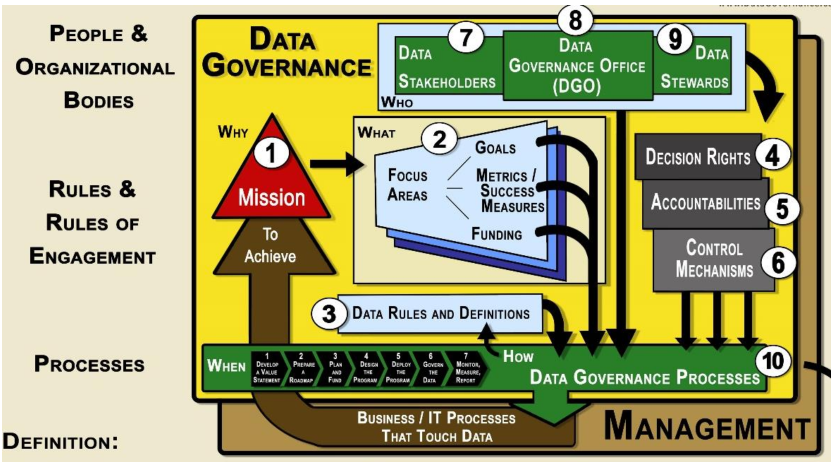


Figure 8 - We find the DGI data governance model is too complex to provide clear insight into data governance

Informatica, a service provider for big data and the cloud, also names a number of components (Standards, Policies and processes, Organisation, Technology) but gives little handles in how to use them (Informatica 2006). Global Data Excellence, a consultancy organisation focused on utilising organisational data, created the Data Excellence Framework. It covers many of the aspects that are included in scientific models, such as data governance roles and the importance of cultivated data assets in the corporate culture (Global Data Excellence 2011).

These models give valuable practical insight into how data governance is used by professionals, but usually do not go beyond a checklist of items or components that should be kept in mind. Therefore they are not suitable for assessing data governance maturity.

4.3.6 Industry data governance maturity models

There are a number of models available that try to assess data governance maturity using the Capability Maturity Model Integration (CMMI; CMMI Architecture Team 2007), all of which are created by industry professionals and are not scientifically evaluated.

The CMMI is a popular generic maturity model that is developed by Carnegie Mellon University. It focuses on optimising operational processes and classifies any process maturity on five well-defined evolutionary plateaus (De Bruin et al. 2006): *none, managed, defined, quantitatively managed, and optimised*. Characteristics are that almost all organisations make a similar progression through the levels and that it is virtually impossible to skip levels (Smith 2015).

Makers of data governance maturity models altered the five levels, such as Kalido (Chen 2010; *Application, Enterprise-repository, Policy, Fully governed*), Microsoft (Salido et al. 2010; *Basic, Standardised, Rationalised, Dynamic*) and DataFlux (DataFlux 2010; *Undisciplined, Reactive, Proactive, Governed*). Within each level, there usually are a number of key areas, such as organisation, process and technology (Kalido) and people, policies and technology (Microsoft; DataFlux).

These three industry data governance maturity models are quite similar. Like any CMMI-based maturity model, they allow organisations to identify their current data governance maturity, determine both interim and long-term goals for improvement and provide the best practices that will move them to the next stage (Smith 2015).

As was illustrated in section 4.2, data governance can be considered multidimensional. While it is possible in CMMIs to work with different key areas, progression to the next level is only possible if *all* key areas meet certain criteria, making CMMIs essentially one-dimensional. By reducing data governance to a one-dimensional concept one does not sufficiently capture its complexity. Researchers of other multidimensional concepts, such as data quality management, support this notion (Hüner et al. 2009).

4.3.7 Maturity for corporate data quality management

Hüner et al. (2009) constructed a framework for Corporate Data Quality Management (CDQM). Although the framework focusses on the *management* of data, it covers some of data governance's elements. The authors defined six enables for corporate data quality management, including a data strategy, a definition of roles and responsibilities, and a data asset overview.

These enablers are included in the CDQM self-assessment. For each enabler, organisations can score themselves on a scale from "nothing has been done" to "proven world-class approach". This self-assessment gives organisation an indication on how they are performing. However, since little guidelines are given on how to score, the results are unreliable and subjective. As the scores are not aggregated, it is difficult to easily see where there is room for improvement and to determine actions to improve. Moreover, the model is not made for data governance specifically and excludes some important aspects such as data access and data lifecycle.

4.4 Comparison of the models

All aforementioned models explain what elements data governance should have and how they relate to each other, but most models do not offer a means to establish data governance maturity.

There are a number of industry data governance maturity models. All of these are based on the CMMI and give an indication of the state of data governance in an organisation. They also provide actionable recommendation. Instructions on how to rate tend to be present, making the models fairly reliable. However, given that they are based on the CMMI, they are only partly multidimensional. None of the industry models appears to be scientifically evaluated.

To the best of our knowledge, there is only one scientific model that has some way of establishing data governance maturity. In the model by Hüner et al. (2009), maturity of data quality management can be assessed. However, the model does not cover all of data governance's aspects and is not reliable, objective and actionable.

Since none of the models we encountered meets all of our requirements, we developed a new data governance maturity method. We will discuss it in the next chapter.

5 DATA GOVERNANCE MATURITY METHOD

As illustrated in chapter 2, the data governance maturity method is designed and evaluated in two cycles. Differences between the initial and the final method are discussed in Appendix C.

Chapter 5 discusses the final version of the method. The method operationalises elements that are present in both scientific as well as professional literature. The method consists of a data governance maturity model, and instructions for how to use it (Figure 9). In section 5.1 to 5.4, we will discuss the model and in section 5.5 we will provide the instructions.

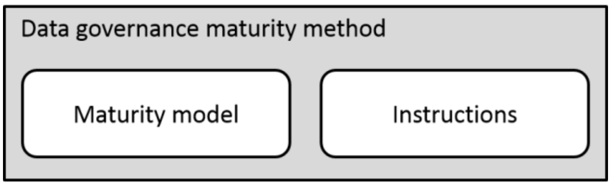


Figure 9 – The data governance maturity method consists of a maturity model, and instructions for how to use the method

5.1 Maturity model

In line with other maturity research (Hüner et al. 2009; De Bruin et al. 2006), the level of maturity describes how mature, competent, capable or sophisticated the processes of a specific area are. A common way to make maturity models is to define evolutionary plateaus, also known as maturity levels (De Bruin et al. 2006). Maturity model using evolutionary plateaus are often based on the Capability Maturity Model Integration (CMMI; CMMI Architecture Team 2007). In this category of models, a number of key areas is defined and if all of these key areas reach a certain point, maturity progresses to the next level. Since all key areas need to reach a certain point in order to progress, these models are essentially one-dimensional (Figure 10). They provide a simple means of comparing maturity levels, but do not adequately represent maturity within complex domains (Hüner et al. 2009; De Bruin et al. 2006), such as data governance.

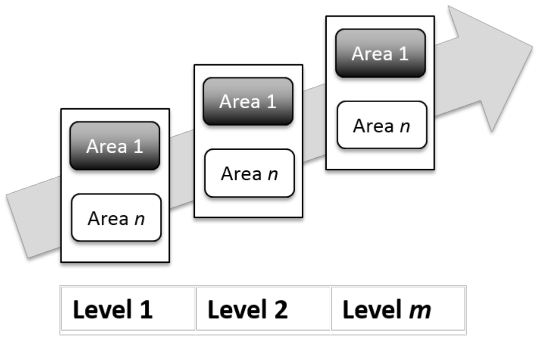


Figure 10 – A maturity model using evolutionary plateaus, for instance the CMMI. Progression to a next level is only possible if all key areas meet certain criteria.

In such complex domains, it is suggested to have areas with separate maturity scores (Figure 11). Jointly, these areas may form a maturity score for the overall entity. Doing so helps organisations in gaining a deeper understanding of their relative strengths and weaknesses in the domain, and to target specific improvement strategies (De Bruin et al. 2006).

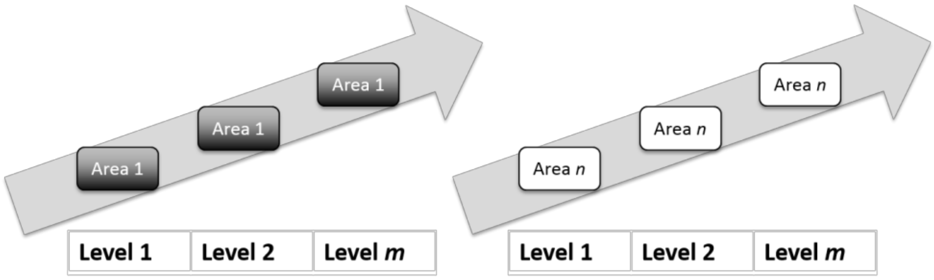


Figure 11 – A maturity model using separate maturity levels for each key area. Such an approach is advised for complex domains, such as data governance.

We developed our maturity model based on the idea of separate maturity scores for several areas. Our maturity model describes the elements of which data governance maturity consists, how they relate to each other and how the scores are computed.

5.2 Structure of the model

As listed in section 4.2, our maturity method should capture the complexity of data governance, but should be decomposed into a maximum of five meaningful units. In order to express a maturity score, we need some kind of a metric.

In software quality, a method to do decompose a domain and consequently attach a metric is Factor-Criteria-Metric (FCM; Figure 13; Marinescu 2005). FCM stimulates researchers to decompose a rather abstract attribute into comprehensible and concrete factors and criteria that can be operationalised into metrics (Aldris et al. 2013).

Given the close connection between data and software (section 3.6 and 3.7), we believe FCM is a good approach to derive data governance maturity metrics.

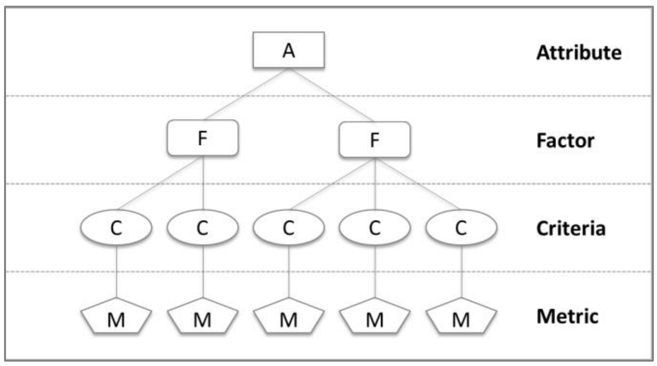


Figure 12 – Illustration of the Factor-Criteria-Metric approach.

The domain for which metrics are to be found is called an attribute. The attribute we want to express is data governance maturity.

- A: Data governance maturity

In section 3.2, we defined that any governance consists of several elements and that each of these elements plays on several domains that are specific to data governance. Data governance maturity should reflect all these, hence we decompose data governance maturity into the factors governance elements and data governance domains.

- F1: Governance elements

- F2: Data governance domains

The factors are explored in the following sections. The final data governance maturity model is illustrated in Figure 13.

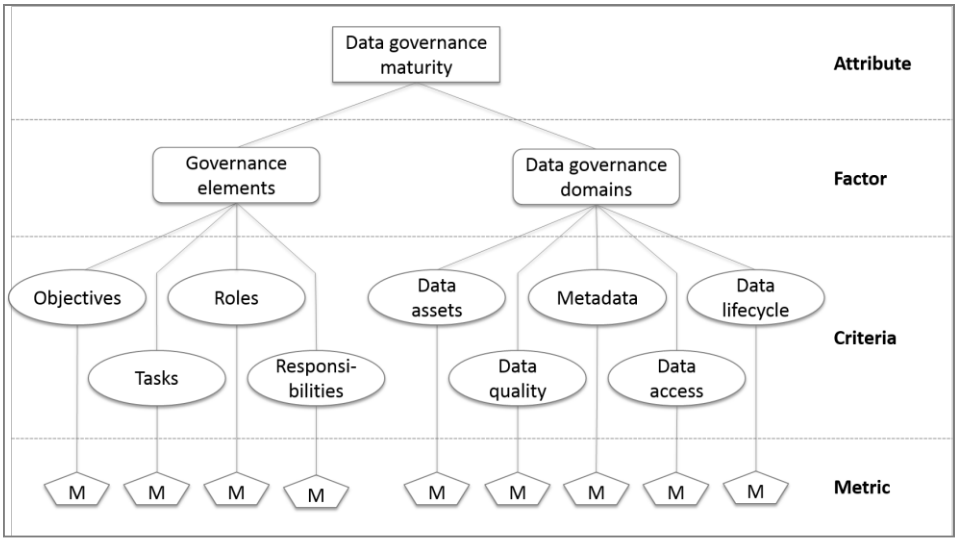


Figure 13 – Decomposition of the data governance maturity model, using FCM

5.2.1 Factor 1: Governance elements

In section 3.2, we defined that any form of governance consists of an organisation's objectives and an appropriate decision-making framework in order to meet these objectives. The decision-making framework consists of the tasks, roles and responsibilities. Based on these, we further decompose governance elements into the following four criteria:

- C1.1: Objectives
- C1.2: Tasks
- C1.3: Roles
- C1.4: Responsibilities

Each will be discussed in the following sub-sub-sections.

5.2.1.1 Criterion 1.1: Objectives

It is widely accepted in business sciences that organisations perform better if clear goals are set (i.a. Kaplan & Norton 2004). For data governance, objectives illustrate what information requirements an organisation has (Weber 2009; Hüner et al. 2009). These *information* objectives form the basis for how *data* should be used.

Organisations are advised to formalise their information objectives in an information strategy document, which should define the scope and goals of the data governance program. The strategy document is a kind of declaration of intent by top management, to take care of the value of data within the organisation. Besides that, the organisation's attitude towards data can be summarised in a brief mission statement, which can be used a simple communication tool (Weber 2009).

5.2.1.2 Criterion 1.2-1.4: Decision-making framework

A means to meet a body's information requirements is to define the data decision-making framework. One can form such a decision-making framework by defining tasks in decision-making, roles that fulfil certain tasks, and responsibilities of these roles (Weill & Ross 2004; Khatri & Brown 2010). These three aspects can be arranged in a matrix (Weber et al. 2009), sometimes called a Responsibility Assignment Matrix (RAM; Table 12). Tasks, roles and responsibilities are discussed in the following sections.

	Role 1	**…**	***Role* n**
Task 1	Responsibility [A|R|C|I]	…	Responsibility [A|R|C|I]
…	…	…	…
***Task* n**	Responsibility [A|R|C|I]	…	Responsibility [A|R|C|I]

Table 12 – The arrangement of tasks, roles and responsibilities.

5.2.1.3 Criterion 1.2: Tasks

To make data useful for the business, certain tasks have to be performed. Just some examples are maintaining the organisation's data asset overviews, keeping overview of all data governance activities, managing access to data assets, making sure that sensors are collecting data, and periodically evaluating guidelines.

The criterion 'tasks' indicates whether a task is executed, regardless if it is assigned to a specific role (next section). The willingness for a person to execute a tasks notwithstanding he or she has the responsibility for it, is called *data stewardship*.

5.2.1.4 Criterion 1.3: Roles

Every task that is defined is to be performed by one or multiple roles. Four major data governance roles emerge in literature (Otto 2011): sponsor, council, owner, and steward.

- ***Sponsor***

  For a data governance program to be effective, data governance requires a sponsor high in the organisation that grants the mandate for the program and continuously stretches the program's importance. Having a sponsor for the data governance program is believed to be one of the most important determinants for the program's success (Economist Intelligence Unit 2008; Dataversity 2013; Weber 2009).

- ***Owner***

  Persons that are owner have the highest level of responsibility over specific data assets. Often, the owner of a business process is implicitly also the owner of a related data asset. An owner typically is from the business. In order to explicitly set data ownership, a data asset overview is needed (see sub-sub-section 5.2.2.1).

- ***Stewards***

  Where data owners are accountable for data assets, they usually delegate responsibility for day-to-day operations to stewards. There may be a lead data steward and several business and technical data stewards. Alternative names for the data steward are trustee, custodian, pilot and information product manager (English 2006; Otto 2011).

  The role data steward should not be mistaken with data steward*ship*, which is part of the corporate culture.

- ***Council***

  A data governance council, or data board or data committee, is usually the way to structure the shared responsibility for data by the business and IT. A council consists of a representation of both the business and IT and its members meet regularly (Weber 2009). At least the data owners and data stewards should be represented in the council (Cheong & Chang 2007). The joint council is responsible for binding decision-making regarding data. Given its strong connecting with IT, the data council should periodically interact with and IT council (Cheong & Chang 2007). Although a data governance council is promoted by almost all scientists and practitioners, Otto (2011) found that it is not necessary for a successful data governance program.

To make sure that people are aware of their roles and to formalise roles, it is advised to define them in function profiles. An example for the role of data steward is given by (English 2006):

> "A steward for management, control and use of information. Maintain quality of information created or maintained within the process or department to meet our information consumers' needs, both within and outside our business areas."

In summary, the criterion 'roles' indicates what role is assigned for a specific task, regardless if the task is executed, and if it is assigned in a function profile.

