Master Thesis
for the study programme MSc Business Information Technology

DATA-DRIVEN IT: TACKLING IT CHALLENGES WITH DATA MANAGEMENT IN A FINANCIAL INSTITUTION

BAS HENDRIKSE

University of Twente
August 2019

Bas Hendrikse: *Data-Driven IT: Tackling IT Challenges with Data Management in a Financial Institution,* Master Thesis, Business Information Technology, August 2019

AUTHOR:

Bas Hendrikse	University of Twente
Study program:	MSc Business Information Technology
Specialisation:	Data Science and Business
Email:	b.hendrikse@alumnus.utwente.nl

SUPERVISORS:

Dr.ir. Hans Moonen	University of Twente
Faculty:	Behavioural, Management and Social Sciences
Email:	j.m.moonen@utwente.nl
Dr. Klaas Sikkel	University of Twente
Faculty:	Electrical Engineering, Mathematics and Computer Science
Email:	k.sikkel@utwente.nl

EXTERNAL SUPERVISOR:

Sven Oosterhoff	Fortran Bank

MANAGEMENT SUMMARY

Information Technology (IT) plays a major role in keeping up in the financial world, reliable data is needed in order to do so. However, it is not always possible to use the available data, and it may not be trustworthy enough for the intended purpose, as it is subject to quality and accessibility issues. It is still largely unexplored what the potential of Data Management would be to tackle such challenges in the IT organisation of a large financial institution in Northwestern Europe. In this study we identify the value of data for IT, we identify challenges with data, we discuss how Data Management could tackle these challenges and we explain how an existing Data Management framework could be adapted such that it is suited for IT.

In our case study with 16 expert interviews we explain that two main challenges arise: Unclear relations between IT landscape components, and quality issues with data. The first challenge can be tackled with a common data model within IT, while the second needs a typical Data Management approach to provide clear accountability, data source administration, distribution and data quality management.

We propose three IT supporting capabilities as focus areas, in addition to the use of an existing Data Management Framework. These capabilities focus on: eliminating the need for key Data Management processes with the use of automatic generation of data; raising understanding of manual data creation by showcasing of the usage of data; and tracking the added value of data and data management.

We present the following key findings:

- Being in control of data needs a shift in mindset;
- Standardisation is an important part of controlling IT data assets;
- Responsibility for data assets is the key to adoption;
- DevOps and CICD lead to more IT control, Data Management enables control of data;
- Traceability is the key to value creation within IT.

The findings are limited by the large influence of organisation-specific context, the limited results per functional area of the case study participants and the defined scope of IT. The findings of this thesis could also be applicable to other financial institutions and IT organisations in companies outside the financial services industry.

In the literature review we describe a foundation for a data-driven organisation with six capabilities; besides Data Management, the importance of Management, Skilled Personnel, Culture, Analytics and Infrastructure is indicated. We found that especially Data Management in combination with DevOps and CICD facilitate improvement in these areas.

ACKNOWLEDGMENTS

This thesis marks the end of my MSc Business Information Technology at the University of Twente. I would like to thank my organisational supervisor Sven Oosterhoff for the support he gave me throughout this final project. Thank you very much for thinking with me during each step of the research process, your critical feedback and motivating me during the time I was at the bank.

Hans Moonen and Klaas Sikkel, my graduation committee, have supported me with their ideas, critical feedback and their encouragement throughout this graduation project, thank you very much for that.

I am grateful for everyone who participated in the case study and helped me form this thesis. My thanks also go out to my colleagues at the bank, who embraced me as part of the team.

I especially want to thank my parents for always being there for me and for supporting me with every major decision I wanted to take during my student life. At last I wish to thank the rest of my family and friends for all their support throughout my student life.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACRONYMS

CICD Continuous Integration and Continuous Delivery

DAMA The Data Management Association

DM Data Management

DMF Data Management Framework

FDO Fortran Data Office

1 INTRODUCTION

The financial world is changing fast and is subject to rising competition from technology entrepreneurs who are entering the financial services market [41]. Customers want fast and reliable service anytime and anywhere, while regulators request more data with higher frequency and higher precision from traditional banks. These financial technology (Fintech) entrepreneurs are not burdened by regulators, legacy IT systems, branch networks or the need to protect existing businesses [53]. Traditional financial institutions are and want to keep up and so are transforming their organisation. Information technology plays a significant role in this transformation [6].

The financial institution Fortran Bank is working on its IT transformation. The institution has previously introduced Agile principles throughout the organisation, which opts to increase the agility of the organisation. The IT organisation, which is mainly responsible for the development, deployment and maintenance of software, takes the next step and is transitioning to the integration of IT Development and IT Operations teams to DevOps, which also needs a new approach to deal with strategic partnerships. This transformation is accelerated by Continuous Integration & Continuous Delivery, Public Cloud and the cleaning of legacy IT landscape.

'Fortran Bank' is a pseudonym for a large financial institution in Northwestern Europe who shall remain anonymous in the continuance of this thesis

Reliable data is an essential asset in this transformation. It is needed to comply with regulation, to improve operational excellence, to improve the customer experience, as well as to innovate. The data available is, however, not always possible to use or reliable enough for the intended purpose. This data is subject to quality and accessibility issues. Data quality is a critical issue that can reduce the likelihood that value will be created from data [58].

Data Management can be used as an approach to get control of data sources on organisation-wide level. Within Fortran, relatively new guidelines for Data Management have been determined for the whole bank by a central organ, but decentralised teams carry the responsibility to translate this into practice for their department. It has remained mostly uninvestigated what it means for their organisation to be *Data Driven*, what the issues with data within IT are and what the potential of Data Management could be for IT within Fortran to tackle these challenges.

This research aims at addressing this gap in knowledge and aims at setting up a Data Management roadmap for the IT organisation in the financial institution Fortran and provides a guiding framework which is suitable for IT.

1.1 THESIS STRUCTURE

- Chapter 2 explains the context and scope of the research.
- Chapter 3 covers the approach of the research.
- Chapter 4 discusses capabilities for large data driven organisations.
- Chapter 5 provides Data Management models present in literature, explains the relation with the model used by Fortran and consults literature on Data Management in IT.
- Chapter 6 explains the methods used for the case study.
- Chapter 7 presents the case study results and discusses the main challenges with data.
- Chapter 8 describes the role of Data Management to tackle these challenges, presents recommendations for the practical application within Fortran and presents a prioritised model with three added IT enabling capabilities.
- Chapter 9 provides feedback from validation participants based on what is discussed in Chapter 7 and 8.
- Chapter 10 reflects on a data-driven capabilities for Fortran IT, provides a discussion on the research questions, explains the key findings of the research and lists the key contributions.

2 BACKGROUND

Background information is needed to understand the context of this research. This chapter describes the organisational scope and serves as a reference for the rest of the thesis.

2.1 ABOUT FORTRAN

Fortran is a listed bank in Northwestern Europe with thousands of employees. The organisation consists out of 7 *business lines*, which are large umbrella departments for the different services the organisation offers. These business lines are composed of smaller *business units*, with their own specific functions. Figure 2.1 displays the bank's organisational structure and the business units of the IT & Operations business line.

Our research focuses on the IT organisation within the bank, which employs 6000 employees, spread over more than 450 teams. The IT department is intertwined with the rest of the organisation. Part of it supports the departments with applications and other IT services. Another part of the IT department is primarily responsible for the creation and maintenance of IT products, for customers as well as for stakeholders within the bank.

This research focuses on the last type of IT departments within Fortran, Figure 2.1 highlights the IT departments. Other departments in the IT & Operations business line are left out of scope; their responsibilities do not fit within our definition of IT.

2.2 ABOUT THE IT DEPARTMENT WITHIN FORTRAN

Fortran needs many software applications to serve their clients, as well as their internal stakeholders. Designated teams are responsible for delivering those software applications, and can be seen as a large 'IT organisation' within the organisation. This organisation is responsible for all stages involved in the value chain, from problem definition to deployment and maintenance. IT can be described as a combination of the business units CTO, CIO and CISO. The following sections describe those units within Fortran.

2.2.1 *IT Business Units*

CORPORATE TECHNOLOGY OFFICE (CTO) The Corporate Technology Office (CTO) provides all tools, procedures and processes to the

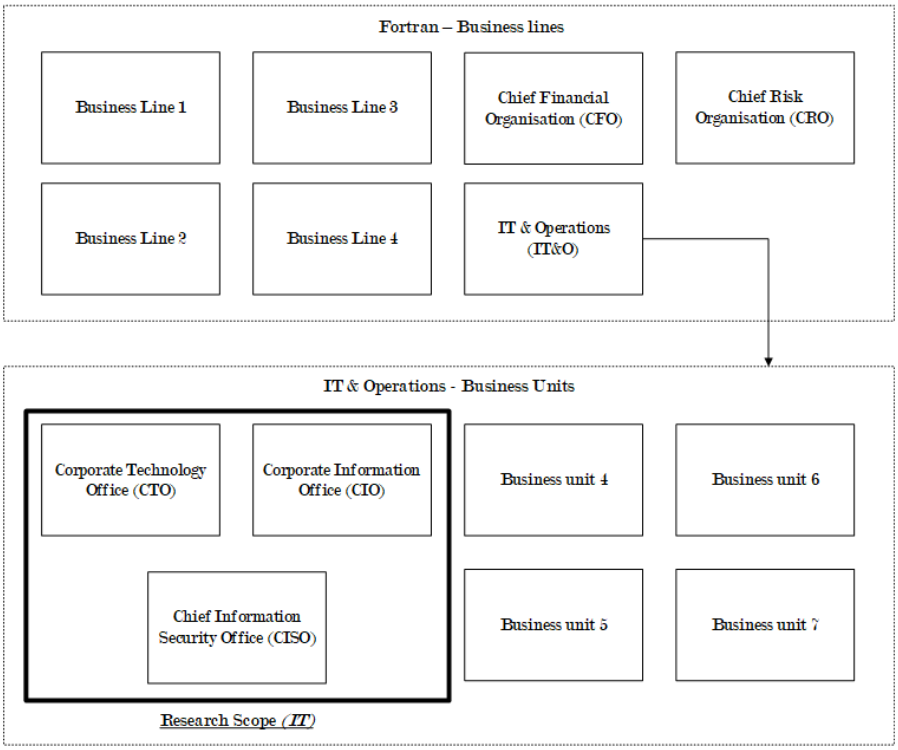

Figure 2.1: Fortran organisational structure and research scope

organisation to design, maintain, manage and improve the way IT is used. The business unit is responsible for managing the running software and hardware, and includes IT support for clients.

CORPORATE INFORMATION OFFICE (CIO) The Corporate Information Office (CIO) is split into two different departments which serve different business units within the bank. The CIO departments are mainly responsible for creating software solutions.

CHIEF INFORMATION SECURITY OFFICE The Chief Information Security Office (CISO) is the department that sets out security guidelines for the whole bank, which includes IT.

2.2.2 *Split of activities within IT*

The IT Value Chain by the Open Group [33] is used by Fortran to depict the IT lifecycle within the bank, the top row in Figure 2.2 displays the IT Value Chain. The main activities of the CIO and CTO organisations can be described with it. Four value streams depict this value chain. The process of developing IT services starts with the first stream *Strategy to Portfolio*, which is about managing the portfolio from business idea to items on the backlog of teams. This stream is about planning the project, evaluating the business strategy and designing the project plan. The next stream *Requirement to Deploy* is

about developing, building, testing and releasing functionality. *Request to Fulfill* takes the software to production and makes IT available to users. The last stream *Detect to Correct* ensures availability and monitors the running IT services. Request to Fulfill is not defined clearly within Fortran IT; the theoretical definitions are placed under the other streams in practice. This stream is, therefore, left out in further reference to the IT value chain in this research.

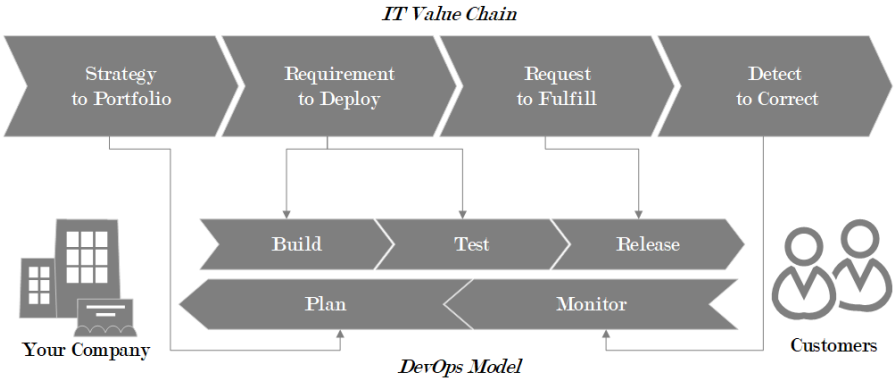

Figure 2.2: Open Group IT Value Chain mapped to AWS DevOps Model [3]

2.3 MAIN INITIATIVES WITHIN FORTRAN IT

2.3.1 *DevOps*

The bank originally split up software development and run activities over separate departments and teams. These teams used to work in isolation, in which the responsibility of a release of software created by development is passed over to operations, which creates a 'wall' between *Dev* and *Ops*. The bank is making a transition to integrate the change and run worlds; this movement is referred to as DevOps.