5.2.1.5 Criterion 1.4: Responsibilities

The specification of responsibilities is commonly expressed using ARCI acronym (Gladden 2007) containing four levels of responsibility in descending order: accountable, responsible, consulted, or informed.

- The role that is ***accountable*** is 'owner' of a task or data asset and is ultimately end responsible. This person should be in the highest management of an organisation. Each task can only have one person accountable (Weber et al. 2009).
- He or she almost always delegates the responsibility for execution of day-to-day tasks to some 'steward' lower in the organisation: the person who is ***responsible***. Multiple persons can be responsible for executing the same tasks (Weber et al. 2009) and responsibility can also be assigned to a group of people.
- The person that is responsible may not have all the knowledge required to execute the task. Therefore, persons that can be ***consulted*** may be defined. These persons do not bear formal responsibility.
- Finally, when a certain task has been executed, it is common to ***inform*** others about it. It may come in handy to define this, but this is not compulsory.

Variations to assigning responsibilities using the ARCI acronym are RACI, which has the same four levels but orders them based on day-to-day involvement rather than level of responsibility, and RASCI, which adds the level 'supportive' in addition to 'consulted'. We choose the order used in ARCI since we believe that a descending order of responsibility it is the most intuitive for defining decision-making.

In large organisations, accountability is delegated over several levels (Figure 14). For example, a CFO may delegate responsibility for day-to-day operations of financial data to the head of finance, who can decide to delegate again it to the financial controller. From the perspective of the controller, the head of finance is accountable, but in reality it is the CFO. Although this may not be the 'official' situation, in reality a staircase of responsibility delegation may form. Naturally, all persons that have some level of responsibility over a data asset should be aware of that and agree with it.

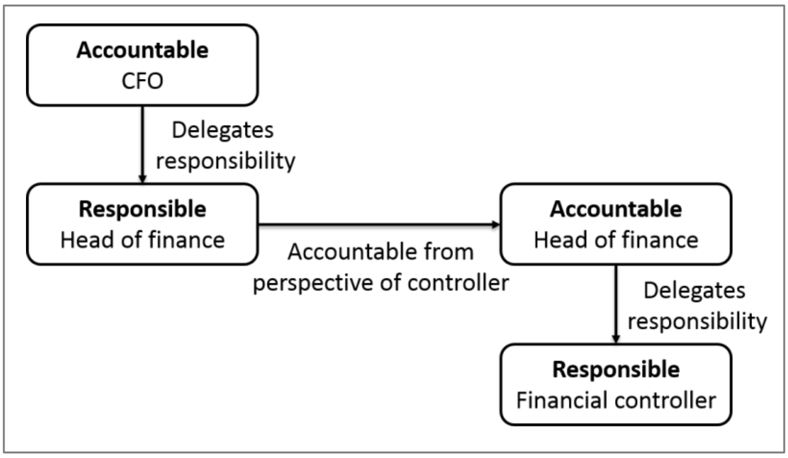


Figure 14 – Staircase of responsibility delegation

In summary, the criterion 'responsibilities' deals with the correctness in defining the responsibility for roles in executing tasks.

5.2.2 Factor 2: Data governance domains

Data is created and used throughout the entire organisation, the governance of data therefore spans several independent but interconnected domains. Khatri & Brown (2010) defined five data governance domains. We slightly altered these domains and define criteria for data governance domains as:

- C2.1: Data assets
- C2.2: Data quality
- C2.3: Metadata
- C2.4: Data access
- C2.5: Data lifecycle

Each will be discussed in the following sub-sub-sections.

5.2.2.1 Criterion 2.1: Data assets

Khatri and Brown named their first dimension 'data principles', setting the boundary requirements for the intended uses of data within the organisation. In essence, data governance is about managing data as an organisational asset (see section 3.1). Data assets should therefore form the core of data governance, which is why renamed the domain. We define data assets as data sources that are of special value to the organisation, because they contribute to meeting an organisation's information requirements (for information requirements, refer to sub-sub-section 5.2.1.1, criterion 1.1. Reference data assets: ISO 2012; Weber 2009).

If external data sources are important to an organisation, these are also data assets.

Prioritisation

Not all data assets are equally important. Just like with regular requirements and assets, these should be prioritised according to their importance to the business (ISO 2012). This importance dictates how much time and money should be spent on governing and managing the data assets.

In information security, the importance of specific data assets is commonly prioritised according to the importance of their confidentiality, integrity and availability. These three aspects are jointly known as the 'CIA triad' (Whitman & Mattord 2011). We use these information security practices to prioritise the key part of data governance: data assets.

Confidentiality indicates who is allowed to have access to certain files; this will usually play at the data access domain. ***Integrity*** discusses how high data's quality should be and what problems will arise if quality criteria are not met. In order to know desired levels of data quality, one should know how it should be interpreted. Therefore, integrity plays at the data quality and metadata domains. ***Availability*** deals with when data should be available and what problems arise when data is not available. Since availability is about infrastructure and backup, this plays at the domain of data lifecycle. It has a close connection to IT.

Data asset overview

Data governance starts with knowing what data assets there are, how important they are and who is accountable and responsible (ISO 2012). Such a data asset overview need not be technical and may be as simple as a table (Table 13).

Data assets	*Accountable*	*Responsible*	*Confide-ntiality*	*Integrity*	*Availa-bility*
Customer information	CCO	Sales	Low	High	Medium
Production hall sensor data	COO	Operations	Medium	High	High
Pipe line designs	CTO	Engineering	High	High	Medium
Etc.					

Table 13 – Example data asset overview

An overview of prioritised data sources, including related responsibilities, sets the standards for data throughout the entire organisation and for the other data governance domains.

5.2.2.2 Criterion 2.2: Data quality

Once we know the intended uses of data within the organisation, we can explore what appropriate levels of data quality are for each data asset.

As we have seen in section 3.5, data of high quality is critical in the ability to turn data into useful information. At least for the most important data sources, there should be clear understanding of the level of data quality that is needed. To do so, a definition of what data quality means in these assets is advised. Such a definition will be different for each data asset. It opens the doors for being able to measure, monitor, and evaluate data quality. A data quality definiton also allows for communicating these results to relevant stakeholders in both business and IT.

Many software tools are available to support organisations in their data quality work. Tools can increase the efficiency of the activities to assess, measure, and adjust for improvement of data quality. Tools may be used to automate parts of the process of evaluation, measurement and verification of data quality, to find business rules, correct data, or prevent errors (Weber 2009).

In section 3.5, we defined high quality as data that is fit for purpose. Data is used in countless different contexts and fitness for purpose differs in all these contexts. Data quality can be expressed using a variety of different attributes, such as accuracy, completeness and relevance (ISO 2008; Cheong & Chang 2007). This makes data quality inherently multidimensional.

5.2.2.3 Criterion 2.3: Metadata

In order to know desired levels of data quality, one should understand how it should be interpreted. In section 3.4 we stated that turning data into information requires interpretation. We now define that metadata provides descriptions about how data should be interpreted to use it as information.

Metadata deals with data about data. Examples are when and by whom it was created, what other piece of data it may be based on, and what it means. A more extended example is illustrated in Example 8.

Just like data quality, metadata means something different in every different context (Duval et al. 2002). One may consider an appropriate level of data quality, whom should have access to it, and when it should be deleted, also part of metadata. Strictly speaking, this indeed is part of metadata, but for comprehensibility we scope metadata to ‘interpretation’. Definitions of data quality, data access and data lifecycle belong to the respective domains.

Metadata can be a simple description in a document, or technical in nature. Metadata helps in fostering a common understanding about the importance of data objects and facilitates communication about technical or business unit boundaries (Weber 2009). As we can see in Example 8, clearly and consistently defined metadata helps in reducing redundancies and increasing accuracy, integrity and consistency (Rahm 2000; Weber 2009).

With the rise of big data and data lakes, data sources that previously were rather independent and used within one department, will now be connected and used organisation-wide. Such techniques make aligned definitions, and with that metadata, extra important.

Metadata

Metadata can be illustrated by data that is stored in two systems. Most client data may be stored in customer relationship management (CRM) system, but it is not uncommon that other client data is stored in other systems, for instance a system that supports the helpdesk. A common action is that a client calls the helpdesk to notify the organisation on a new address. The helpdesk operator makes the change in the helpdesk systems, but now, client information in the CRM is not accurate anymore.

To prevent integrity issues between data in these two systems, one may decide to make one system ‘master’ and the other ‘slave’. This means that client data in the one system is submissive to client data in the other. This can mean that client data can only be altered in the first system, or that it can be altered in two systems but that in case of a conflict the first one overrules the second. Another option is to add a third system that contains master data for the two other systems, which is called master data management (MDM; Figure 15). In MDM, core entities of the enterprise are defined, including customers, prospects, services and turnover (Loshin 2007). These definitions can be used by other systems.

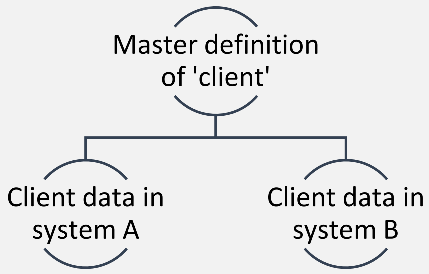


Figure 15 – The definition of client (metadata about ‘client’) using master data management

Example 8 – An example of metadata

5.2.2.4 Criterion 2.4: Data access

Data access refers to which users should have access to certain data. It should be based on the definition of unacceptable uses of data within the organisation, and compliance requirements for auditability, privacy and availability (Khatri & Brown 2010). Organisations can use international standards, such as ISO 27000 for information security, to derive their data access guidelines from (ISO 2012). Data is typically stored on media that are also physically accessible. Therefore, data access guidelines should also include physical data access (Khatri & Brown 2010).

Whereas many organisations struggle with rather vague and multidimensional concepts such as data quality and metadata, most organisations have specific security guidelines that include information security (PricewaterhouseCoopers 2012).

5.2.2.5 Criterion 2.5: Data lifecycle

Managing data as a product with a lifecycle is one of the key principles of data quality management, which in turn inspired data governance (section 3.5.3). Just like a physical product, data typically goes through a number of stages. It is created, used, needs maintenance, may be lost during an infrastructure crash, and will eventually need to be deleted or archived. Guidelines for data quality, metadata, data access and data lifecycle should be specified to all these stages.

Guidelines for these stages play a key role in operationalizing the data principles into IT infrastructure, making data lifecycle the domain that has the closest connection to IT (Khatri & Brown 2010). We derive typical stages for data lifecycle from COBIT, which is a popular framework for IT governance and IT management (IT Governance Institute 2008).

Creation

A piece of data is created at some time. Since a good start is half the battle, guidelines about that data are to be present right when data is created. An example is given in Example 9.

> ***Example: registering new members***
>
> An operator registers a new member in the member system. Data quality guidelines should state that all zip codes in the member system are to be stored in the form “1234 AB”. The operator can take these guidelines into account.

Example 9 – An operator registering new members in an information system

Usage

Most data will be used multiple times during its lifecycle, and there should be guidelines available for that. For instance, client information may be accessible to sales persons but not to all consultants.

Alteration and maintenance

Just like physical articles, data will get altered during its lifecycle, which may result in needed maintenance. If someone decides that the zip code format changes to “1234AB” to match it to another system, all zip codes need to be updated.

Backup

Data needs to have some kind of availability. Since infrastructure can fail, it should be backed up. Specialised tools such as hierarchical storage management (HSM) can support this process (Rouse 2015).

Deletion or archiving

Data may need to be deleted or archived. Archiving periods are usually based on regulatory requirements, for instance, sensitive information may need to be deleted within a certain period of time. On the other hand, in The Netherlands, financial information should remain available for at least seven years and also supervisory bodies demand that certain information is available at any time.

5.3 Metrics of the model

In the previous section, we defined the structure of the maturity model. We will now discuss the arrangement of the factors and criteria, and how the scores are calculated.

5.3.1 Arrangement of the factors and criteria

In the previous section, we defined factors that are decomposed into several criteria. We consider these factors and related criteria two dimensions of data governance maturity. We arrange them in a two-dimensional matrix, meaning that each governance element is present in each data governance domain (Figure 16).

Governance elements	Data governance domains				
	Data assets	Data quality	Meta-data	Data access	Data lifecycle
Objectives	A1	A2	A3	A4	A5
Tasks	B1	B2	B3	B4	B5
Roles	C1	C2	C3	C4	C5
Responsibilities	D1	D2	D3	D4	D5

Figure 16 – Arrangement of the criteria and factors

5.3.2 Metric per cell

The cells represent the intersection of two criteria of separate dimensions. For each cell, we drafted a number of questions. Each question can be answered on a five-point Likert scale. The score per cell is the percentage of the maximum score per cell.

5.3.2.1 Likert-scale

Each question is to be answered on a five-point Likert scale. In an earlier version of the model, we used a seven-point Likert scale. Participants indicated that this gave them too much options, after which we decided to lower it to a five-point scale. Research has shown that this will likely not distort the data in terms of variation about the mean and skewness and that a five-point scale is the minimum scale for sufficient discriminating power (Dawes 2008; Preston & Colman 2000).

5.3.2.2 Scale descriptions

The description given to the scale points differs for several questions. There are five different descriptions for the Likert scale, summarised in Table 14.

Example question	*Minimum*	*Neutral*	*Maximum*
Is there a data asset overview?	No	Yes, implicitly	Yes, explicitly
Are external data assets considered?	Is there, or not needed		Needed, but not there
Does each data asset have an owner defined?	No	Yes, implicitly	Yes, in function profile
Is the overview of data assets maintained?	Yes		No
How many persons with some level of responsibility over data assets, are aware?	None	Half	All

Table 14 – Overview of the descriptions on the Likert scale

5.3.2.3 Ordinal, interval and ratio scales

Likert-scales are ordinal, meaning that items on the scale are *ordered*. For instance 'strongly agree' is better or higher than 'agree' and 'agree' is better or higher than 'neutral'. However, the difference between two items need not be the same, for instance the difference between 'strongly agree' and 'agree' maybe be higher or lower than difference between 'agree' and 'neutral'.

This limits the amount of descriptive statistics that can be applied. For instance, one is allowed to calculate frequency distributions ("*19%* of the people scored *strongly agree* on question 3") or mode ("the most scored item on question 3 is *neutral*"), but not mean and standard deviation ("the average response over all questions for this respondent is *2.3* with a standard deviation of *0.2*"), correlation ("participant 1 and 2 have a strong positive correlation of *0.8*) or summation and fraction ("the respondent scored *81* points with a maximum of *101*, which is *80%*").

For mean, standard deviation, and summation at least an interval scale where the difference between two items is the same is needed. For calculating a fraction, a ratio scale with an absolute zero is required. Since we desire to use descriptive statistics such as fraction, we define that the items on the Likert-scale linearly correspond to natural numbers, making it an interval scale. Since we use an absolute zero, it becomes a ratio scale (Table 15).

Example question	*0 points*	*1 point*	*2 points*	*3 points*	*4 points*
Is there a data asset overview?	None		Neutral		Maximum

Table 15 – Linear mapping between the ordinal and ratio scale

5.3.2.4 Percentage of the maximum score

The score *per cell* is illustrated as the percentage of the maximum score (Table 16). Each cell contains multiple questions, of which the given scores are summed. This number is divided by the maximum score, which is the number of questions in that cell multiplied by the maximum score per question.

Cell A1	*0 points*	*1 point*	*2 points*	*3 points*	*4 points*	*Score*	
Question 1	X					0	
Question …		X				1	
Question n					X	4	
					Total score	5	+
					Number of questions * maximum score per question	3 * 4 = 12	/
					Percentage of the maximum score	42%	

Table 16 – Calculation of the score per cell

5.3.2.5 *Metric per class and total*

The scores per cell can be aggregated into a score per class and a total score. These aggregated scores are the weighted average of the score per cell and the importance (weight) per cell. We determined the importance based on the theoretical framework we outlined.

A common wisdom in business science is that programs that have a clear ***objective***, perform better (sub-sub-section 5.2.1.1). As stated in section 3.2 and sub-sub-section 5.2.2.1, data governance is about treating data as an organisational asset, hence about ***data assets***. Finally, one may also say the primary reason for a data governance program is achieving high ***data quality*** (section 3.5 and sub-sub-section 5.2.2.2). We consider these three classes to be of greater importance to data governance than the others and weigh them with a factor 2. The importance per cell is determined by multiplying the importance of the related classes (Figure 17).

			Data governance domains				
			Data assets	Data quality	Meta-data	Data access	Data lifecycle
			2	2	1	1	1
Governance elements	Objectives	2	4	4	2	2	2
	Tasks	1	2	2	1	1	1
	Roles	1	2	2	1	1	1
	Responsibilities	1	2	2	1	1	1

Figure 17 - The weights to calculate the weighed totals

The sub-total per class is determined using the weighted average of all cells in that class. The total score is calculated using the weighted average of all cells (Figure 18).

Governance elements	Data governance domains: Data assets	Data quality	Meta-data	Data access	Data lifecycle	**Weighed total**
Objectives	%	%	%	%	%	%
Tasks	%	%	%	%	%	%
Roles	%	%	%	%	%	%
Responsibilities	%	%	%	%	%	%
Weighed total	%	%	%	%	%	%

Figure 18 – The model used to score data governance maturity on the individual cells, the classes and the whole attribute

5.4 Questions of the model

The questions that are to be scored in order to determine an organisation's data governance maturity are discussed in the following sub-sections. They are summarised in Table 17 and Table 18.

5.4.1 Data assets

A1 – Objectives for data assets

Knowing an organisation's information requirements and their connection to business processes, data assets and applications that connects these, helps in understanding how data is used in an organisation and who should be responsible for certain parts.

Questions for this cell are if information requirements and data assets are known, if there is a connection between them, if external data assets are considered, and if data assets are prioritised according to confidentiality, integrity and availability.

B1 – Tasks for data assets

Tasks that need to be performed are derived from the objectives and include maintaining an information requirements and data asset overview and maintaining the connection between these. Questions are targeted at indicating if these tasks are performed.

C1 – Roles for data assets

The aforementioned tasks can be assigned to specific roles. Moreover, each data asset needs to be assigned an owner and one or more stewards (delegated responsibility), someone high in the organisation[13] should promote the data governance program (sponsor) and a data council can be considered[14].

D1 – Responsibilities for data assets

The aforementioned tasks and roles should be assigned an appropriate steward for the ARCI acronym. Questions include if data ownership is actually assigned to persons rather than departments, if ownership is assigned to highest management, if persons that delegated responsibility actually support their delegate, and if all persons that have some level of responsibility over data assets are aware of that responsibility and agree with it.

5.4.2 Data quality

A2 – Objectives for data quality

Objectives for data quality are derived from the objectives for data assets and include if the importance of data quality to the business is known and if that importance is broken down for the several data assets.

B2 – Tasks for data quality

Tasks that are to be performed for data quality include maintaining the importance of data quality to the business, being able to measure data quality for all data assets and actually monitoring data quality in these assets.

C2 – Roles for data quality

Roles for data quality include ownership for data quality for several data assets. This will likely be the same as ownership of the data asset itself, but organisations are free to choose differently. Other questions are about maintaining the importance of data quality to the business, making data quality definitions in order to measure, and actually monitoring data quality.

D2 – Responsibilities for data quality

The questions are the same as with data assets, but specified to data quality.

5.4.3 Metadata

A3 – Objectives for metadata

Objectives for metadata are if the importance of metadata to the business is known and if it is split for the several data assets.

B3 – Tasks for metadata

Tasks for metadata are about if there is an active translation between technical data and its meaning, if descriptions and meanings of data are documented, and if there is a periodic evaluation of all metadata.

C3 – Roles for metadata

Metadata roles include ownership of all metadata, maintaining the description and meaning of data assets, and the periodic evaluation.

---

[13] Recall: high in the organisation, but not necessarily on the executive level (sub-sub-section 5.2.1.4).

[14] Recall: a data council can help, but is no determinant for a program's success (sub-sub-section 5.2.1.4).

	Data assets	*Data quality*	*Metadata*	*Data access*	*Data lifecycle*
Obectives	Are information requirements known?	Is the importance of data quality to the business known?	Is the importance of metadata to the business known?	Are there company guidelines for data access?	Are there company guidelines for data life cycle?
	Is there a data asset overview?	Is that importance split for the valuable data assets?	Is that importance split for the valuable data assets?	Do company guidelines include physical data access?	Do company guidelines include compliance requirements for retention & archiving?
	Are information requirements connected to specific data assets?			Do data access guidelines include compliance requirements for auditability?	
	Are external data assets considered?			Do data access guidelines include compliance requirements for privacy?	
	Is the importance of data assets prioritised according to confidentiality?			Do data access guidelines include compliance requirements for availability?	
	Is the importance of data assets prioritised according to integrity?				
	Is the importance of data assets prioritised according to availability?				
Tasks	Is the overview of information requirements maintained?	Is the importance of data quality to the business being maintained?	Is there a translation between data (technical) and interpretation (business)?	Are guidelines for data access being maintained?	Are guidelines for data lifecycle being maintained?
	Is the connection between information requirements and data assets maintained?	Can data quality be measured for all valuable data assets?	Is there documentation of descriptions and meanings of data?	Is there a periodic evaluation of data access?	Is there a periodic evaluation of the data lifecycle for all valuable data assets?
	Is the overview of data assets maintained?	Is data quality being monitored for all valuable data assets?	Is there a periodic evaluation of metadata for all valuable data assets?		Is there a periodic maintenance of data for all valuable data assets?
					Is there a periodic archiving or deletion of data for all valuable data assets?

Table 17 – Overview of the model's questions (part 1)

	Data assets	Data quality	Metadata	Data access	Data lifecycle
Roles	Does each data asset have an owner and steward defined in function profiles?	Is the accountability for the level of data quality for all valuable data assets listed in function profiles?	Is the accountability for the metadata for all valuable data assets listed in a function profile?	Is the accountability for the data access for all valuable data assets listed in a function profile?	Is the accountability for the data lifecycle for all valuable data assets listed in a function profile?
	Is the steward for maintaining the information requirements overview listed in a function profile?	Is the steward for maintaining the importance of data quality to the business listed in a function profile?	Is the steward for maintaining data descriptions and meanings for all valuable data assets listed in a function profile?	Is the steward for the periodic evaluation of data access for all valuable data assets listed in a function profile?	Is the steward for the periodic evaluation of data lifecycle for all valuable data assets listed in a function profile?
	Is the steward for maintaining the connection between information and data listed in a function profile?	Is the steward for measuring data quality for all valuable data assets listed in function profiles?	Is the steward for periodic evaluation for all valuable data assets listed in a function profile?		Is the steward for the periodic maintenance of data for all valuable data assets listed in a function profile?
	Is the steward for maintaining the data asset overview listed in a function profile?	Is the steward for monitoring data quality for all valuable data assets listed in function profiles?			Is the steward for the periodic archiving or deletion of data for all valuable data assets listed in a function profile?
	Does someone in the organisation sponsor a proper usage of data?				
	Is having a data council considered?				
	Is having business stewards considered?				
Responsibilities	How many entities that are listed accountable (A) for data assets are a person?	How many entities that are listed accountable (A) for data quality are a person?	How many entities that are listed accountable (A) for metadata are a person?	How many entities that are listed accountable (A) for data access are a person?	How many entities that are listed accountable (A) for data lifecycle are a person?
	Is data ownership (A) in het highest management?	How many persons accountable support their person responsible?	How many persons accountable support their person responsible?	How many persons accountable support their person responsible?	How many persons accountable support their person responsible?
	How many persons accountable support their person responsible?	How many persons that have some level of responsibility over data quality, are aware?	How many persons that have some level of responsibility over metadata, are aware?	How many persons that have some level of responsibility over data access, are aware?	How many persons that have some level of responsibility over data lifecycle, are aware?
	How many persons that have some level of responsibility over data assets, are aware?	How many persons that have some level of responsibility over data quality, agree?	How many persons that have some level of responsibility over metadata, agree?	How many persons that have some level of responsibility over data access, agree?	How many persons that have some level of responsibility over data lifecycle, agree?
	How many persons that have some level of responsibility over data assets, agree?				

Table 18 – Overview of the model's questions (part 2)

D3 – Responsibilities for metadata

The questions are the same as with data assets and data quality, but specified to metadata.

5.4.4 Data access

A4 – Objectives for data access

Data access objectives indicate if there are company guidelines for data access, including physical access to the servers, and compliance requirements for auditability, privacy and availability.

B4 – Tasks for data access

Data access tasks are about maintaining these data access guidelines and periodically evaluating them.

C4 – Roles for data access

Roles for data access indicate ownership for data access[15], and responsibility for a periodic evaluation.

D4 – Responsibilities for data access

The responsibility questions are the same as with data assets, data quality and metadata, but specified to data access.

5.4.5 Data lifecycle

A5 – Objectives for data lifecycle

Data lifecycle objectives questions indicate data lifecycle guidelines, including compliance requirements for retention and archiving.

B5 – Tasks for data lifecycle

Tasks for the data lifecycle include maintaining guidelines, periodically evaluating them, a periodic maintenance on the data assets, and a periodic archiving and retention.

C5 – Roles for data lifecycle

Data lifecycle roles indicate ownership for data lifecycle and responsibility for a periodic review of the guidelines, periodic maintenance and periodic archiving and retention.

D5 – Responsibilities for data lifecycle

Responsibility questions for data lifecycle are similar to those for data assets, data quality, metadata and data access.

5.5 Instructions

Instructions on how to use a method enhance that the method will be used as proposed by the authors. Organisations that want to assess data governance maturity are advised to follow the seven steps that are summarised in Table 19.

[15] Whereas ownership for data quality and metadata almost exclusively lies at the owners of the related data asset, ownership of data access is quite commonly assigned to a specific security officer.

Step	*Short description*
1. *Role assignment*	Assigning roles for the maturity assessment
2. *Data collection*	Collecting all the necessary facts
3. *Execution*	Performing the maturity assessment
4. *Interpretation*	Interpreting the results and summarising them in a report
5. *Validation*	Validate the interpretation with data stakeholders
6. *Decision-making*	Decide whether data governance maturity is sufficient.
7. *Diffusion*	Changing the current state, if needed and desired.

Table 19 – Summary of the proposed steps for a data governance maturity assessment

5.5.1 Step 1: Role assignment

The organisation that wants to assess its data governance maturity should prepare for the assessment by assigning two primary roles involved with the assessment. One person should execute the assessment and someone high in the organisation should sponsor the assessment.

Assessor

The assessor should execute the maturity assessment. The assessor needs to be familiar with governance, which will probably mean he or she is not too low in the organisation and has a sufficient conceptual thinking level. Since the data will be collected from several people, the assessor should have sufficient communication skills. The person needs to collect information from both business and IT, hence needs to be familiar with both. A person with all these skills may operate in a function as business analyst or enterprise architect.

If these skills are not present in the organisation, or if the organisation values an objective assessment, an external consultant can perform the assignment. However, informal communications are often needed to gather all required information and the external consultant should be supported in finding these.

In large organisations, multiple assessors may be assigned. We will assume one assessor.

Sponsor

Since time of multiple people is needed to perform the assessment and significant changes may be needed afterwards, someone high in the organisation should sponsor the project. He or she should let the organisation know that the project is important and that people should cooperate.

5.5.2 Step 2: Data collection

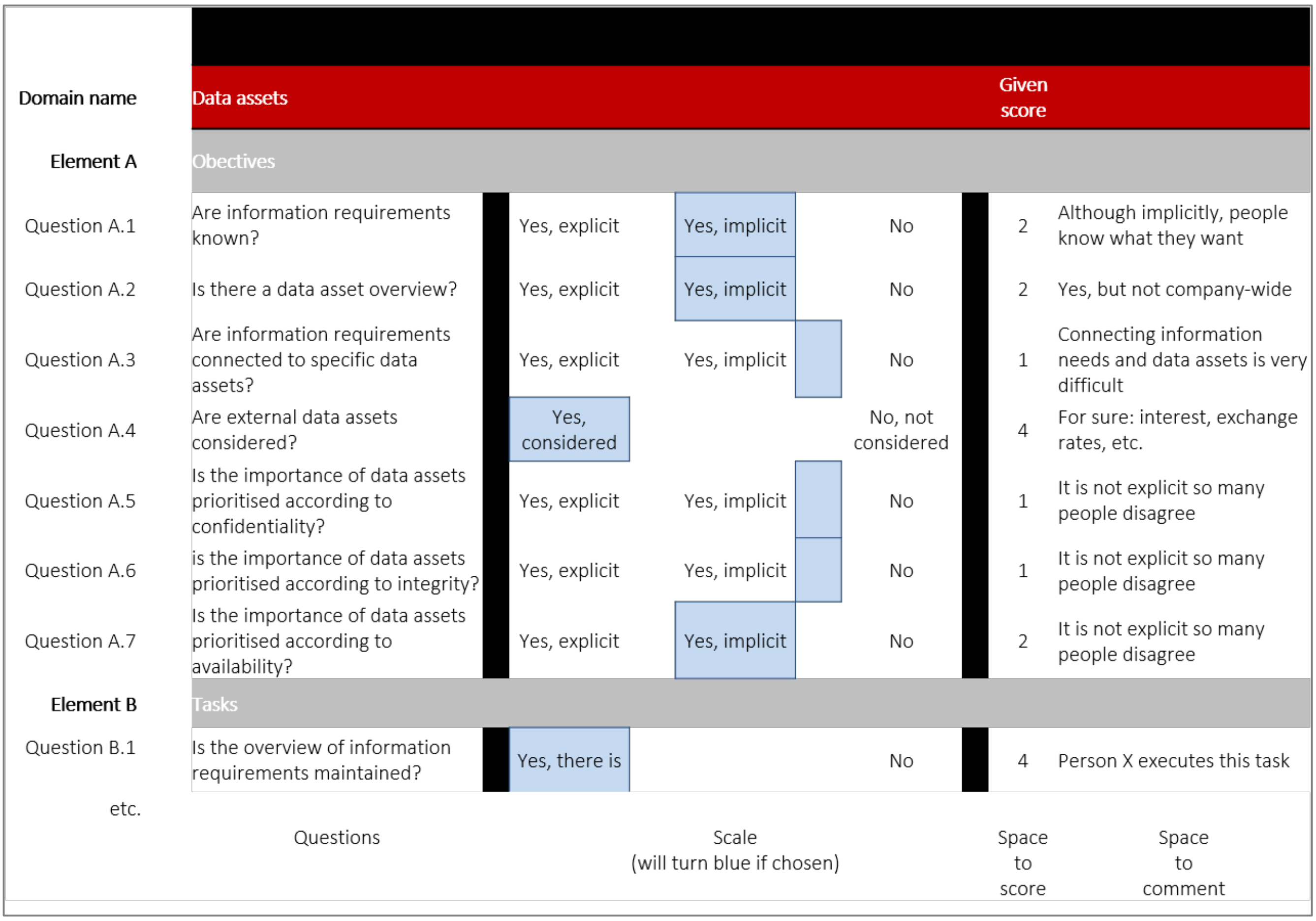
The next step is to collect all the data needed to execute the assessment. What data is required can be determined from the maturity model's question (section 5.4). Data is typically collected by interviewing relevant stakeholders and by studying documentation, for instance a data asset overview or security guidelines.

The assessor is advised to start by talking with executive management about IT and data challenges for the future. Next, the assessor should talk to the business process, data asset and application owners. Finally, the assessor should interview persons that are, either explicitly or implicitly, responsible for the specific data governance domains, which are functions as enterprise architects, business analysts, business intelligence specialists, security specialists and database administrators.

5.5.3 Step 3: Execution

Once all required information is collected, the scorecard needs to be filled in. To promote objectivity, the assessor should do this together with someone else who is familiar with the governance of data.

The scorecard is meant to enter in all the collected data and to calculate the scores per cell, the subtotals and totals. It is an Excel file with one tab for each domain (Figure 19). Each tab is vertically divided into the elements and related questions. Horizontally, one finds the questions, the scale, space to score the question, and space to comment. When a score is given, the relevant part of the scale lights up in blue.



Domain name	Data assets				Given score	
Element A	Obectives					
Question A.1	Are information requirements known?	Yes, explicit	Yes, implicit	No	2	Although implicitly, people know what they want
Question A.2	Is there a data asset overview?	Yes, explicit	Yes, implicit	No	2	Yes, but not company-wide
Question A.3	Are information requirements connected to specific data assets?	Yes, explicit	Yes, implicit	No	1	Connecting information needs and data assets is very difficult
Question A.4	Are external data assets considered?	Yes, considered		No, not considered	4	For sure: interest, exchange rates, etc.
Question A.5	Is the importance of data assets prioritised according to confidentiality?	Yes, explicit	Yes, implicit	No	1	It is not explicit so many people disagree
Question A.6	is the importance of data assets prioritised according to integrity?	Yes, explicit	Yes, implicit	No	1	It is not explicit so many people disagree
Question A.7	Is the importance of data assets prioritised according to availability?	Yes, explicit	Yes, implicit	No	2	It is not explicit so many people disagree
Element B	Tasks					
Question B.1	Is the overview of information requirements maintained?	Yes, there is		No	4	Person X executes this task
etc.						
	Questions	Scale (will turn blue if chosen)			Space to score	Space to comment

Figure 19 – The method's scorecard

Once all questions are answered, the scores are automatically calculated. Now scores can be imported in the statistics tool R. A chart with horizontal bar charts for each cell, sub-total and total is produced. The size of the bar corresponds to the calculated score. An example chart is illustrated in Figure 20.