The mismatch between Dev and Ops is sometimes called the 'Wall of Confusion'

We found that DevOps can be described as an overarching term for several best practices, including a change in team composition and CICD (Section 2.3.2). The DevOps process models we found online are quite similar to the IT Value Chain model, such as [26], [3] and [5]. We created a mapping between the IT Value Chain streams and the different stages of the DevOps model by Amazon Web Services [3], see Figure 2.2.

DevOps changes the work teams perform. As a result, the composition of the organisation changes, since different roles are required. Some call the bank an off-shoring organisation, meaning that much of IT is outsourced to vendors in other countries. A large number of teams that work separately (and remotely) at vendors, will work together in teams of the bank when DevOps is adapted. A challenge within IT is to adapt to those organisational transformations.

2.3.2 *Continuous Integration and Continuous Delivery*

The goal of Continuous Integration and Continuous Delivery (commonly abbreviated as CICD) is to integrate tools such that the development process can be automated. It was found that the terms Continuous Delivery and Continuous Deployment are used interchangeably in literature [50]. Continuous Integration is a practice in which developers integrate their work frequently with each other. The work can be automatically deployed and released at any moment [50]. The goal at Fortran is to provide every team with the skills and tools to implement CICD practices. The main aspect of this implementation is to connect all tooling in order to create an automated pipeline. Each part of the IT Value Chain (see Section 6.3.1) can be linked with each other, in order to speed up the development process, reduce the number of manual actions and to get an end-to-end overview of how the final product has been created.

An end-to-end overview of the IT value chain can be created once these tools are connected.

IT is introducing the use of Public Cloud as a hosting platform for their applications. Most applications still run on-premise, meaning that own infrastructure is used in data centres.

3

RESEARCH DESIGN

This research project makes use of a literature review and a case study in order to address the gap in knowledge about data-driven organisations, to find out what the challenges within IT are and to construct a roadmap based on an existing Data Management framework. The study is performed according to a design science methodology, in order to redesign a Data Management framework. We followed the methodology from Wieringa [56], which presents a design cycle that guides efforts in order to redesign an artifact. The design cycle is described by the three main phases *problem investigation*, *treatment design*, and *treatment validation*. The design cycle is presented here as Figure 3.1, exclamation marks are design problems and question marks are knowledge questions. The design cycle is part of the engineering cycle, which also includes the phases *treatment implementation* and *implementation evaluation*.

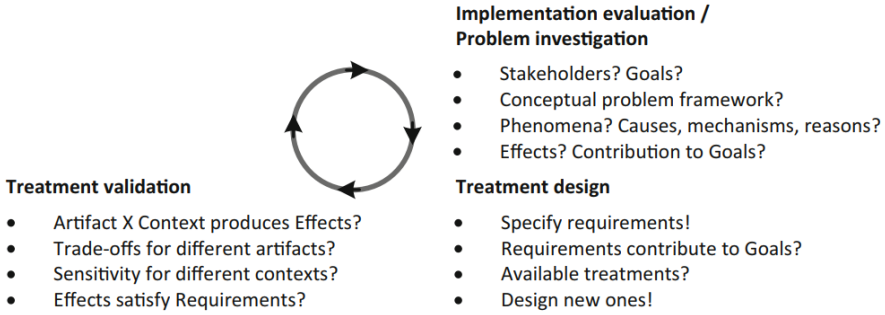

Figure 3.1: Design cycle as presented by Wieringa [56]

3.1 RESEARCH QUESTIONS

Wieringa [56] provides a template for formulating design problems. The template follows the following format: Improve <a problem context> by <(re)designing an artifact> that satisfies <some requirements> in order to <help stakeholders achieve some goals>. We formulate our design problem in Figure 3.2.

This research objective helps us formulate the main research question. Which we can formulate as follows:

What constitutes a usable capability model for Data Management in an internal IT organisation in a financial institution like Fortran?

Subquestions break this large question into more manageable pieces. The first step of the design cycle investigates the Problem Context.

Improve the usage of data for value generating processes
by redesigning a Data Management Capability model
that satisfies requirements that fit those of an internal IT organisation in a financial institution
in order to provide a Data Management roadmap.

Figure 3.2: Design problem formulation for this research

This step helps create a grounded basis for the rest of the research and will help define background information. In order to design a suitable Data Management framework, first three knowledge questions are formulated.

To define the problem context, it is first desired to find out why organisations want to make use of their data and what should be in place to facilitate value generation with data. This first research question aims to point out important aspects that need to be taken into account when organisations want to define themselves as data-driven.

1. What are key capabilities that support large data-driven organisations?

A Data Management capability framework is used as a basis for a design that is suitable for IT. Therefore, it is necessary to understand what Data Management is about and be able to discuss differences with what has been presented in literature. The following question is defined:

2. What is the current state and research agenda of Data Management Capability models?

The capability model that is designed is focused on the IT organisation of Fortran. It is a necessity to investigate what the responsibilities of the IT departments are in order to understand the context. It particularly should be known how data is used, what data is used, how data is shared, and what the challenges are that makes it challenging to make use of data effectively. We, therefore, find out what problems arise with their use of data, we define these problems as *data challenges*.

3. What are challenges with the use of data in the IT organisation at Fortran?

Data management has proven to be effective to get control of Fortran's data assets, but it is required to know what specific capabilities need to be in place to effectively implement it in its IT organisation. This reflection on the data challenges is done as part of the case study at Fortran.

This research also serves as a guide for the department that is responsible for Data Management of the IT organisation at Fortran. The desired situation is discussed and gives insight into priorities for Data Management for IT within Fortran.

4. How can Data Management contribute to challenges with data in the IT organisation at Fortran?

The results of the previous questions can be used to reflect a bank-wide Data Management capability model in order to be fit for the IT organisation in particular. The following question focuses on how the model can be adapted with what has been found in literature and during the case study:

5. How can a capability framework be redesigned to support Data Management in the IT organisation of Fortran?

3.2 OVERVIEW

Following the design science methodology according to Wieringa [56], the questions are grouped with regards to the stages of the design cycle. Figure 3.3 shows an overview of the relationships between the questions. The arrows indicate that the results of one step will be used in the successive step. Table 3.1 displays the research methods per research question.

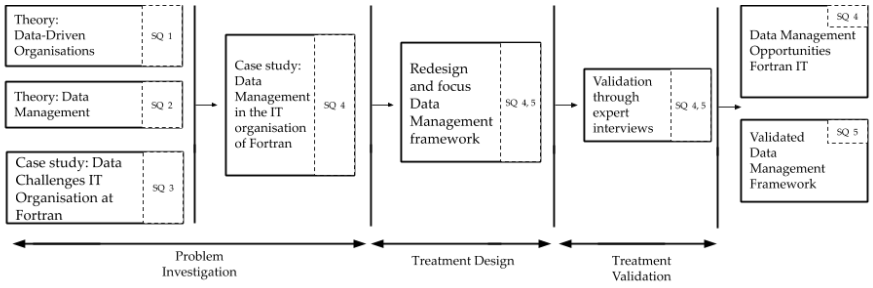

Figure 3.3: Research Design

QUESTION	METHOD
SQ1	Literature review
SQ2	Literature review
SQ3	Case study
SQ4	Case study
SQ5	Framework design, Validation

Table 3.1: Research methods per subquestion

3.3 SOURCES

An empirical study is performed in order to collect insights which can be used to redesign a Data Management framework for the IT organisation at Fortran. The results of the study are based on multiple sources.