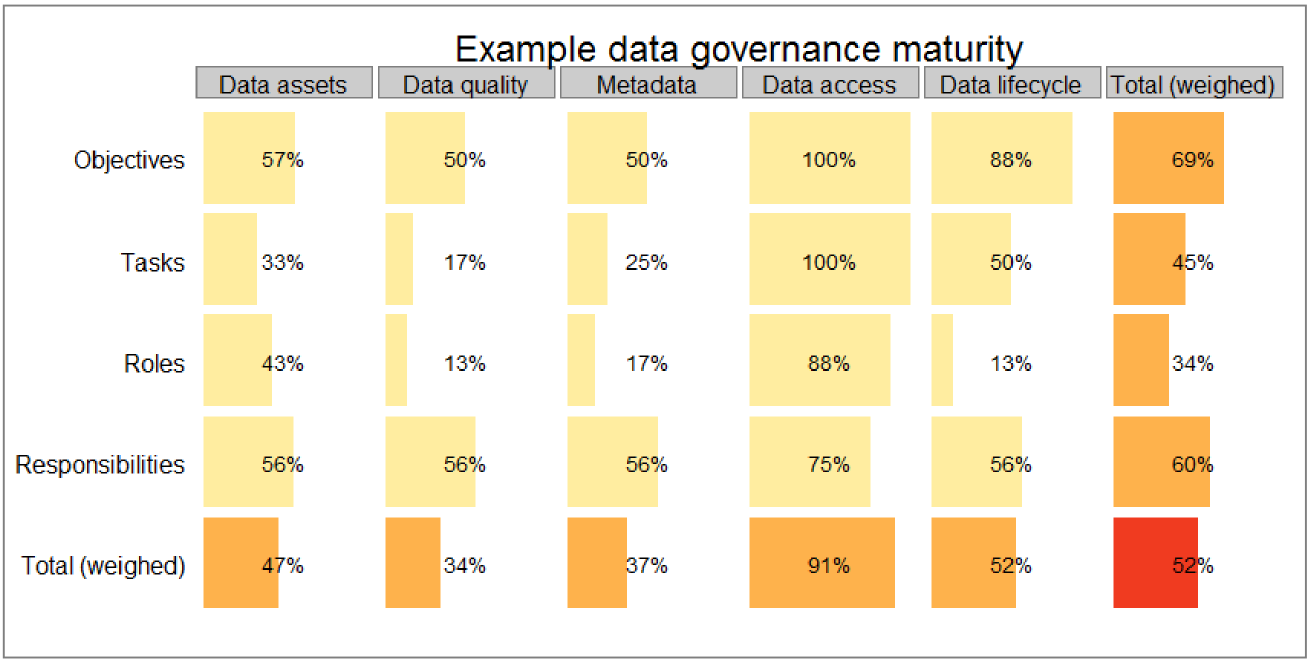


Figure 20 – Example data governance maturity scores

5.5.4 Step 4: Interpretation

The data governance maturity model is a tool to quickly assess and illustrate data governance maturity. The model is purely descriptive, meaning it does not state whether improvements are needed. Therefore, interpretation of the model's outcomes is always needed. To emphasise the difference between the objective, descriptive model and the subjective, prescriptive interpretation, we separate these parts over different steps.

In future work, we aim to structure the interpretation step with an interpretation framework (section 7.5). For now, we provide the assessor with instructions on how to interpret results from the maturity model.

As-is data landscape

The assessor should start by analysing the current data landscape within the organisation. This includes considering what data assets there are and how they are used, how the data organisation is composed, and what problems arise concerning data.

To-be data landscape

Organisations that are assessing their data governance maturity will have goals on what to achieve with data. Goals may be classified as reasons to implement a data governance program (section 3.1, Otto 2011).

Gap analysis

Once both the as-is and to-be data landscape are clear, the assessor should analyse how similar these two landscapes are and identify gaps.

Data governance maturity scores

Now that gaps are identified, the assessor should check if these can be connected to specific data governance maturity scores.

List these as actionable recommendations

Finally, for gaps that can be linked to specific maturity scores, the assessor should recommend on how to bridge the gap. Organisations should never strive for 100% on all classes, but find the right match between needs and how much effort it takes to get there. Recommendations can be rated on their expected effort (low-medium-high) and expected benefit (low-medium-high; Table 20). After that, the commendations can be arranged in an effort-benefit matrix (Figure 22). A proposed order of execution can also be given (Table 20).

An example of how the steps until deriving to the actionable recommendations can be executed, is illustrated in Example 10.

Report

The assessor can summarise his interpretation of the maturity results in a report. An example outline is illustrated in Figure 21.

1. Table of contents
2. Management summary
3. Explanation of the model
    a. Domains and elements
    b. Usage of the model
    c. Difference with CMMI's
4. Research method
5. As-is data landscape
6. To-be data landscape
7. Results
    a. Data assets
    b. …
    c. Responsibilities
8. Gap analysis
9. Recommendations
    a. Table
    b. Effort-benefit-matrix
    c. Recommendation 1
    d. …
    e. Recommendation *n*
    f. Proposed order
10. Limitations
11. Appendices
    a. Interviewee list

Figure 21 – Example outline of the report

5.5.5 Step 5: Validation

To ensure that the assessor collected the right information and to discuss the interpretation he or she made, the assessor should send a draft report to the persons that he interviewed and discuss their feedback. The assessor may decide to formally measure their validity, such as illustrated in section 2.8.

All persons' feedback should be processed and the report should be finalised.

5.5.6 Step 6: Decision-making

If management agrees with the assessor's interpretation and recommendations, it should decide who should be responsible for what part of the change, and assign appropriate budgets if necessary.

5.5.7 Step 7: Diffusion

A large part of data governance is about awareness and organisational change. If management decided that change is needed, the assessor should actively spread the results and involve persons in the change process. A good way to do so is by organising an improvement workshop.

As-is data landscape

There are two data assets: customer data and financial data. The owners are the process owners of sales and finance, respectively. There are two database administrators that form the data organisation. All data operate independently. If connections are needed, the database administrators make these manually.

To-be data landscape

The organisations desires to more quickly gather management information, for instance how much money each customer spends.

Gap analysis

The data sources for customer and financial data need to be integrated, but are currently not.

Data governance

Using the results chart (Figure 20), the assessor can easily sport areas that should be improved, some of which are needed to bridge the gap. For instance:

- Objectives for data assets are only 57%. Knowing the link between information requirements and data assets will result in knowing how exactly the integrated data source will be used;
- Most tasks and roles are low (e.g. 13 and 17%). Clarifying these will result in better knowing who will operate the new data source, for instance its archiving;
- Metadata scores low, between 13 and 50%. Using a data dictionary containing all relevant business definitions, we know wat definitions of the client are present in both sources.

Actionable recommendations

The identified areas should be improved. The assessor can determine actionable recommendations (Table 20) that can be arranged in an effort-benefit matrix (Figure 22).

#	*Recommendation*	*Benefit*	*Effort*	*Order*
1.	Explicitly state information requirements, data assets, and the link between them	High	High	1
2.	Explicitly state accountability and responsibility for data	High	High	2
3.	Draft metadata guidelines	Medium	Medium	3

Table 20 – Prioritised requirements

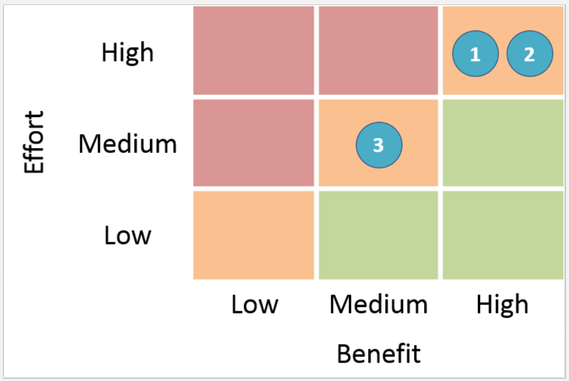


Figure 22 – Effort-benefit matrix

Example 10 – How to execute the first steps of the maturity method

5.5.7.1 Improvement workshop

First, the assessor and the sponsor should determine the workshop's goal. An example goal is to plan how to execute the recommendations. Attendees should be at least the data sponsor, a technical and a business person involved with data, an architect and business analyst, and persons involved with specific data governance domains (Table 21).

Role	*Task in the workshop*
Data sponsor	Without a sponsor, the workshop's outcomes have no place in the organisation. The data sponsor should stretch necessity for the workshop, be the connection to highest management and assign persons to execute tasks that arise during the workshop
Technical persons involved with data	Offer technical insights into difficulties with data
Business persons involved with data	Suggest information requirements and offer business insights into difficulties with data
Architect, if any	Discuss the existing data and information architecture
Business analyst, if any	Provide actual problems; help in translating between business and IT
Persons involved with the domains, if any	Provide insight into data quality, metadata, data access and data lifecycle practices.

Table 21 – Attendees of a data governance improvement workshop

The workshop should start with brief introduction of the workshop's aim, why data governance helps with a better usage of data and some terminology. Terminology includes the difference between information and data and the four levels of responsibilities (ARCI).

To make the rather conceptual and formal concept of data governance more specific to the participants, we advise the assessor to present items he noticed during the maturity assessment. These may be persons that are not aware of their responsibilities, several different answers to the same question, or problems with poor data quality.

Next, the assessor should introduce the data governance maturity method by explaining the domains and elements. After this, he or she can introduce the process that was followed to derive the maturity score, show the results chart and highlight the most striking areas.

Finally, there is the part where we cover the organisation's goal for the workshop. One possibility is to jointly come up with the organisation's most important data assets, including who should be accountable and responsible. Another option is to jointly stocktake problems with data that can be improved with data governance.

Given these contents, the workshop should last at least two hours.

6 RESULTS

The method was evaluated in two case studies at a large and a small IT-intensive organisation. We will discuss how possible users of the model value the data governance maturity method and what data governance looks like at these case study organisations.

6.1 Collection of the results

Qualitative results are collected from the evaluation session, case study, improvement workshop and feedback on the report. These qualitative results are enriched with quantitative results, which are based on answers on the evaluation questionnaire and scorecards that were collected during the evaluation sessions. With fifteen and ten participants, depending on the type of test, we will not be able to draw statistically significant conclusions, which is why we will only use them to enrich the qualitative results. However, these quantitative results are supported by remarks that the persons gave us during the case studies.

Fifteen persons from Bank and ANWB participated in the evaluation sessions. With all, we discussed their feedback on the report, after which they conducted the validation questionnaire. Ten of these persons also participated in the acceptance part, which consisted of using the method's scorecard and afterwards answering some acceptance questions about their experiences. We calculated inter-rater reliability using the filled-in scorecards from the acceptance tests.

Although the case study at Software Improvement Group was conducted using an earlier version of the method, we gathered enough data to later assess their data governance maturity using the new method. Therefore, most qualitative results gathered at Software Improvement Group are included in the results. Quantitative results at Software Improvement Group are excluded.

6.2 User groups and organisations

Where applicable, the results are broken down into results per organisations and per function category. We created three function categories, illustrated indicated in Figure 23. Horizontally, one can see the three levels in an organisation: strategy to determine the organisation's future plans, tactics that supervises operations, and operations that executes most of the actual work (Cheong & Chang 2007). Vertically are the enterprise architecture layers (see section 3.6), indicating how IT (bottom four layers) supports the business (top layer). The three function categories are executives, pivots, and others.

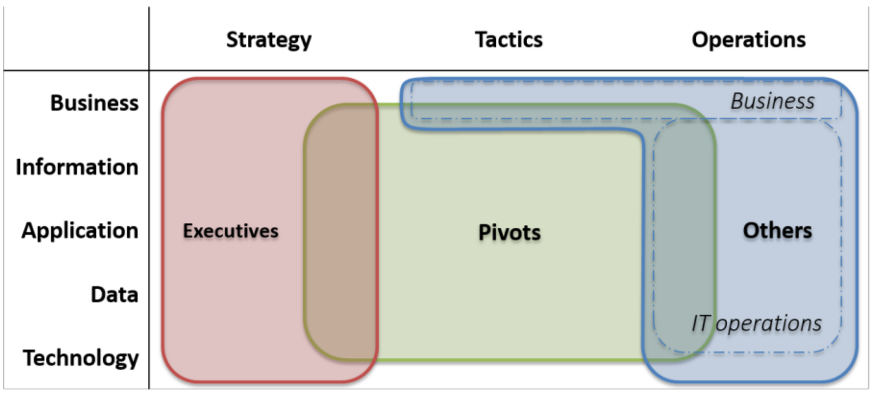


Figure 23 – Three function categories, used to breakdown the results

Executives, both at business and IT, are supposed decide what to do with the method's outcomes, so we primarily value their opinion for validity. Due to time constraints, most executives did not participate in the acceptance test. Example executive functions are Chief Financial Officer (CFO) and Chief Information Officer (CIO).

The group *pivots* consists of persons that have a function operating between the business and IT, for example business analysts and enterprise architects. We expect people in this group to understand the method and its terminology and are likely candidates to perform the assessment. Therefore, we primarily value their opinion on acceptance of the scorecard.

The group *others* consists of two subgroups: persons that work on the business side, and those on the IT operations side. Example functions are financial controller and security specialist. Persons involved with business generally have relatively little knowledge of IT; persons on IT operations will portably not have enough overview of the organisations to understand implications of the method. Since for each subgroup there is a low number of participants and we expect both subgroups to have difficulty with the method, we merged them into one group. We expect that validity and acceptance scores in this group will be relatively low.

6.3 Construct validity

This section deals with if participants believe the outcomes of the method reflect data governance within their organisation, if they agree with the authors' score and if they prefer this method to their current method.

6.3.1 Method's validity

Using the developed data governance maturity method, we were able to assess data governance maturity at the two case study organisations. Employees of these organisations think the method is valid. People' reactions were that "it is a great help in asking the right questions", "it is easy to mark areas for improvement" and that "it can be used as a checklist for improving data practices, even by persons lower in the organisation". People value most its structured explanation of the current state and its actionable recommendations.

Our time investment at each organisation was about 50 hours. This indicates that although data governance is a complex and multidimensional notion, only a relatively small time investment is needed to identify what current data governance looks like and to mark areas for improvement.

The positive attitude towards the method is reflected in the questionnaire results (Figure 24).

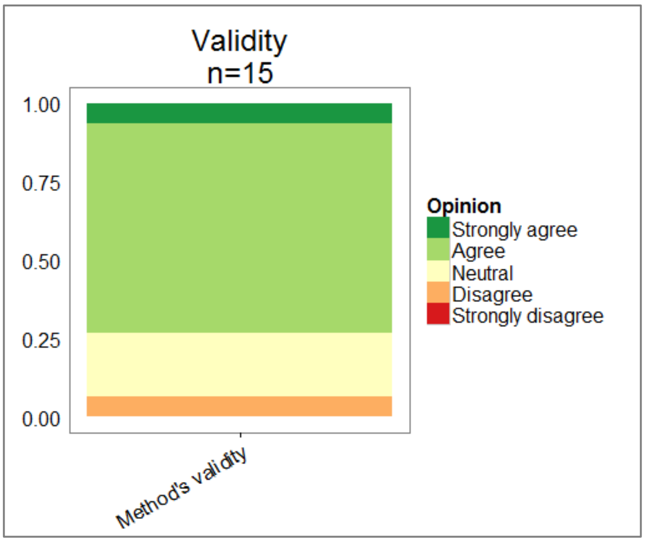


Figure 24 – Questionnaire results on the method's validity

We can see that when asking for the method's validity, 64% agrees and 7% even strongly agrees. Persons that found the method difficult to use, mostly had a neutral (21%) rather than a negative attitude towards its validity.

Persons that disagree (7%) question if the method consists of the right questions to determine data governance maturity. One person that did not participate in the evaluation sessions told us he strongly disagrees with the method since it provides insufficient support to interpret the results. We also experienced difficulty in advising on improvements.

Another person suggested to allow organisations to choose their own weights. This makes it harder to benchmark organisations, which is why we think it is not a good suggestion.

The method separates description of the maturity results from interpreting them by covering these in different steps (section 5.5). In future work, we suggest to structure the interpretation step by developing an interpretation framework (Figure 25). We further elaborate on this in section 7.6.

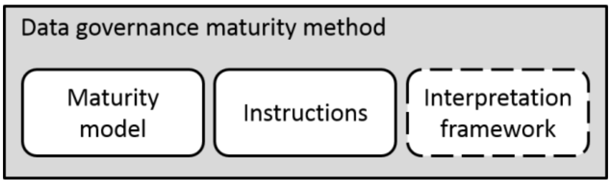


Figure 25 – In future work, we aim to extend the method with an interpretation framework

6.3.2 Agree with authors

As reflected in the Figure 26, 71% of the persons agree with the score that the authors gave about their organisation.

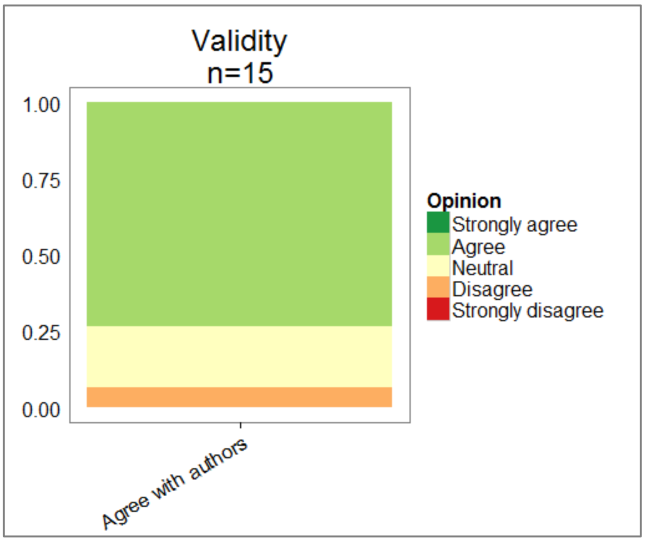


Figure 26 – Questionnaire results if participants agree with the authors

We noticed that, even when participants disagree with the authors' score, there is consensus on the classes that score high and those that score low. This indicates that, despite disagreement, general conclusions about an organisation's data governance maturity still hold.