SEMI-STRUCTURED INTERVIEWS Stakeholders within the IT business units are interviewed on the main projects, their Data Management needs and their perception of the current Data Management model. Stakeholders mainly include IT managers, IT transformation leads, team leads and software engineers.

INTERNAL DOCUMENTS The intranet of Fortran presents a lot of specialised documents. These documents present models, roles, metrics, guidelines, and more useful information that can be used as input for an adapted model. Some of the figures are taken over in this thesis.

MEETINGS During the research, we participated in numerous meetings about Data Management. The meetings provided insight into the current state of Data Management at the IT departments, as well as the Data Management efforts within other departments.

4

DATA-DRIVEN ORGANISATIONS

With constantly increasing amount and sizes of datasets, organisations are eager to turn data into value. Successful organisations come up with valuable use cases which exploit data, in order to support their business goals. But on the other hand many executives report that their companies are lacking in data and analytics [45].

Organisations want to make better use of data and want to become a 'data-driven organisation'. To become one, they need to know what it means and what is required to have in place to be data-driven. It is however hard to determine what should be done in the company to adopt this new organisational goal. There are many ways to measure if an organisation is data-driven [46], existing benchmarking models do not take organisational specific factors into account [44].

In this chapter we provide insight on what it means to be a data-driven organisation and what organisational capabilities need to be in place to align the value creation process from data.

4.1 METHODOLOGY

We used academic search engine Scopus to find academic literature on the topic. Articles are selected through iterations with different search terms. The search process follows a similar approach as described by Wolfswinkel, Furtmueller, and Wilderom [57], but is performed in a non-exhaustive manner, which conforms with the nature of a narrative review.

We found that many articles about data are domain specific and many academic articles go into technical depth, this is why the subject area filter was limited for all search queries to show only results in the 'Business, Management and Accounting' domain. We first used the keywords "data driven" to select papers which can provide context to the research field and provide insight in other concepts that could be interesting to investigate. Search results were selected for relevance by assessing the article's title, abstract, year and business journal. We did not select results that were about technological infrastructure; had a too specific context scope; or were published before the year 2010.

Metadata about the papers were logged, such as key concepts; reason for selection; a relevance score was assigned; and it was registered how the paper was found. After more concepts showed to be relevant, a similar process was performed for the keywords *"big data" implementation* (sorted on most cited first), *"data analytics" implementation* (sorted on most cited first), *data "capability model"* (all 104 results were

considered) and *data "maturity model"*. More literature was found by using back- and forward citations from selected relevant sources, as well as found through paper suggestions from literature hosting platforms (such as *ScienceDirect*) and through reading business journals of relevant sources.

Non-academic articles and technical reports about the field of research were also included in this report to provide an up-to-date overview from a practical perspective. This type of literature was acquired through searching for managerial business journals and white papers, as well as by using bench-marking studies on maturity models as a source.

4.2 DEFINING A DATA-DRIVEN ORGANISATION

4.2.1 *Definitions in literature*

In their 2011 article, Patil defines a data-driven organisation as one that

> "... acquires, processes, and leverages data in a timely fashion to create efficiencies, iterate on and develop new products, and navigate the competitive landscape." – Patil [46]

We extract three stages from this definition: 1. Sourcing Data, 2. Processing Data, 3. Using Data; with the goals: 1. to improve quality of processes, 2. use it as a driver to improve products, 3. find opportunities for new products, and 4. to find out what competition is doing. The author describes that assessing how mature an organisation is in its goal to become data-driven can be done by looking how effective data is used within the organisation. This definition is however relatively old, and ambiguous. The article explains how the first data science teams were formed at the time, while the field has matured a lot in the meanwhile.

In the article of Fabijan et al. [25]. They do not specify a literal definition of a data-driven organisation, but mainly refer to companies which use data to improve their processes and products. This article is an example that the definition of a data-driven organisation is highly context dependent within current literature.

These articles are both focused on their own context, but either give clues that a data-driven organisation is one that is able to successfully create organisational specific value by sourcing, processing and using data.

Buitelaar [10] wrote his master's thesis on data-driven organisations. With the use of an iterative design approach, the author presents a novel Data-Driven Maturity Model. This framework covers known theory and practise, and describes the journey to fully become data-driven. Buitelaar describes data-driven organisations as those who excel in turning data into action.

> "Being data-driven as an organisation means supporting your decisions with data-backed intelligence. But being data-driven is also about transcending isolation and integrating data-driven activities into your business processes. The goal is to enable all employees, not just business analysts or data scientists, to explore and exploit data. Data-driven organisations are those who have successfully empowered employees with data-driven capabilities: enabling them to optimize and innovate."
> – Buitelaar [10]

The author uses the better studied topics *business intelligence*, *business analytics*, *big data*, *data science* and *data-driven marketing* as the basis for the theoretical background of data-driven organisations. His definition points out the importance of enabling employees throughout the organisation to work with data and giving them the responsibility to create their own solutions.

In the following sections we present concepts which effectuate these principles.

4.2.2 *Related definitions*

The growing maturity of the data field has lead to the emergence of new concepts. Some of which share a similar theoretical basis, these data-related terms are highly correlated with each other. As this may lead to confusion and because lessons can be learned from work which is based on similar, but slightly different concepts, we first introduce terms we came across in the field of data and introduce their concepts.

Provost and Fawcett [47] also saw this parallel in the field. In their article the author presents their definition of *data science* and explain the relationships with other related concepts such that a better understanding of what data science has to offer can be created.

Buitelaar [10] based its data-driven maturity model mainly on Business Analytics and Business Intelligence, these two terms will be explained first.

4.2.2.1 *Business Intelligence and Analytics*

The main driver for data-driven decision making is analytics. Analytics translates data into actionable information, such that decisions do not have to be solely based on instinct, but also is based on facts. The term business intelligence describes a large combination of software and processes that can be used to collect, analyse and distribute data, such that it can support better decision making [20].

Chen, Chiang, and Storey [14] treat those two terms as a unified term, which describes its evolution through the emergence of technological advancements, which started in the database management

field. The authors explain the evolution from *BI&A 1.0* to *BI&A 2.0* due to web intelligence, web analytics, web 2.0, and the ability to mine unstructured user generated content. The article, published in 2012, marks Big Data as the enabler of BI&A 3.0.