6.3.3 Prefer over current

As reflected in Figure 27, 71% prefers the provided method to assess data governance maturity to the current available method: 36% agrees and also 36% even strongly agrees.

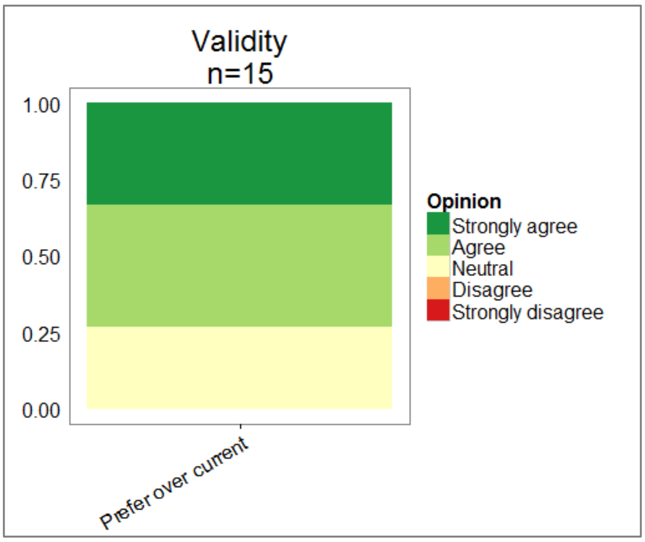


Figure 27 – Questionnaire results if the participants prefer the developed method over their current method

Participants indicated that this is largely because their organisation currently has no method to assess data governance maturity. A director of Bank stated that this was the first time that someone attempted a structured approach to inventory responsibility over data and was very pleased with the results.

6.3.4 Function category breakdown

We combined the three validity questions in one chart and broke the results down for the three function groups we defined in section 6.2 (Figure 28).

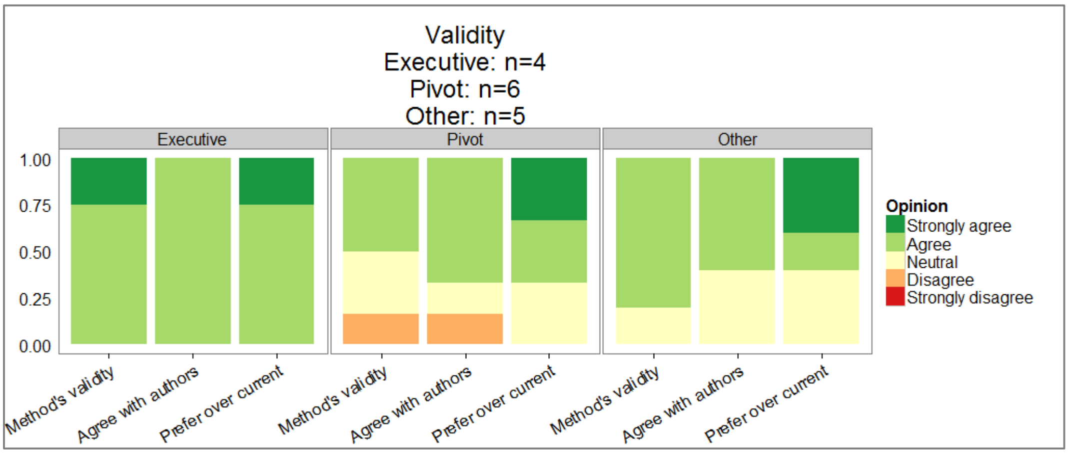


Figure 28 – Validity results, broken down for the function categories

6.3.4.1 Executives

In Figure 28, we can see that executives (*n*=3) without an exception find the method valid; across all validity questions 78% agrees and 22% even strongly agrees with the validity of the method.

Executives' heir agreement on the method's validity means there is a good chance they will use the outcomes to improve data governance in their organisation.

6.3.4.2 Pivots

Pivots (*n*=6) predominantly rate the method's validity positively (61% agrees or strongly agrees across all validity questions). However, 11% of the questions has a negative score. This is the lowest score of the three function categories. We contribute the relative negativity to pivots' high knowledge of the intersection of business and IT, which allows them to see flaws that others do not.

6.3.4.3 Others

Other than expected, most of the 'others' find the model valid (n=5; 67% across all validity questions), whereas we thought persons in this category would have difficulty with it and would rate it negatively. This tells us that the method's results are understandable to a wider range of functions than expected.

6.3.5 Organisation breakdown

The results are also broken down per organisation. As illustrated in Figure 29, both Bank and ANWB are predominantly positive: 77% and 70% across all validity questions, respectively.

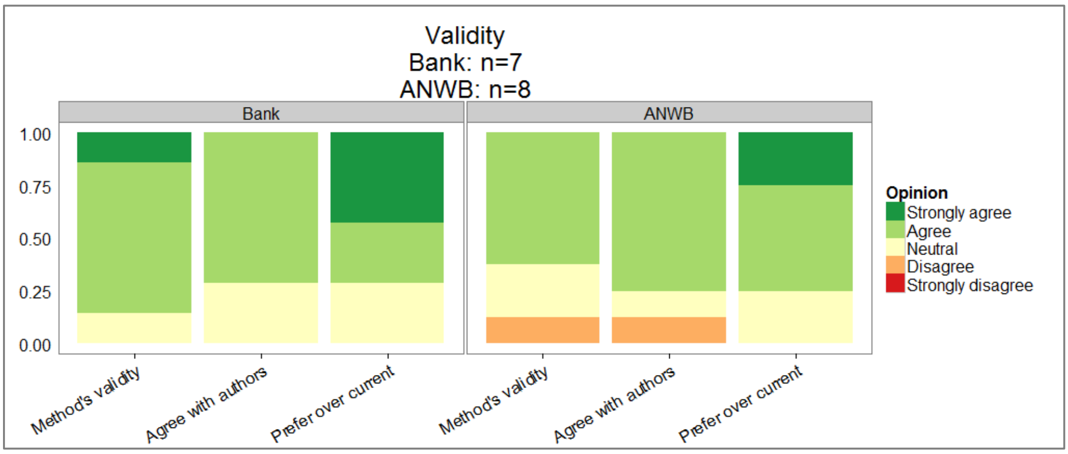


Figure 29 – Validity results, broken down for the organisations

Executives at Bank stated that the method can help them improve their data practices and asked us to present the results to the management team, in addition to organising a three-hour improvement workshop in which nine persons participated.

Validity scores at Bank are slightly higher than at ANWB. ANWB is a much larger and more diverse organisation than Bank (see section 2.10). Due to this, eight interviews turned out to be not enough to capture sufficient complexity of ANWB's data governance, which we believe is the reason for more negative validity scores. In such an organisation, one should either focus on a part of the organisation and/or plan more interviews. We estimate that for the whole of ANWB, about twice as much interviews and twice as much time is needed.

6.3.6 Distinctness

We noticed that the Pareto principle is applicable to data governance: with a relative small amount of time, organisations can get a lot done, but getting things perfect takes significant time. We believe that this principle is insufficiently covered by the method.

To increase the distinctness between organisations, we can use for instance an exponential scale where the difference between two steps increases as the score gets higher. We wish to explore this in future research (section 7.5)

6.3.7 Split if participated in acceptance testing

As indicated in Figure 30, persons that did not participate in the acceptance test scored more extreme than persons that did participate.

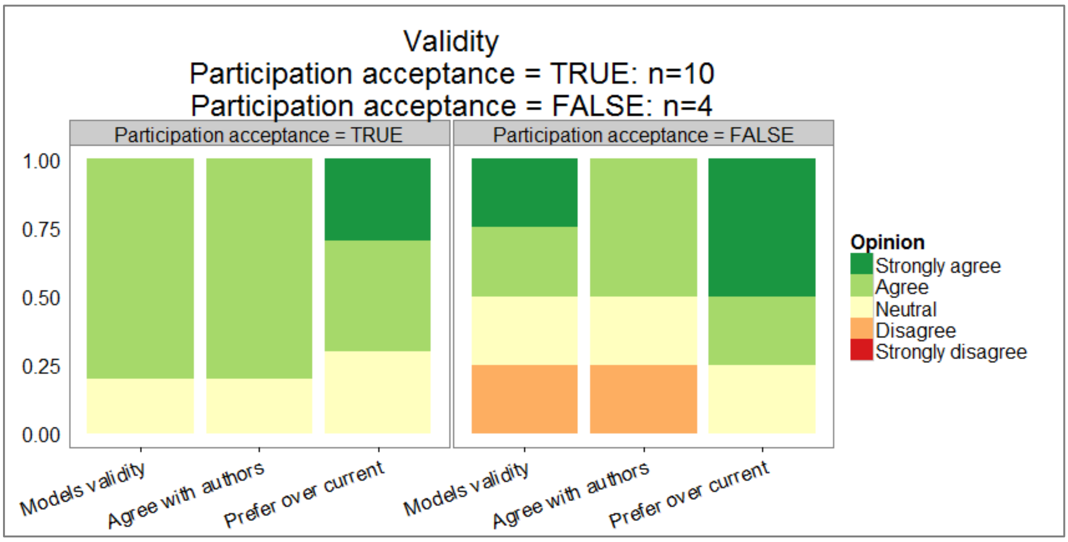


Figure 30 – Validity results, broken down for participation in the acceptance tests

We believe this has two causes. (1) Executives mostly did not participate in the acceptance test, and executives score positively (see sub-sub-section 6.3.4.1), and (2) persons that have a negative attitude did not want to invest time in the acceptance test. We understand that decision, but would like to have seen how such a negative attitude affects acceptance scores.

6.4 Acceptance

We also tested whether persons would accept using the scorecard needed to perform the assessment.

6.4.1 General

Without an expectation, persons easily understood the structure of the scorecard (see sub-section 5.5.3). However, all persons needed to get used to the questions. They asked quite some clarification, such as “what if there is a data asset overview in department X, but not at Y” and “what if person Z knows that she is responsible for maintaining the data asset overview, but it is not listed in his function profile?” After finishing questions for the first domain, most people got familiar with the method and the next domains went much smoother.

These qualitative results are reflected in the results of the acceptance questionnaire (Figure 31).

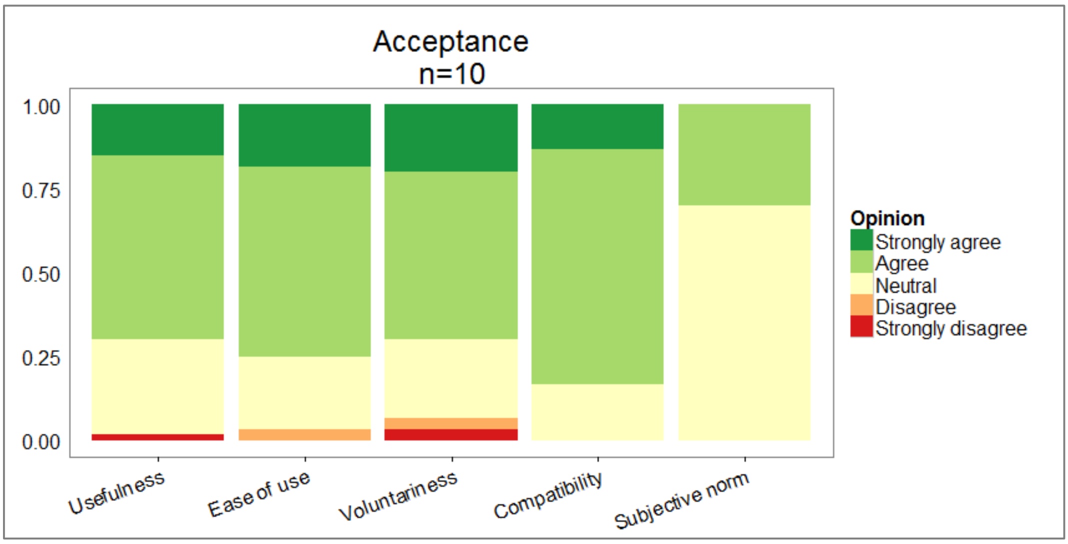


Figure 31 – Results of the acceptance questionnaire

Of 10 participants, 70% believes that the scorecard is useful and 75% finds the scorecard easy to use. Persons that rated these two items negatively find the scorecard and method valid but believe that it will make their job more difficult and that it requires significant mental effort.

70% of the persons believe they would use the method voluntarily, although they find it hard to imagine as there currently is nothing like the method in their organisation. Some persons wonder if their supervisor would enforce them to use the method. 83% of the persons believe the method fits nice within their current way of working.

Nearly all participants mentioned that subjective norm is not a good measure for acceptance, since data governance maturity is new to them and they cannot image what their peers would think. This is reflected in a high number of neutral scores (53%). We consider the subjective norm not applicable and removed them from further acceptance results.

People are generally willing to accept the method, as a whole 74%[16] would accept using the method.

6.4.2 Function group breakdown

Results that are broken down per function group are illustrated in Figure 32. For executives, the number of participants is too low (*n=1*) to include them in the function breakdown results.

[16] This is excluding subjective norm. Including subjective norm, the result would have been 68%.

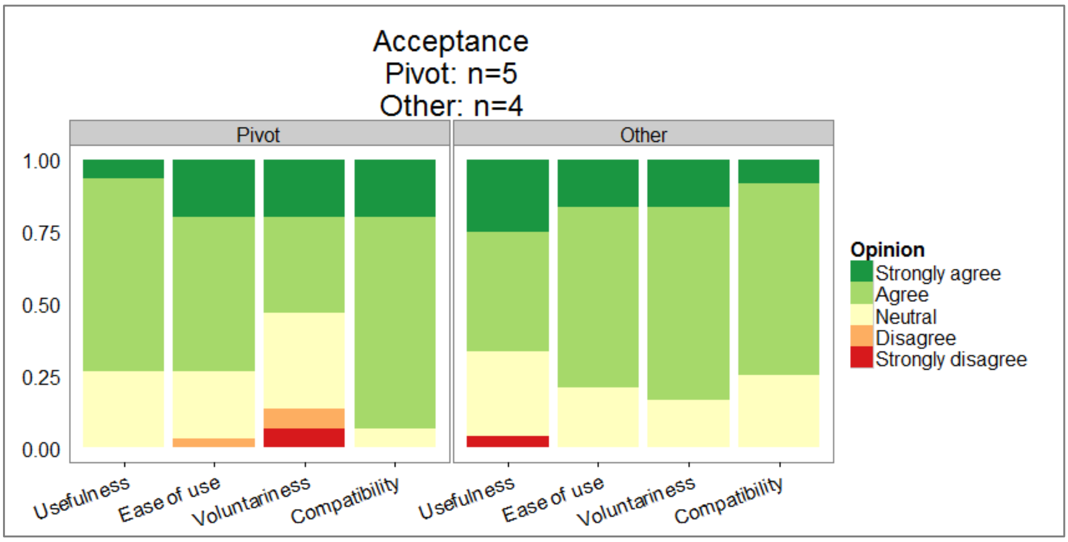


Figure 32 – Acceptance results, broken down for the function groups.
Executives are removed because n=1
Subjective norm is removed because it is not applicable

We noticed that, unlike expected, those in the group 'other' score slightly higher than pivots (70% vs. 66% over all questions). Since 'others' have significantly lower knowledge of data governance, they need to think more actively about the scorecard, which we believe is the reason for their slightly higher scores. Another factor is that pivots have high knowledge of data governance, hence easy spot flaws (also see sub-section 6.3.4).

These results indicate that although the method's scorecard is targeted at pivots, a broader range of functions can use it.

6.4.3 Organisation breakdown

As illustrated in Figure 33, acceptance results are broken down per organisation. Acceptance scores across all questions are 62% at Bank and 76% at ANWB.

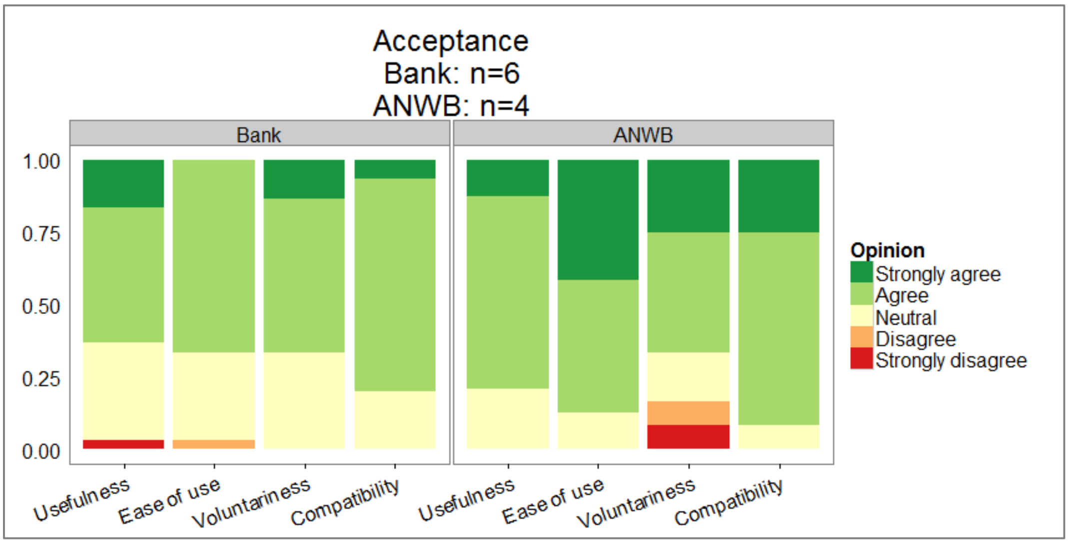


Figure 33 – Acceptance results, broken down for the organisations.
Subjective norm is removed because it is not applicable

Whereas validity is substantially higher at Bank than at ANWB (see sub-section 6.3.5), acceptance is substantially higher at ANWB. We expect this has two causes. (1) Persons that find the model less valid, did not participate in the acceptance tests. (2) ANWB is a larger and more diverse organisation, making the need for governance bigger (Weber 2009). This bigger need is reflected in higher acceptance scores.

6.5 Inter-rater reliability

As was explained in sub-sub-section 2.8.3.4, ten participants also used the method to assign a maturity score to their organisation. Inter-rater reliability (IRR) is measured using intra-class correlation (ICC) between rater pairs within the same organisation. The raters' scores are derived from the filled in scorecards of the acceptance tests. An IRR of *1* indicates a perfect correlation between the raters, *0* indicates no correlation. Since the scorecards are filled in for specific organisations, they cannot be combined.

Inter-rater reliability for the two case study organisations is reflected in Table 22Figure 39.

Organisation	*Raters (n)*	*Subjects*	*IRR*
Bank	6	73	0.39
ANWB	4	73	0.41

Table 22 – Inter-rater reliability at Bank and ANWB

Inter-rater reliability at Bank and ANWB is similar: 0.39 and 0.41, respectively. This means that raters tend to agree, however, there is variation. We contribute the variation to two causes: (1) Every rater is different. The level of knowledge and their familiarity with data governance will differ between raters; this likely also affects the answers they give. Besides that, raters may simply disagree. (2) We noticed that some questions could be answered in several ways.

Unfortunately, we cannot tell from the data we collected what cause has the highest influence on the IRR. Moreover, these results are not statistically significant and therefore only apply to this specific group. We aim to address these issues in future research (section 7.6).

6.5.1 Governance element breakdown

IRR differs substantially between the elements (Table 23). Especially on tasks, IRR is high (0.42 at Bank, 0.69 at Gamma). IRR is rather similar for objectives (0.29 at Bank, 0.24 at Gamma) and responsibilities (0.20 and 0.11), but differs a lot on tasks (0.43 vs. 0.69) and roles (0.07 vs. 0.43). Average IRR on governance elements is 0.30.

Organisation	*Element*	*Raters (n)*	*Subjects*	*IRR*
Bank	Objectives	6	18	0.29
Bank	Tasks	6	15	0.42
Bank	Roles	6	20	0.07
Bank	Responsibilities	6	20	0.20
ANWB	Objectives	4	18	0.24
ANWB	Tasks	4	15	0.69
ANWB	Roles	4	20	0.43
ANWB	Responsibilities	4	20	0.11

Table 23 – Inter-rater reliability at Bank and ANWB,
broken down for the governance elements

The IRR of 0.07 for roles at Bank is close to zero, which indicates no correlation. We have no explanation for this. Besides a high IRR for tasks at both Bank and Gamma, we see no clear trend. We have no explanation for the large difference in IRR between elements. We are confident that the future research to which we referred in the previous section will shed more light on the issue.

6.5.2 Data governance domain breakdown

IRR also differs substantially between the data governance domains (Table 24). The lowest IRR is 0.10 for data lifecycle at ANWB; the highest is 0.47 for data lifecycle at Bank. There is no clear domain with a substantial high or low IRR. Average agreement on the domains is 0.23, compared to 0.30 at the elements.

Organisation	*Domain*	*Raters (n)*	*Subjects*	*IRR*
Bank	Data assets	6	21	0.24
Bank	Data quality	6	13	0.28
Bank	Metadata	6	12	0.29
Bank	Data access	6	13	0.17
Bank	Data lifecycle	6	14	0.47
ANWB	Data assets	4	21	0.13
ANWB	Data quality	4	13	0.14
ANWB	Metadata	4	12	0.25
ANWB	Data access	4	13	0.23
ANWB	Data lifecycle	4	14	0.10

Table 24 – Inter-rater reliability at Bank and ANWB, broken down for the data governance domains

We contribute the high IRR (0.47) for data lifecycle at Bank to the nature of the organisation: due to regulatory requirements, data lifecycle is important to a bank and all employees were notably aware of that importance.

Case studies with a higher number of participants are needed to clarify the difference in IRR between the elements.

Results that we discussed in the previous sections were primarily collected to answer the research questions. We think the case studies yielded interesting results that do not directly relate to these questions; we will discuss them in the following sections.

6.6 Requirements of a data governance maturity method

In section 4.1 we stated ten requirements for a data governance maturity method. Having discussed what persons think of it, we can see if it meets these requirements (Table 28).

Requirement	*Summary on result*	*Reference section*	*Meeting the requirement*
1. Assess formal data governance programs	(No results)	N/A	Unknown
2. Assess ungoverned data practices	Validity results: 73%	6.3	Yes
3. Produce a maturity score	Design: maturity model	5.2	Yes
4. Multidimensionality	Design: maturity model	5.2	Yes
5. No more than five classes per factor	Design: maturity model	5.2	Yes
6. Interpret the maturity results	Design: interpretation instructions	5.5.4; 6.3.1	Partly
7. Actionable recommendations	Design: interpretation instructions	5.5.4; 6.3.1	Partly
8. Accepted by users	Acceptance results: 74%	6.4	Yes
9. Reliable results	Inter-rater reliability results: 0.4	6.5	Partly
10. Scientifically evaluated	Process: evaluation sessions	6	Yes

Table 25 – How the data governance maturity method meets the requirements

Unfortunately, we did not test if the model assesses the maturity of ***formal data governance programs*** (requirement 1). We should test this in future case studies (section 7.5).

At organisations with ***ungoverned data practices***, people agree that the method ***assesses maturity*** of these practices (requirement 2; 73% overall validity; *n=15*).

Certain requirements are met by the method's very design. Once the scorecard is filled in, it ***produces maturity scores*** (requirement 3). Given the several dimensions and classes and their arrangement in a matrix, the method's maturity model captures data governance's ***multidimensionality*** (requirement 4). The model has ***no more than five classes per dimension*** (requirement 5).

Also by design of the method, there are means to ***interpret the results*** (requirement 6) to derive ***actionable recommendations*** (requirement 7) for improvement. Although most participants remarked that they think our interpretation and recommendations are correct, some believe they are insufficient (see sub-section 6.3.1). We conclude that we only partly met these two requirements and that we should address these issues in future research (section 7.5).

Assessors ***accept*** the method: 74% would use it (*n=10*; requirement 8).

The inter-rater reliability is 0.4 (*n=10*), indicating that the results are moderately ***reliable*** (requirement 9). We only partly met this requirement and should address reliability in future research (section 7.5).

Finally, using the evaluation sessions we ***scientifically evaluated*** the method, thereby meeting requirement 10.

6.7 Governance elements

As explained in section 6.1, qualitative results about data governance maturity at Software Improvement Group are gathered using a previous version of the method but are mostly applicable. These results are included in the following sections.

Results about the governance elements are based on the three case studies. Due to confidentiality, the exact case study results are only available as confidential attachments, however, we can indicate how persons responded to specific areas, how certain areas correlate, and what we believe are trend lines.

We should keep in mind that our results are only based on three organisations that have similar levels of data governance maturity (we will elaborate more on this in section 7.5), hence should not generalise the results.

6.7.1 Objectives

We noticed that all three organisations mostly know what they want to achieve with data, but hardly formalised these objectives. We believe that this lack of direction causes several problems in data practices, such as employees that are not treating data with care since they do not know it is important.

6.7.2 Tasks

Although specific data governance roles are frequently not present, there are enough people that *feel* responsible and do execute the tasks. In other words: strong stewardship for data is present at all organisations. In next versions of the method, we want to elaborate more on responsibility that people feel, rather than they officially have.

6.7.3 Roles

Data council

Software Improvement Group, Bank nor ANWB has a data council. Software Improvement Group and ANWB do not desire one; Bank is setting one up. We agree with this: given Software Improvement Group's limited size a data council would be overkill; ANWB has aversion against large governance mechanisms such as councils; Bank indeed needs a data council because a large number of decisions concerning data have to be made due to the data lake project.

These results replicate that, unlike what is promoted by most researcher and practitioners, a data council is not obligatory for decision-making around data (Otto 2011).

Data steward

Also data stewards are not present at the three organisations, but all three are considering it. For Software Improvement Group and ANWB, this is due participation in this study; Bank already had plans. We believe data stewards are beneficial for all.

Agile product owner

A role that is not included in the maturity model but mentioned frequently during the case studies is the product owner (PO) of an agile project. All three organisations embraced agile methodologies; Bank and ANWB did so to bring the business and IT closer together. This is also an important goal of data governance.

In agile development, the PO is the project's key stakeholder. He or she is the decision-maker regarding what functionality a product will have, and functionality typically contains data. Hence, PO's naturally have some level of responsibility over data. We experienced that most POs are not aware of this. This may be the reason for our observation that POs often do not instruct teams to update the data model based on the delivered functionality. The data model is considered important for turning data into information (Whitten & Bentley 2005).

Agile teams in general, and the product owner specifically, clearly play some role in data governance. However, these did not occur during our literature study. The relationship between agile and data governance should be further explored.

6.7.4 Responsibilities

At Software Improvement Group, Bank and ANWB, often any level of responsibility for data quality, metadata and data lifecycle is derived from responsibility for data assets. Responsibility for data assets in turn is often derived from responsibility for applications and business processes.

The results from the questionnaire indicate that about 50% of the persons that have any level of responsibility over data assets are insufficiently aware of this. The same percentage goes for ownership that is set to departments rather than roles or persons, and ownership that cannot be traced to the highest management. All this results in confusion and disagreement about responsibility, affecting the efficiency with which data is processes and the effectiveness with which it can be used.

6.8 Data governance domains

Just like the governance elements results, results for data governance domains are based on the three case studies we conducted.

6.8.1 Data assets

We noticed that the enterprise architecture practices we included – the organisation's data assets and their connection to information requirements, business processes, and applications – are useful communication tools. Especially after discussing the organisations primary data assets, participants began to think about the importance of data to the organisation and ways to properly govern it. It looks like data assets give organisations the required handles to comprehend data governance.

This method's focus on data asset caused one downside: whereas Software Improvement Group had a data asset overview, Bank and ANWB did not, which made the process to rate these organisations on this domain more difficult.

6.8.2 Data quality & metadata

We noticed a strong connection between the domains data quality and metadata. Although the questions for both classes are distinct, final scores are remarkably similar. Moreover, at all three organisations these are the classes with the lowest scores. Examples of problems that arose at the case study organisations are inconsistent client data between different systems (data quality problem) and difficulty in retrieving data, due to a missing company-wide definition of the business concept 'client' (metadata problem).

We believe that data quality and metadata are the most difficult classes of the model and organisations generally do not sufficiently understand them. We expect this is caused by the nature of the two classes: data quality inherently is multidimensional (see sub-sub-section 5.2.2.2) and metadata has a broad scope (see sub-sub-section 5.2.2.3).

6.8.3 Data access

On data access, all three organisations have their highest score. We contribute this to most organisations including security in their corporate governance, for instance with a chief security officer that also covers information security.

Each participating organisation has dedicated governance elements for the data access domain, such as policies whom should have access to what information and specific data access stewards (Figure 34). This differs for tasks, roles and responsibility for data quality, metadata and to a lesser extent for data lifecycle, which are usually derived from, and integrated with, guidelines for data assets (see sub-section 6.8.1),

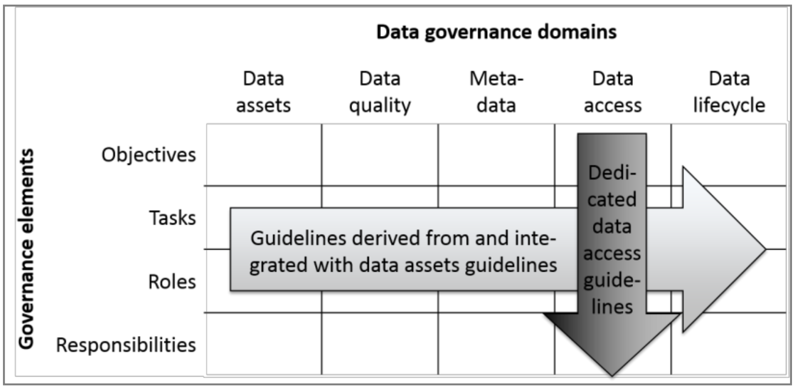


Figure 34 – Participating organisations have dedicated data access guidelines, whereas most other guidelines are derived from data assets guidelines

6.8.4 Data lifecycle

The data lifecycle domain is the most technical domain and therefore also has the strongest connections to IT. Data lifecycle tasks are mostly executed by database administrators and on the operational level. However, these tasks are mostly part of regular IT operations rather than explicitly part of data governance.

6.9 Data governance at SIG, Bank and ANWB

Although Software Improvement Group, Bank and ANWB are very different organisations, their IT aspirations with, and governance over, data are remarkably similar. Software Improvement Group, Bank and ANWB all identified the integration of several data sources as one of their primary IT targets. All develop using agile methodologies and none have a formal data governance program.

Little specific roles involved with data are defined, but tasks are executed due to a strong sense of data stewardship. Ownership and responsibility for data at all three organisations implicitly is derived from application and business process ownership, and all now consider to make data ownership and responsibility more explicit.

At all three organisations, data objectives are known but not stated explicitly. All score high on data access and low on metadata and data quality. Although none of the data governance programs is formal, this is never causing problems in day-to-day operations. However, the lack of a formal data governance program is likely causing problems for future data aspirations.

The similarity in current data governance maturity at Software Improvement Group, Bank and ANWB, and considerations they make to improve, indicate that data governance may be more generic than found by for instance Weber et al. (2009). This is more in line with Cheong and Chang (2007), whose findings suggest that basic structures for data governance apply. We do not promote that data governance at every organisation should look the same. Rather, the core of data governance practices may be similar across most organisations. Based on that core, organisation-specific data governance should be built.

6.10 Data governance & organisational change

In sub-section 3.2.3, we defined data governance as "[…] the set of processes […] that serve two purposes: specifying an organisation's data objectives, and specifying a decision-making framework that is appropriate for meeting these objectives." This definition containing the specification of data objectives and a decision-making framework gives data governance an explicit nature.

During the three case studies, we noticed that most person that are unaware of their explicit or implicit responsibility over data are willing to take responsibility, but lack the knowledge to actually do so. Participants remarked that achieving a high maturity level is not possible without an organisational culture where persons take responsibility for data, regardless of a formal responsibility. This is also reflected in scientific literature (English 2006; Lucas 2010).

Data governance is not only involved with a set of formal processes, but also with implicit responsibilities and organisational change. We considered these in section 3.5.3, but implicit responsibilities and organisational change were absent in our definition and therefore insufficiently covered in our maturity method. Organisations noted that they would like to receive more specific advice on these issues.

Although the method's validity is rated high, we should look at how to incorporate organisational change and implicit responsibilities into it, for instance by giving specific guidelines for training programs.

6.11 Data governance organisation

Most researchers scope data governance as governance about data and information, serving the business (see section 3.7). Although this may be data governance's *scope*, we found that the data governance *organisation* spans all five enterprise architecture layers (Figure 35).

For instance, a function such as business analyst may be responsible for collecting *information* requirements from the *business* and connecting these to *data*. A database administrator may gather this data from several *data* sources using several *applications* and processes these into *information*. Finally, a network administrator may need to alter the *infrastructure* in order to deal with specific requirements, for instance for archiving of *data*.

We conclude that functions involved with data governance go beyond the business, information and data layer, and covers all of the five organisational layers. This knowledge can make it easier to place data governance within an organisation.