> "[BI&A] is often referred to as the techniques, technologies, systems, practices, methodologies, and applications that analyze critical business data to help an enterprise better understand its business and market and make timely business decisions." – Chen, Chiang, and Storey [14]

The concept Business Intelligence & Analytics (BI&A) can as such be described as the combination of processes and assets in such a way that data can be translated into insightful information, that can support the decision making process.

Business Analytics (BA) and Business Intelligence (BI) are not new concepts, the term "Business Intelligence" was introduced in the late 1980's [20]. The growth curve of companies using analytics once was steep, but this growth is flattening out [36]. Despite this, in their 2014 research, Kiron, Prentice Kirk, and Boucher Ferguson also report that according to their survey with over 2037 business executives it was shown that 87% of organisations still want to step up their use of analytics to make better decisions. In total 39% of the respondents agreed and another 26% strongly agreed with the argument that their organisation relied more on management experience than data analysis when addressing key business issues.

Even though business analytics and business intelligence have been around for some time, it seems to be still relevant today and still are susceptible to technological advancements such as Big Data. Organisations want to make data-driven decisions, but are still working on evolving their maturity level in the field.

4.2.2.2 *Data-driven decision making*

Data-driven decision-making (DDD) can be described as the act of making decisions based on data, instead of pure intuition [47].

There is evidence that DDD is linked with better firm performance, Brynjolfsson, Hitt, and Kim [8] found that firms that adopted a data-driven way of decision making have up to 5-6% higher productivity and output, compared to what would be expected if the same investment would be made for other information technology usage.

It has also been found that the share of US manufacturing organisations that adopted a data-driven decision-making approach has tripled to 30% in between 2005 and 2010 [9].

Newer sources also present evidence that DDD is still relevant. Rejikumar, Aswathy Asokan, and Sreedharan [48] found in their study, among 173 practising managers in Indian industries, that the main

reason that managers do not adopt a data-driven approach to decision making is due to the lack of confidence on the technological readiness. When empowering managers with appropriate technical and analytical skills by training, enables them to enhance their 'absorptive capacity' to adopt data-driven approaches. The authors find factors that can contribute to increasing the confidence to take innovative practices such as data-driven decision making: resource availability regarding capital, infrastructure, and trained workforce [48].

4.2.2.3 *Big Data*

The term Big Data has been extensively used by all sorts of organisations to describe their effort to turn data into value, however what is meant with this term is not always similar.

McAfee and Brynjolfsson [43] explain a difference between traditional analytics and the big data movement. The main differences can be described with the use of three V's: Volume, Variety and Velocity. Volume is used to describe the possible large size of data, Variety the different types of files that can be processed and Velocity the speed at which data can be processed to turn it into value. Since the introduction of these data-management challenges by Laney [39], more dimensions have been added to this definition, including Value and Veracity [21].

In their research about big data analytics and firm performance, Wamba et al. [55] provide the following definition:

> "Big data analytics capability (BDAC) is broadly defined as the competence to provide business insights using data management, infrastructure (technology) and talent (personnel) capability to transform business into a competitive force"

We have learned that the principles to describe an effective Big Data strategy are about data issues that are also relevant in a context where Volume, Variety and Velocity are less important. This statement can be supported by the explanation of El-Darwiche et al. [24]. They view the concept Big Data as an evolution through various stages described by 'buzz words' like *data mining* or *business intelligence*, each with the goal to create meaningful information for business purposes from raw data. Figure 4.1 shows the relationships between the concepts that evolved into big data.

> "Big data, may appear all-enveloping and revolutionary. However, the essential principles for exploiting its commercial benefit remain exactly the same as they were in previous moves toward increased data-driven decision-making."
> – El-Darwiche et al. [24]

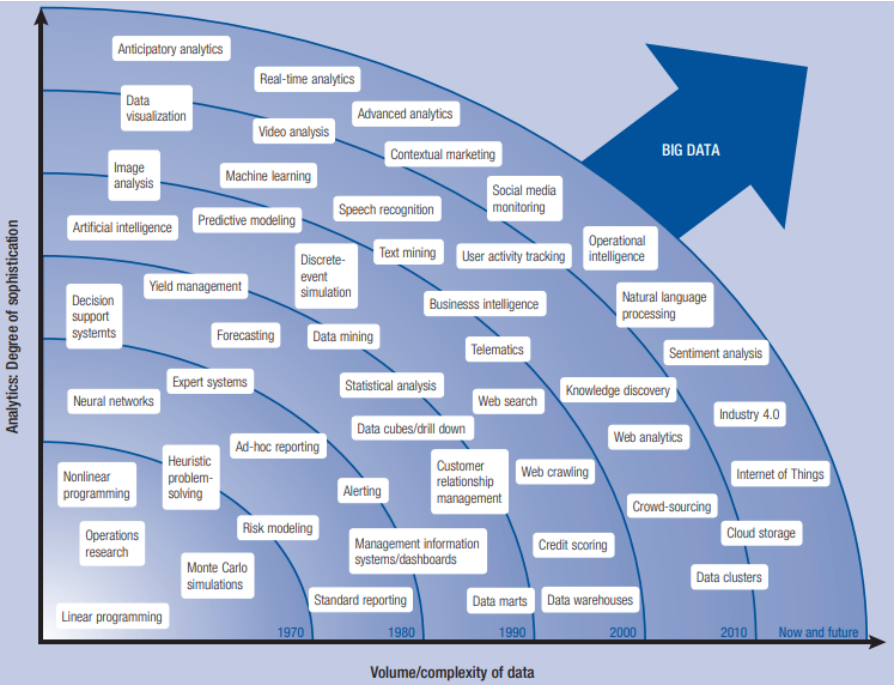

Figure 4.1: Concepts that evolved into Big Data according to El-Darwiche et al. [24]

4.3 MODELS FOR DATA-DRIVEN ORGANISATIONS

Literature provides us capability models on the concepts Business Intelligence and Big Data, these models describe the cornerstones of an organisation that leverages the potential value of data. Some focus on concept specific aspects, while others provide a more high level overview of what is needed to become data-driven. In this section we review capability models about the subjects presented in the previous section and will reflect on their foundations.

Organisations can find it useful to create structure around a program, define the organisation's goals around it and create a vision which can be communicated across the organisation [28]. A maturity model can help those organisations with guidance to evolve their capabilities. These models are sometimes referred to as capability maturity models, or simply capability models.

We only extract the dimensions that are used to support the model, but do not evaluate stages of maturity presented in those frameworks.

Wamba et al. [55] conducted a review on big data analytics (BDA) capabilities in which they showed the relationship between firm performance and BDA.

The research model describes big data analytics with the constructs infrastructure, management and personnel. These are further described by constructs about technical compatibility, project management, and domain knowledge [55].

Another model is published in a report by the professional services firm EY, which conducted a survey with 270 senior executives and was used as a basis for a big data capability framework [23]. This white paper is mainly focused on identifying obstacles which industry comes across with the introduction of big data. They indicate that value creation with data is supported by decision making, technology, analytics, ownership and accountability (data governance), and security.

Keppels [34] performed research on Business Intelligence maturity models. The author decided to use the Business Analytics Capability Framework (BACF) by Cosic, Shanks, and Maynard [16] as a basis for further research. Keppels praises the framework for its strong theoretical basis, but points out the lack of operationalisation. The model by Cosic, Shanks, and Maynard presents 16 capabilities, grouped under four capability areas: Governance, Culture, Technology and People.

Although each of the three models provide capabilities from an unique perspective, the key areas within the models show overlapping concepts. The following section will add more concepts to this convergence of theory.

Braun [7] benchmarked big data maturity models. The models were evaluated based off multiple criteria, namely: Completeness of the model structure, the quality of model development and evaluation, ease of application and Big Data value creation. It was concluded that the maturity model by Halper and Krishnan [28] was overall the best, followed by IDC [32] and El-Darwiche et al. [24]. The first and last of these provide a visual model with explanations, while IDC only provides an online maturity test.

The model of Halper and Krishnan [28] consists of a sequence of maturity stages. To assess the maturity of a company's big data strategy criteria were used which were grouped under the key cornerstones: Infrastructure, Data Management, Analytics, Governance and Organisation. The criteria are presented here in Figure 4.2.

El-Darwiche et al. [24] provide an article which largely consists out of industry examples. The article touches aspects which are used to describe capabilities which we previously found, such as launching a data-driven decision culture or training talent, but does not provide a structured presentation of key capabilities.

As previously introduced in section 4.2.1, Buitelaar [10] reviewed maturity models in the field of data-driven organisations. In the master's thesis the author presents a review on maturity models published in literature and present a novel data-driven maturity model. The author saw the need for a formally built and validated model in the field of data-driven maturity and analytics. Most maturity models are not academic publications, but are covered in grey literature, such as white papers and blogs. These sources cannot be established if it is validated by peers. Buitelaar reviewed maturity models published between the

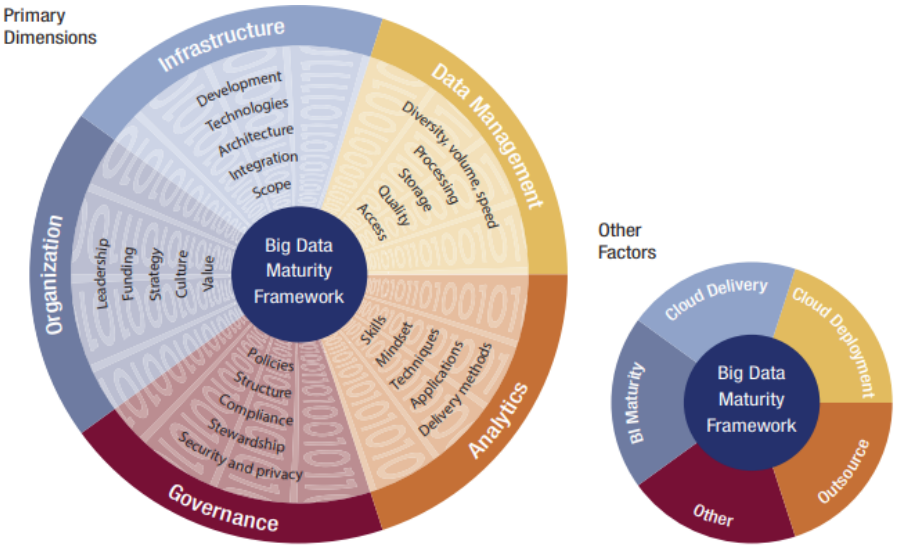

Figure 4.2: The features used for scoring on the big data maturity model according to Halper and Krishnan [28]

years 2007 and 2017. Based off criteria found in literature eight key dimensions which cover the most important principles of data-driven maturity are presented: Leadership, Data, Culture, Metrics, Strategy, Skills, Agility and Technology. The author adds another two dimensions with special focus to the importance of integrating analytics throughout the whole organisation: Integration and Empowerment. These are dimensions which are different than the main capability areas of other models. However, we can group most of these dimensions from Buitelaar under the previously introduced categories. The Leadership and Strategy dimensions can be grouped under Management area identified by other models, Data under Data Management, Metrics under Analytics, Skills under Personnel, Empowerment under Culture. Although the focus of Agility and Integration is clearly different, the main principles can be described with the use of dimensions Management, Culture and Analytics.

4.3.1 *Conclusions*

Both maturity models and capability models provide areas which companies should focus on to become more data-driven. In this section we discussed models in literature and identified their main focus areas. We grouped the main focus areas of the capability and maturity models discussed in this section and present those areas models in Table 4.1. We found that management, organisational culture, infrastructure, personnel, analytics, data management, governance and security were marked as important factors by the frameworks.

Management	[23], [28], [35], [55]
Organisational culture	[16], [35]
Infrastructure / Technology	[16], [28], [35], [55]
Personnel / People	[16], [35], [55]
Analytics	[23], [28]
Data Management & Governance	[16], [23], [28], [35]
Security	[23]

Table 4.1: Key areas in capability and maturity models in literature

4.4 DISCUSSION

Data-driven is a term that previously is typically used to describe an approach that focuses on the use of data within highly organisational specific processes. It is clear that organisations can benefit from making better use of data, as it may lead to increased firm performance and better decision making. Although many organisation recognise the need to adopt data-driven practises, many are still lacking in their strategy and are unfamiliar with the approach they should take to become a data-driven organisation. Due to the organisational specific factors, application of data-driven practises in one organisation may not be applicable in another context. Literature provides us with limited definitions on what it means for an organisation to be data-driven, the general consensus is depicted as as an organisation that is able to effectively turn data into value. It is stressed that organisations need to use data to back decisions, integrate it in organisational processes and to give employees the opportunity and tools to create their own solutions.

We found that the relatively new concept 'data-driven' emerged from previously studied concepts, such as business intelligence, business analytics, data-driven decision making and big data. These concepts are highly related and are described as direct enablers of each other. In this sense, Business Intelligence & Analytics effectuate Data-Driven Decision Making and the former continues to transform based on technological advancements, such as Big Data. Since these concepts are so much related, their theoretical basis are as much alike.

Literature provides us with capability and maturity models about these topics, which turned out to share similar key capability areas. The transition from a company with data to a data-driven company is enabled by more than technology alone. Other factors, such as organisational culture, skilled people, management buy in, analytics solutions and data governance are hugely influential for a successful adoption of a data-driven way of working.

The foundation of those capabilities is based off concepts which share similar concepts at their basis, but still have different purposes

and angle of perspective. Big Data is for example an advancement that is a lot newer than Business Analytics itself, still the capabilities which we found need to be in place for both. For this reason we believe that highlighted capabilities are largely insusceptible to further technological advancements in the analytics field.