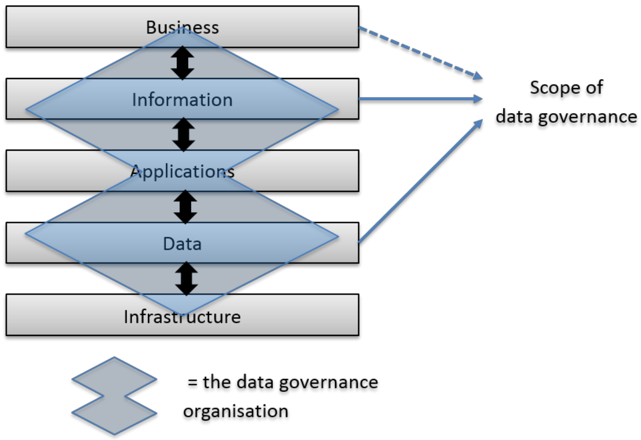


*Figure 35 – The data governance organisation,
illustrated using the five enterprise architecture layers*

6.12 Data for innovation

In section 3.1, we discussed five reasons for organisations to treat data as an organisational asset, hence to implement a data governance program (Otto 2011). We found that his list should be extended with 'enabling innovation'.

All reasons mentioned by Otto are about improving the current situation, for instance increasing the satisfaction of the customers that an organisation has, and to increase operational efficiency. Sometimes, however, a drastic disruption of the current situation is desired to break out of existing patterns. Such drastic disruptions are called innovations and can be triggered by data (Schumpeter 1934; Sathi 2012).

Especially at ANWB and to a lesser extent at Bank and Software Improvement Group, data will be used to drastically change the as-is situation. ANWB wants to use data to completely reinvent itself, rather than improving the current services. We believe that innovation can be an enabler for data governance.

6.13 Research method results

In this section, we will briefly reflect on the research method.

6.13.1 Case study

The case study was executed as planned and we believe it was a useful way to test the method and to gather data for the evaluation session. Good questions to get people to talk about the subject are about how happy they are with the current situation, and how they would like to improve it. Within organisations, not all persons agreed or had the same level of knowledge, which is why it is important to ask the same question to multiple people (control questions).

Also, as indicated in section 6.3.5, at ANWB it was difficult to gather enough information during the seven interviews. This resulted in a lower validity score. In such an organisation, one should either focus on a part of the organisation and/or plan more interviews.

Our focus on persons that have a pivot function between business and IT resulted in much valuable feedback, but we believe we did not give sufficient attention to persons either from the business or IT. How they value the method should be explored in further research.

6.13.2 Evaluation sessions

We believe the evaluation sessions we held are good ways to gather feedback. Most people read the report we sent in advance and everybody was very willing to comment it. However, one person decided after reading the report that he did not want to participate in the acceptance part of the evaluation session, since he felt uncomfortable with it. He explained that he was too much on the operational side of the organisation to have sufficient information to judge any governance. Although we agree that he may not have sufficient information, we would have liked him to have participated to test how such insufficient knowledge affects his acceptance and reliability scores.

We should keep in mind that the maker of the method also conducted the evaluation sessions, and that people their natural tendency is to provide socially-acceptable answers. Although the persons would have been more negative without the authors being present, we believe the results would not have been drastically different.

6.13.3 Report

The report with the provided outline was a valuable tool to explain the method, elaborate on the results and to provide actionable recommendations. However, for some the report was too conceptual and therefore difficult to understand. This was especially the case at persons lower in the organisation.

Some persons mentioned that they would also like our advice on who should perform the given recommendations.

6.13.4 Improvement workshops

Software Improvement Group, Bank and ANWB were happy with the results and wanted to improve their data governance maturity, using inter alia an improvement workshop. The workshops at Software Improvement Group and Bank had taken place; it is scheduled at ANWB. All participants at Bank and most participants at Software Improvement Group found the workshop useful and specific actions are determined to create awareness of data governance within the organisations.

7 CONCLUSIONS

In the previous chapters we illustrated what data governance is, listed requirements for a data governance maturity method, developed a method that meets these requirements and illustrated what possible end users think of the developed method. We conclude the research by answering the research questions and by suggesting future work.

7.1 Research question 1 – Data governance

The first research question is formulated as "*How is data governance placed in the context of data management, information, data quality, enterprise architecture, IT governance and corporate governance?*" The answer to the question is given in this section.

Data governance

Data as an organisational asset (section 3.1) – A high percentage of organisations believe that data can be a valuable organisational asset. The reason is that data can be processed into information, which is needed to ensure compliance, enable better decision-making, improve customer satisfaction, increase operational efficiency, support business integration and enable innovation.

Society legitimisation (sub-section 1.2.2) – Organisations find implementing a program difficult. 34% of the respondents in a large survey note that they do not know where to start, indicating that there is a strong need for maturity assessment methods. 43% of the organisations considers seeking assistance from an outside organisation, or have already done so.

Data governance (section 3.2) – Organisational assets require some form of governance, which is why organisations are adopting data governance programs. Data governance is the set of processes within an organisation that serve two purposes: specifying data objectives, and specifying a decision-making framework that is appropriate for meeting these objectives. The framework consists of tasks, roles and responsibilities that all play on the domains data assets, data quality, metadata, data access and data lifecycle.

Organisational placement (section 3.2) – Organisations need to make three decisions about placing main responsibility for a data governance program. (1) The first is about how responsibility is divided between an IT department and the business. A common way is to make IT responsible for retrieving and securely storing data, whereas the business is responsible for the data's content. A close collaboration of the business and IT is one of the core aspects of data governance. (2) Organisations should also decide how high in the organisation data governance responsibility should be placed. No clear trend about this can be found in literature. (3) Lastly, organisations should decide on placing data governance centralised or decentralised. The best way tends to be a federalist approach: only important strategic decisions are made by the highest levels of the organisation and departments operate largely independently.

Data management & data governance (section 3.3)

Data governance deals with the future state of the data in an organisation and the road on how to get there, whereas data management focuses on managing data operations.

Information & data governance (section 3.4)

As mentioned in the first paragraph of this section, data can be processed into information. Organisations require information, which is non-technical in nature, but it needs to be derived from technical data sources. Organisations are treating *data* as an enterprise asset since that is the way to get the *information* they desire. We consider both data and information in the scope of data governance, hence do not distinguish between data governance and information governance.

Data quality & data governance (section 3.5)

To unlock the potential of information that is hidden in data, the data needs to be of sufficient quality. Ultimately, one may say that achieving high quality data is the primary reason for a data governance program. Methods to improve data quality include establishing a corporate culture where every employee is willing to be accountable for a set of business information, called data stewardship, and assigning specific data stewards that are responsible for managing the quality of a set of data on a daily basis.

Enterprise architecture & data governance (section 3.6)

According to enterprise architecture principles, the IT organisation consists of five stacked layers: business, information, applications, data, and infrastructure. Knowing an organisation's data assets, how they are used in business processes, and from what applications they are derived, is critical in understanding how data is used in organisations and who should be responsible. Remarkably, these advantages are hardly present in popular scientific data governance literature

IT governance, corporate governance & data governance (section 3.7)

Data governance is a relatively new field of which much information is derived from IT governance. Both deal with assets in the five enterprise architecture layers, however, data governance deals with data assets whereas IT governance deals with IT assets such as applications, desktops and switches. IT and data governance are interconnected but independent disciplines.

The scope of IT and data governance is limited to specific types of assets, but corporate governance deals with all assets in an organisation. Both IT and data governance are generally considered subsets of corporate governance.

7.2 Research question 2 – Requirements to assess data governance maturity

We formulated the second question as *"What are requirements for a method to assess data governance maturity, and do existing data governance models meet these requirements?"* We will now elaborate on the question.

Requirements (section 4.2)

The method should assess the maturity of formal data governance programs and of ungoverned data practices. It should have some kind of metric, a means to interpret the results and it should be possible to derive actionable recommendations. To capture the full complexity of data governance, it should be multidimensional. For comprehensibility, each dimension should be decomposed into a maximum of five units. For users, the method should be easy to use and objective. Moreover, the model should be scientifically evaluated.

Data governance models (section 4.3)

We explored eleven models, both scientific and commercial, that consider the multidimensionality of data governance and decompose it into a maximum of five units, but do not provide a means to measure data governance maturity. There are maturity models for data governance specifically. However, since they are based on the Capability Maturity Model Integrations, they are not multidimensional. Also, they are not scientifically evaluated and objective nor reliable. The academic corporate data quality management self-assessment (Hüner et al. 2009) comes closest to meeting our criteria. However, it measures data quality management maturity, which is close to data governance but not the same. The model is unreliable and it is hard to determine actionable improvements.

Comparison with the requirements (section 4.4)

None of the models we explored meets all of our requirements. The next research question discusses the method we developed and how it meets most of these requirements.

7.3 Research question 3 – Data governance maturity method

We formulated the final research question as *"Can a new method help intended end users assess their data governance maturity, and how do they value that new method?"* To answer it, we will first discuss the data governance maturity method, after which we summarise what intended end users think of it. We conclude with checking if the method meets our requirements.

Method (chapter 5)

We developed a method to assess data governance maturity. It consists of a data governance maturity model, and instructions for how to use the model.

The maturity model describes the elements of which data governance maturity consists, how they relate to each other and how the scores are computed. The model is created using Factor-Criteria-Metric and consists of two dimensions: governance elements and data governance domains. The dimension 'governance elements' is divided into the classes objectives, tasks, roles and responsibilities; 'data governance domains' is divided into data assets, data quality, metadata, data access and data lifecycle. The dimensions and related classes are arranged in a two-dimensional matrix where the cells contain between two and seven questions. Each question can be answered on a five-point Likert scale. The score per cell is the division of the summed scores in that cell by the maximum score per cell (section 5.1 to 5.4).

The method goes beyond a model describing data governance maturity. It provides instructions for interpreting maturity results to derive actionable recommendations. Moreover, the method advises on the process for data collection, execution, validation, decision-making, diffusion, and the assignment of roles (section 5.5).

Construct validity (section 6.3)

We conducted two case studies at a large and a small IT-intensive organisation, of which both do not have a formal data governance program. Hence, we used the method to assess the organisations' ungoverned data practices. 73% of the participants (n=15) in the evaluation sessions find the method valid; within the group of executives (n=3) this number rises to 100%. Employees of the two case study organisations valued most the structured explanation of the current state and the actionable recommendations.

Our decision to position data assets as the core of the method greatly helped organisations in understanding the concept of data governance.

Our assessment is regarded as more correct at the first organisation than at the second (77% vs. 70%). The reason is likely the larger and more diverse nature of the second organisation.

Even persons that are not familiar with the method's terminology (5=6) understood the essence of the method within half an hour and rated it as valid (67%).

Participants found, however, that the method currently provides insufficient support to interpret the results. Therefore, in future work we aim to structure the interpretation step with an interpretation framework that connects organisational targets to maturity scores.

Acceptance (section 6.4)

74% of the participants in the evaluation sessions (n=10) would accept using the method. This is the case both for persons that are experienced with governance, and persons that do not have such experience.

Acceptance is the highest at the second organisation (76%, compared 62% at the first organisation). We expect that persons in a large and diverse organisation such as the second organisation find the need for a data governance program higher, resulting in higher acceptance scores for the method.

Inter-rater reliability (section 6.5)

Although there is variation between raters, they tend to agree. This is reflected in an inter-rater reliability score of 0.4 (n=10). We contribute the variation to two causes: (1) every rater is different. The level of knowledge and their familiarity with data governance will differ between raters, this likely also affects the answers they give. Besides that, raters may simply disagree. (2) Some of the method's questions are ambiguous.

Requirements of a data governance maturity method (section 6.6)

Having discussed what persons think of the data governance maturity method we developed, we can see if it meets the ten requirements of such a method that we stated in section 7.2.

The method meets most of the requirements we listed. We did not test if the model assesses the maturity of formal data governance programs. At organisations with ungoverned data practices, people agree that the method assesses maturity of these practices (73%; n=15). Once the scorecard is filled in, it produces maturity scores. Given the several dimensions and classes and their arrangement in a matrix, the method's maturity model captures data governance's multidimensionality. The model has no more than five classes per dimension. There are means to interpret the results to derive actionable recommendations for improvement. Although most participants remarked that they think our interpretation and recommendations are correct, some believe they are insufficient. We conclude that we only partly met these two requirements. Assessors accept the method: 74% would use it (n=10). The inter-rater reliability is 0.4 (n=10), indicating that the results are moderately reliable. We only partly met this requirement. Finally, we scientifically evaluated the method.

7.4 Other conclusions

We found interesting results that are not specific to the research questions, which will be summarised in this section. As explained in section 6.1, results from the case study at Software Improvement Group are included.

Governance elements and data governance domains (section 6.7 and 6.8)

Although most organisations know what they want to achieve with data, this is hardly made explicit. This lack of direction causes several problems in data practices. The results from the questionnaire at three organisations indicate that about 50% of the persons that have any level of responsibility over data assets are insufficiently aware of this. Whereas many data governance roles are not made explicit, tasks are being executed because people feel responsible. The role of agile product owner was not found in data governance literature, but was encountered frequently during the case studies.

Data quality and metadata are the two domains that score the lowest at all three organisations. Moreover, they closely correlate. We believe this is because these two domains are complex, resulting in that organisations do not understand them sufficiently. The data lifecycle domain usually plays on the level of IT operations and is considered to be the most technical. On the data access domain, scores are the highest at all organisations.

Data governance at Software Improvement Group, Bank and ANWB (section 6.9)

Data governance at Software Improvement Group, Bank and ANWB is remarkably similar, indicating that the core of data governance practices may be similar across most organisations. Based on that core, organisation-specific data governance should be built.

Data governance and organisational change (section 6.10)

Data governance is not only involved with a set of formal processes, but also with implicit responsibilities and organisational change. These are insufficiently covered in our maturity method. Although the method's validity is rated high, we should look at how to incorporate organisational change and implicit responsibilities into it, for instance by giving specific guidelines for training programs.

The data governance organisation (section 6.11)

We conclude that functions involved with data governance go beyond the business, information and data layer, and covers all of the five organisational layers. This knowledge can make it easier to place data governance within an organisation.

Data for innovation (section 6.12)

We also identified a sixth reason to perform data governance: enabling innovation, focusing on radical rather than incremental change.

Research method summary (chapter 2)

We first studied data governance literature and talked to Software Improvement Group consultants, after which we drafted a first version of the method. We used it in a pre-case study at Software Improvement Group. Although Software Improvement Group is no representative organisation since the employees are highly-educated and familiar with the value of data and governance, the case study helped us in understanding implications of the method in organisations. The results were used to improve the method, after which it was both qualitatively and quantitatively tested at Bank and ANWB. At three case study organisations, the results are actively diffused through reports, workshops, and presentations.

Research method conclusion (section 6.13)

We found the case studies, including the evaluation sessions, valuable in testing the method. Feedback on the report with the given outline, suggests that it is a valuable tool to explain the method, give interpretation to the results and provide actionable recommendations. Everyone that participated in the improvement workshop at Bank ($n=9$) valued the workshop as useful.

7.5 Discussion

The results of the maturity method are encouraging, but we should keep some limitations in mind.

Generalisability of a case study

Qualitative research using cases studies is aimed at understanding a specific attribute, and it focuses on a small number of cases. This creates a bias, for instance, all our organisations do not have a formal data governance program, show similarities in their data practices, and are Dutch. Since the number of participants is only ten or fifteen, depending on the test, the quantitative results we gathered are not statistically significant. One should therefore not generalise these case study results until the method has been applied at substantially more organisations (Lucas 2010; Cheong & Chang 2007).

More in-depth case studies at divergent organisations are needed, for instance at organisations that have a formal data governance program. To test our validity, acceptance and reliability results, also quantitative research using many organisations and a high number of participants is needed. These can be used to create benchmarks, for instance per industry. Benchmarks allow for easy comparison between organisations.

Interpretation framework

Participants found that the method provides insufficient support to interpret the results. We also experienced difficulty in advising on improvements.

In section 5.5, we emphasised the difference between the objective, descriptive model and the subjective, prescriptive interpretation of the model. In future work, we aim to structure the interpretation step by developing an interpretation framework.

Using this framework, we suggest that organisations should state what its most important reasons are to govern data, using the six motivations for a data governance program[17]. For each motivation, an appropriate score on each of the model's classes should be determined. The maturity scores can then directly be connected to organisational goals, which makes it easier to identify what improvements are needed.

Inter-rater reliability

In section 6.5, we concluded that although raters tend to agree, there is substantial variation. We also discussed that we cannot tell what the primary causes is for the relatively low inter-rater reliability: natural differences between raters, or ambiguity in some questions.

We aim to address these issues in four directions for future research. (1) First, inter-rater reliability tests should be conducted with a larger number of people, say forty. (2) We also want to see what the influence of natural differences between raters is, for instance by handing different raters the same factual information and see how they interpret it. This includes testing if IRR differs between the function groups we identified in section 6.1. (3) Thirdly, we should re-examine the questions that turned out to be ambiguous. (4) Lastly, we are curious if the construct validity, acceptance and inter-rater reliability results differ when organisations perform a self-assessment to rate themselves.

Weights

The weights in the maturity model with which we calculated the score per class and the total score, are based on our literature study (sub-sub-section 5.3.2.5). Different weights may result in maturity scores that better reflect reality, hence it should be investigated what suitable weights are. As stated in sub-section 6.3.1, we think weights should not be company-specific as this makes it harder to benchmark organisations.