4.4.1 *Definition*

With the use of those capabilities and with literature we provide the following definition of a Data-Driven Organisation.

> A data-driven organisation can be described as one that is successfully able to turn data into value, with the use of management alignment, organisational culture, infrastructure, skilled personnel, analytics solutions, effective data management & governance.

4.4.2 *Capability model*

Based on the findings of this exploratory research we can construct an unified capability model for data-driven organisations. The model should be used as a common thread to be able effectively use data. We have mapped the main relationships between the different capabilities, to illustrate how they interact in a data-driven organisation. The model is presented as Figure 4.3.

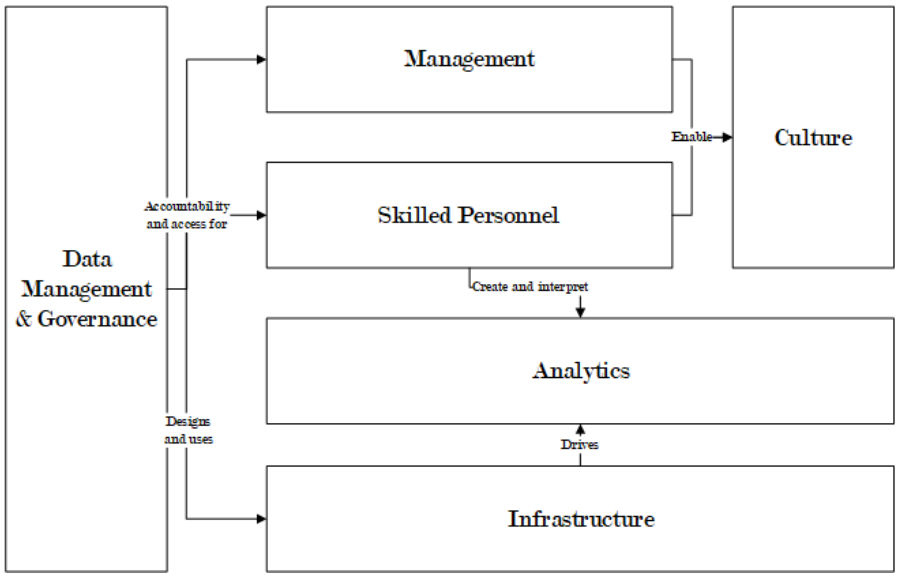

Figure 4.3: Capability model for data-driven organisations.

5 DATA MANAGEMENT

One of the key capabilities from the previous chapter is Data Management. According to DAMA International [17], Data Management (DM) "is the development, execution, and supervision of plans, policies, programs, and practices that deliver, control, protect, and enhance the value of data and information assets throughout their lifecycles.".

In contrast to the high level organisational capabilities described the previous chapter, Fortran focuses on tackling practical data challenges with the use of Data Management. In this chapter we perform a literature review about industry standards with the goal to test the validity of the Data Management Framework by Fortran and to be able to relate it to the Data Management approach of other organisations. We also explain its overlap with the Data-Driven model as defined in Chapter 4 and highlight the lack of literature on Data Management in IT.

The definition of DAMA International includes a wide range of organisational aspects, it not only covers infrastructure, but also describes the change for the way of working. While in more technical domains, such as in software engineering, Data Management is often about how data assets can be processed on operational level (for example in [40] and [29]). We use the definition and concepts of *DAMA-DMBOK2* [17] to describe Data Management.

5.1 METHODOLOGY

5.1.1 *Industry literature on Data Management*

The need for better Data Management can be addressed with the use of frameworks created to guide organisational transformation efforts. Metamodels provide an overview and guidelines to implement Data Management throughout the organisation. Metamodels were found with the use of the academic search engine Scopus with the key words *"Data Management AND framework"*, filtered on *Business, Accounting and Management*. The first 100 results were considered, but did not return industry standards. Another search with the keyword *"DMBOK"* was used, but also did not retrieve results with other industry standards. A Google search on *data management models* was used to find frameworks comparable to *DAMA-DMBOK2* [17], the first two pages with results were considered.

5.1.2 *Academic literature on Data Management in IT*

This research focuses on Data Management for IT departments in banks. The term *IT (Information Technology)*, is broad, and may have many different meanings. In our context we define IT as the department in an organisation which creates, deploys and maintains software solutions. We use the definition of Data Management as described in the first paragraphs of this chapter.

A literature scan was performed on Data Management practices for IT. Software development can be described as activities to create, design, deploy and support software [30]. Software Engineering is closely related term, but focuses more on applying engineering principles to create software for specific functions [31]. These activities are similar to the main responsibilities of the IT departments.

First we queried the search engine *Scopus* for the search terms *"data management" AND "software engineering"*. Only papers published from the year 2012 onward were considered. All 225 search results went through a selection process. The process followed a similar methodology as described by Kitchenham [37] in which a selection was made based on title and abstract and in a second iteration based on the introduction and conclusions, selection criteria can be found in Table 5.1.

INCLUSION	EXCLUSION
Data generation in software engineering or development	Technical data infrastructure
Data usage in software engineering or development	Database management
Data management in DevOps	Blockchain
Data management and Public Cloud	

Table 5.1: Inclusion and exclusion criteria

The first iteration resulted in a limited selection of results, but were later discarded in the second iteration. Most of the found results were technical and described Data Management as a way to organise datasets in databases and in code at operational level, unlike the more high level concept of Data Management as described in the first paragraphs.

We also performed a search on the search terms *"Data Management" AND "Software Development"*. Just like the previous search, only papers from 2012 onward were considered (126 results total), a similar methodology was used and the same inclusion criteria were used. The

first iteration resulted in a selection of 5 papers, 2 were discarded in the second iteration.

5.2 DATA MANAGEMENT MODELS

Data Crossroads performed a comparative analysis on six Data Management maturity models. Each of the models provide a guideline for Data Management and Data Governance. Seven key subject areas were defined: Data, Data and System Design, Technology, Governance, Data Quality, Security and Related Capabilities. The author created a metamodel which displays the key capability areas per maturity model in one overview, it is adapted here as Figure 5.1.