Focus on the intersection of IT and business

Since data governance plays on the intersection of business and IT, in our case studies and evaluation sessions we focused on functions operating in this intersection (sub-section 6.13.1). We spoke to a relatively low number of people from *either* the business or IT and need to explore more how they value the method.

Measurement scale

In sub-sub-section 5.3.2.3, we discussed that the questions in the scorecard are answered on an ordinal scale. We want to study if other ordinal scales to answer the questions are better, for instance ordinal scales without a mid-point a ten-item scale (Garland 1991; Preston & Colman 2000).

To increase the statistics possibilities, we defined that the ordinal items linearly map to natural numbers with an absolute zero, making it a ratio scale (see sub-sub-section 5.3.2.3). We should keep in mind that our results on the ratio scale are derived from a scale where the difference between two items may not be the same.

Finally, we want to explore what the influence of exponential mappings, where the difference between two steps increases as the score gets higher, is on the method's distinctness (see sub-sub-section 6.3.6).

Organisational change

We choose to scope the research by using a definition that neglects organisational change and implicit responsibilities in data governance. Organisations indicated that they would like to have more specific advice on this issue (section 6.10). We should try to incorporate that in the method, for instance by giving specific guidelines for training programs.

[17] These are the five motivations mentioned in section 3.1, complemented with 'enabling innovation' (section 6.12).

Besides, we are curious in ways to measure implicit responsibilities, awareness of that, and if these scores correlate with data governance maturity scores.

Evaluation sessions

One of the method's authors conducted the evaluation sessions, and persons' natural tendency is to give socially acceptable answers. This creates a bias, although we do not think that results with a different person will drastically differ.

Documents

The assessment we designed is based on interviews, which makes the assessor dependent on what the interviewees tell him or her. This creates a certain bias. We suggest to complement the interviews with the studying documents such as existing guidelines, a data assets overview and function profiles.

Data governance knowledge

Much of the data governance knowledge we used is extracted from IT governance and has not been researched properly (section 1.2.1). More research on data governance is needed to validate items that were used in the method.

7.6 Future research

Besides the suggestions for further research that were based on the limitations in the previous section, we have the following suggestions.

Data governance and enterprise architecture

We found that the enterprise architecture practices we used strongly help in understanding data governance (sub-section 6.8.1). To the best of our knowledge, no research has been conducted on the intersection between the two fields. We suggest to further explore the relation between enterprise architecture and data governance, for instance if organisations that have enterprise architecture programs also score high on data governance maturity.

Agile and data governance

We frequently encountered agile and the role of product owner (PO; sub-section 6.7.3). The PO is the decision-maker regarding what features a product will have, and since functionality typically contains data the PO plays some role in data governance. Moreover, we noticed that agile teams usually do not update the data model, whereas the data model is considered critical in turning data into information. Agile teams in general, and the product owner specifically, play a substantial role in data governance but did not occur during our literature study, hence should be further explored.

Data governance basis

Our results indicate that the core of data governance practices may be similar across most organisations. This is in line with (Cheong & Chang 2007), who found that basic structure for data governance apply, but partly contradicts results (Weber et al. 2009), promoting a contingency approach. More research is needed to explore this inconsistency.

Data governance maturity and firm performance

Research on IT governance shows a correlation between type of governance (for instance centralised, decentralised or federated) and firm performance (Weill & Ross 2004), and between governance maturity and firm performance (Luftman & Kempaiah 2007). We found no such research for data governance and are curious if the same holds.

The correlation between several domains and elements

In section 6.7 and 6.8, we saw that certain domains and elements score similarly. Due to time constrains and a limited number of participants, we did not look at correlation patterns in questions, domains and elements. Such patterns can help in understanding data governance and reducing the number of questions needed to assess data governance maturity. We aim to explore correlation patterns in questions, domains and elements in future research.

7.7 Final conclusion and considerations

To the best of our knowledge, this thesis presents the first scientifically evaluated method to measure data governance by assessing its maturity. The method provides a means to perform the assessment and to interpret the results. It was tested at two organisations that both valued it as valid and would use it to assess data governance maturity.

Governing data begins with knowing what data assets there are, and how these are used in the business processes. This means that an architecture of the data assets should be made. Moreover, these data assets should be prioritised according to confidentiality, integrity and availability. It looks like data assets give organisations the required handles to comprehend data governance.

By providing a means to measure data governance, we give organisations handles to improve their data governance and to more successfully treat data as an organisational asset. Moreover, it resulted in a large list of possible future research to increase maturity of data governance research.

8 REFERENCES

AIS, 2015. AIS submission guidelines. Available at: http://icis2015.aisnet.org/en/paper-submissions/paper-submission.

Van Aken, J.E., 2005. Management research as a design science: Articulating the research products of mode 2 knowledge production in management. *British Journal of Management*, pp.19–36.

Aldris, A. et al., 2013. Measuring the degree of service orientation in proprietary SOA systems. *Proceedings - 2013 IEEE 7th International Symposium on Service-Oriented System Engineering, SOSE 2013*, pp.233–244.

Basel Committee on Banking Supervision, 2006. International Convergence of Capital Measurement and Capital Standards. *Group*, II(June 2004), pp.1–26. Available at: http://www.bis.org/publ/bcbs128.pdf.

Baxter, P., Susan Jack & Jack, S., 2008. Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. *The Qualitative Report Volume*, 13(4), pp.544–559.

Beck, K. et al., 2001. Manifesto for Agile Software Development. *The Agile Alliance*, 2009(December 14), p.2006. Available at: http://agilemanifesto.org/.

Bellinger, G., Castro, D. & Mills, A., 2004. Data, Information, Knowledge, and Wisdom. *Gene*, pp.1–3. Available at: http://www.systems-thinking.org/dikw/dikw.htm.

De Bruin, T. et al., 2006. Understanding the Main Phases of Developing a Maturity Assessment Model. , pp.6–8.

Chen, W., 2010. Kalido Data Governance Maturity Model.

Cheong, L.K. & Chang, V., 2007. The Need for Data Governance : A Case Study. *ACIS 2007 Proceedings*, pp.999–1008.

CMMI Architecture Team, 2007. *Introduction to the Architecture of the CMMI ® Framework*,

Data Governance Institute, DGI Data Governance Components Framework.

DataFlux, 2010. The Data Governance Maturity Model: Establishing the People, Policies and Technology That Manage Enterprise Data. , 44(0), pp.0–10. Available at: http://www.sas.com/offices/NA/canada/lp/DIDQ/DataFlux.pdf.

Dataversity, 2013. The 2013 DATAVERSITY Snapshot Survey on Data Governance.

Davis, J.H., Schoorman, F.D. & Donaldson, L., 1997. Towards a stewardship theory of management.

Dawes, J., 2008. Do data characteristics change according to the number of scale points use? An experiment using 5 point 7 point and 10 point scales. *International Journal of Market Research*, 50(1), pp.1–20.

Dember, M., 2009. Case Study : Gaining Efficiencies as a Result of Implementing a Data Governance Program.

Demise, N., 2006. OECD principles of corporate governance. *Corporate Governance in Japan: From the Viewpoints of Management, Accounting, and the Market*, pp.109–117.

Denyer, D., Tranfield, D. & van Aken, J.E., 2008. Developing Design Propositions through Research Synthesis. *Organization Studies*, 29(3), pp.393–413.

Duval, E. et al., 2002. Metadata principles and practicalities. *D-Lib Magazine*, 8(4), pp.1–13.

Economist Intelligence Unit, 2008. The future of enterprise information governance.

English, L., 2006. Information Stewardship: Accountability for Information Quality. *Information Impact International, Inc*, pp.1–16. Available at: http://iaidq.org/publications/doc2/english-2006-11-information-stewardship-white-paper-v3.pdf.

Ernst & Young, 2011. Regulatory compliance: adapting to a pressurized environment. Available at: http://www.ey.com/Publication/vwLUAssets/InSights_June_2011_Regulatory-compliance_adapting-to-a-pressurized-environment_CJ0180/$FILE/InSights_June 2011_Regulatory-compliance_adapting-to-a-pressurized-environment_CJ0180.pdf.

GAO, 2001. A Practical Guide to Federal Enterprise Architecture. *Public Law*, 1(February 2001), p.112. Available at: http://www.citeulike.org/group/15536/article/9666776.

Garland, R., 1991. The mid-point on a rating scale: Is it desirable? *Marketing Bulletin*, 2, pp.66–70. Available at: http://marketing-bulletin.massey.ac.nz/V2/MB_V2_N3_Garland.pdf.

Gartner, 2007. "Dirty Data" is a Business Problem, Not an IT Problem.

Gladden, R.R., 2007. Project management: strategic design and implementation (5th). *Project Management Journal*, 38(1), p.124. Available at: http://0-search.ebscohost.com.brum.beds.ac.uk/login.aspx?direct=true&db=buh&AN=24844793&site=ehost-live&scope=site.

Global Data Excellence, 2011. Data Governance : A Business Value-Driven Approach A White Paper.

Hüner, K.M., Ofner, M. & Otto, B., 2009. Towards a Maturity Model for Corporate Data Quality Management. *Proceedings of the 2009 ACM symposium on Applied Computing - SAC '09*, p.231. Available at: http://portal.acm.org/citation.cfm?doid=1529282.1529334.

Informatica, 2006. Data Governance.

ISO, 2012. ISO / IEC 27002 — Information technology security techniques. , 2012.

ISO, 2008. Software engineering — Software product Quality Requirements and Evaluation (SQuaRE) — Data quality model (25012). , 2008.

IT Governance Institute, 2008. *Cobit 4.1*, Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:COBIT+4.1#4.

Kaplan, R. & Norton, D., 2004. *Focusing Your Organization on Strategy-with the Balanced Scorecard*, Available at: http://www.paca-online.org/cop/docs/Kaplan+Norton_Balanced_Scorecard_-_3_articles.pdf [Accessed September 30, 2014].

Khatri, V. & Brown, C.V.C., 2010. Designing data governance. *Communications of the ACM*, 53(1), p.148. Available at: http://dl.acm.org/citation.cfm?id=1629210 [Accessed September 22, 2014].

King's College London, Statistics advisory service. Available at: http://www.kcl.ac.uk/ioppn/depts/biostatistics/SAS/faqs9.aspx#a9_4.

Kitchenham, B., 2004. Procedures for performing systematic reviews. *Keele, UK, Keele University*, 33(TR/SE-0401), p.28. Available at: http://csnotes.upm.edu.my/kelasmaya/pgkm20910.nsf/0/715071a8011d4c2f482577a700386d3a/$FILE/10.1.1.122.3308[1].pdf\nhttp://tests-zingarelli.googlecode.com/svn-history/r336/trunk/2-Disciplinas/MetodPesquisa/kitchenham_2004.pdf.

Kooper, M.N., Maes, R. & Lindgreen, E.E.O.R., 2011. On the governance of information: Introducing a new concept of governance to support the management of information. *International Journal of Information Management*, 31(3), pp.195–200. Available at: http://linkinghub.elsevier.com/retrieve/pii/S0268401210000708 [Accessed July 15, 2014].

Kuechler, W. & Vaishnavi, V., 2008. On theory development in design science research: anatomy of a research project. *European Journal of Information Systems*, 17(5), pp.489–504. Available at: http://www.palgrave-journals.com/doifinder/10.1057/ejis.2008.40.

Logan, D., 2012. Hype Cycle for Enterprise Information Management, 2012. *Gartner Research Report ID Number: G00167472*, (July), pp.1–54.

Loshin, D., 2007. Data Governance for Master Data Management and Beyond. Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Data+Governance+for+Master+Data+Management+and+Beyond#4 [Accessed September 22, 2014].

Lucas, A., 2010. Corporate Data Quality Management From Theory to Practice. In *5e Confrence ibérique des systemes de gestion de l'information*.

Luftman, J. & Kempaiah, R., 2007. An update on Business-IT alignment: "A line" has been drawn. *MIS Quarterly Executive*, 6(3). Available at: http://misqe.org/ojs2/index.php/misqe/article/view/135 [Accessed September 22, 2014].

Maes, R., 1999. Reconsidering Information Management Through A Generic Framework. *PrimaVera Working Paper Series*, 20(2), pp.146–146. Available at: http://primavera.fee.uva.nl/PDFdocs/99-15.pdf.

Marinescu, R., 2005. Measurement and quality in object-oriented design. *21st IEEE International Conference on Software Maintenance (ICSM'05)*.

Miller, G.A., 1956. The Magical Number Seven. *Psychological Review*, 63, pp.81–97.

Österle, H. et al., 2010. Memorandum on design-oriented information systems research. *European Journal of Information Systems*, 20(1), pp.7–10. Available at: http://www.palgrave-journals.com/doifinder/10.1057/ejis.2010.55 [Accessed November 5, 2014].

Otto, B., 2011. a Morphology of the Organisation of Data Governance. *European Conference on Information Systems (ECIS)*, p.13.

Peterson, R.R., Callaghan, R.O. & Ribbers, P.M.A., 2000. Information technology governance by design.

Preston, C.C. & Colman, A.M., 2000. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta psychologica*, 104(1), pp.1–15.

PricewaterhouseCoopers, 2012. *Annual Information Security Survey*,

Rahm, E., 2000. Data Cleaning : Problems and Current Approaches. *Informatica*, pp.1–11.

Riemenschneider, C.K. et al., 2002. Explaining Software Developer Acceptance of Methodologies : A Comparison of Five Theoretical Models. , 28(12), pp.1135–1145.

Rouse, M., 2015. Data Lifecycle Management. , (Dlm), p.2015. Available at: http://rcs.columbia.edu/rdlm.

Russom, P., 2006. Taking data quality to the enterprise through data governance. *The Data Warehousing Institute, Seattle*, (March). Available at: http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Taking+Data+Quality+to+the+Enterprise+through+Data+Governance#0.

Salido, J. et al., 2010. A Guide to Data Governance for Privacy , Confidentiality , and Compliance. *Microsoft Trustworthy Computing*, Part(April), p.35. Available at: https://privacyassociation.org/media/pdf/knowledge_center/Guide_to_Data_Governance_Part5_Moving_to_Cloud_Computing_whitepaper.pdf.

Sathi, A., 2012. *Big data analytics: Disruptive technologies for changing the game*, Mc Press.

Schumpeter, J.A., 1934. *The Theory of Economic Development*, Available at: http://books.google.co.uk/books?id=-OZwWcOGeOwC&printsec=frontcover&dq=schumpeter+theory+economic+development&hl=en&ei=7Q4xTJjbEYegsQaUj6W2Ag&sa=X&oi=book_result&ct=result&resnum=1&ved=0CDAQ6AEwAA#v=onepage&q&f=true.

Sen, a., Ramamurthy, K. & Sinha, a. P., 2012. A Model of Data Warehousing Process Maturity. *IEEE Transactions on Software Engineering*, 38(2), pp.336–353.

Shrout, P.E. & Fleiss, J.L., 1979. Intraclass correlations: uses in assessing rater reliability. *Psychological bulletin*, 86(2), pp.420–428.

Smith, A.M., 2015. Data Governance Maturity - An Overview. , (10), pp.4–5.

Spewak, S. & Hill, S., 1993. Enterprise architecture planning: developing a blueprint for data, applications and technology. Available at: http://dl.acm.org/citation.cfm?id=138413 [Accessed January 5, 2015].

Tallon, P.P., Ramirez, R. V. & Short, J.E., 2013. The Information Artifact in IT Governance: Toward a Theory of Information Governance. *Journal of Management Information Systems*, 30(3), pp.141–178. Available at: http://mesharpe.metapress.com/openurl.asp?genre=article&id=doi:10.2753/MIS0742-1222300306 [Accessed September 6, 2014].

Wang, R.Y. et al., 1998. Manage Your Information as a Product. , p.15. Available at: http://web.b.ebscohost.com/ehost/detail/detail?sid=94782af3-8ac9-4884-a7b0-9840ac131058@sessionmgr110&vid=1&hid=112&bdata=JnNpdGU9ZWhvc3QtbGl2ZQ==#db=buh&AN=887820 [Accessed September 16, 2014].

Wang, R.Y. & Strong, D., 1996. Beyond accuracy: What data quality means to data consumers. *Journal of management information systems*, 12(4), pp.5–33. Available at: http://www.jstor.org/stable/40398176 [Accessed September 22, 2014].

Weber, K., 2009. Data Governance Reference Model.

Weber, K. et al., 2008. Organising Accountabilities for Data Quality Management-A Data Governance Case Study. *Data Warehousing*. Available at: http://www.alexandria.unisg.ch/export/DL/Robert_Winter/204595.pdf [Accessed September 22, 2014].

Weber, K., Otto, B. & Osterle, H., 2009. One Size Does Not Fit All — A Contingency Approach to Data Governance. , 1(1).

Weill, P. & Ross, J.W., 2004. *IT governance - How top performers manage IT decisions rights for superior results*,

Whitman, M.E. & Mattord, H.J., 2011. *Principles of Information Security*,

Whitten, J. & Bentley, L., 2005. *Systems Analysis and Design Methods*,