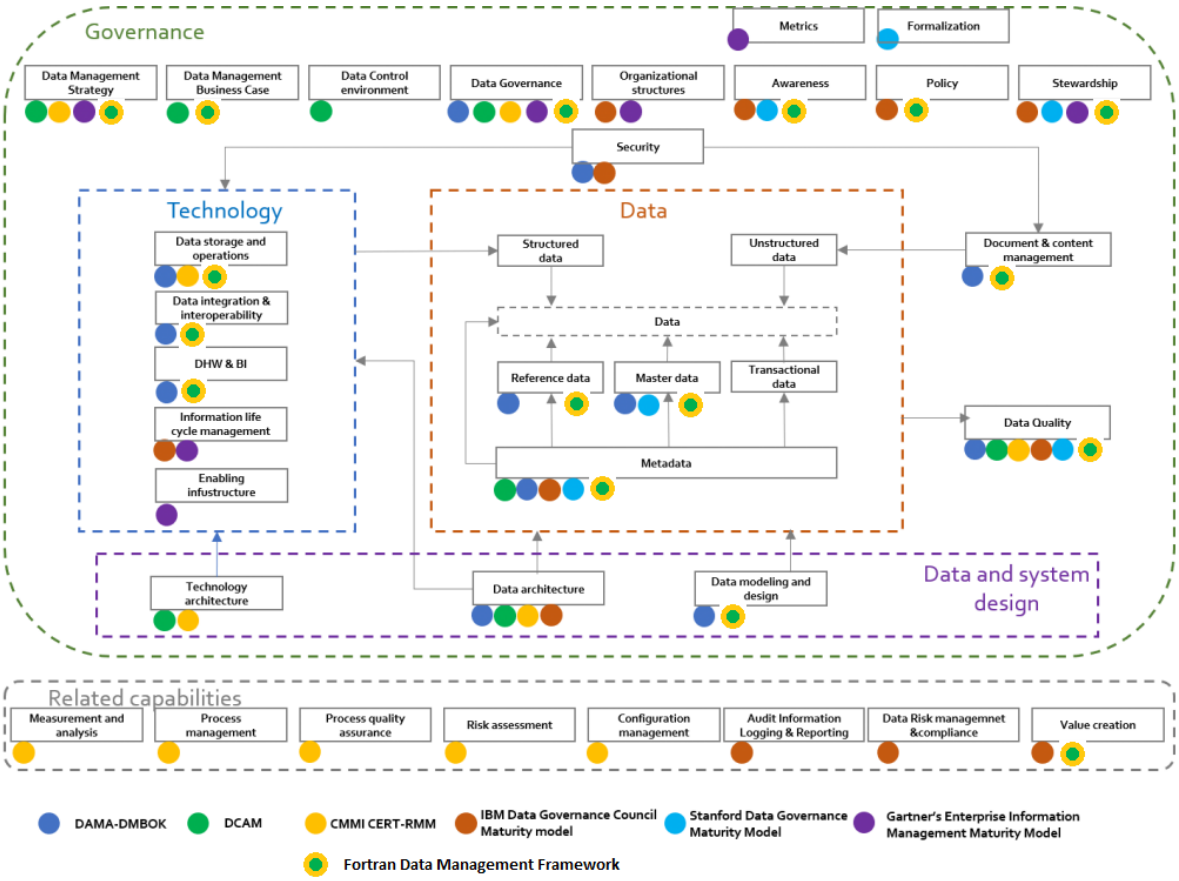

Figure 5.1: Comparison Data Management Metamodels by Data Crossroads [18] combined with the Fortran Data Management Framework.

We can see that some key areas overlap. The largest consensus in the Data Management models can be found in the areas of data governance, data management strategy, stewardship, metadata (management), data quality and data architecture. It is worth noting that the DMBOK model does not provide a key capability area for Data

Management strategy and stewardship. The DMBOK model however, describes Data Stewardship as a sub-category of their central Data Governance capability.

The next sections will describe the different metamodels.

5.2.1 *DAMA-DMBOK2*

As previously introduced, Fortran used this framework as input for their own model.

The Data Management Association (DAMA) published a framework for DM, their book describes trends and guidelines for Data Management [17].

The DMBOK model presents a wheel with Data Management knowledge areas, adapted here as Figure 5.2. Data Governance is placed as the centre of the wheel, as governance is required for consistency and balance between the functions. The other areas are placed in a circle around the centre, displaying the knowledge areas that are necessary for mature Data Management. The descriptions of the areas are adapted in Table A.2.

DAMA recognised that the desire by organisations to create and exploit data has increased, and so has the need for Data Management practices. The association created the functional framework with guiding principles, widely adopted practices, methods and techniques, functions, roles, deliverables and metrics. The framework as well helps establish a common vocabulary for Data Management concepts.

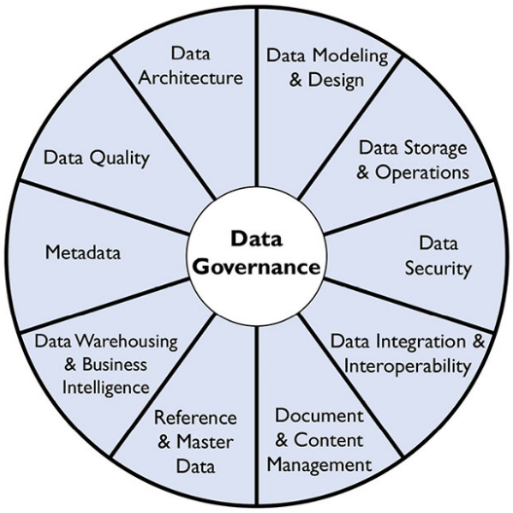

Figure 5.2: The DAMA Wheel adapted from DAMA International [17]

5.2.2 *Other frameworks*

The DCAM model [19] was originally created for Data Management in financial institutions. Although it includes industry specific components, Data Crossroads recognises it suitable for other industries

as well [18]. DCAM does not see Data Management as solely an IT function, the model views IT as part of the organisational ecosystem.

The Capability Maturity Model Integration (CMMI) is an organisation that collects and creates industry best practises in order to provide organisations with guidelines and maturity assessments on key their 'business challenges'. The CMMI created a Data Management maturity model as well. Just like DMBOK and DCAM, this model is not freely available, and so limited information about it is available. The CMMI Data Management Model provides six key capability areas which can be used to identify strengths and gaps, and provides best practises to leverage data assets [11].

Other recognised frameworks include Gartner's Enterprise Information Management Maturity Model, Stanford Data Governance Maturity Model and the IBM Data Governance Council Maturity model. These mainly focus on data- or information governance, and as such are not as extensive as the previously described models. According to Data Crossroads the differences between the approaches of the models is not clear [18], and as such might not be as useful as larger frameworks such as DMBOK, DACM or CMMI.

5.3 FORTRAN DATA MANAGEMENT FRAMEWORK

This section is part of the problem investigation stage in the design cycle, as shown in Figure 3.1. The artifact in this research is the Fortran Data Management Framework. We describe it as an architectural conceptual framework, which can be seen as a set of definitions of concepts, often called constructs [56]. Constructs are used to describe the structure of the artifact and its context. The architecture of the framework is known, the mechanisms and constructs are already described, ready to be implemented. This research focuses on the usage of the Fortran Data Management Framework in another context. Namely in the context of IT, instead of the whole organisation.

The Data Management Framework is presented in Figure 5.3. The descriptions of the capabilities are presented in Table A.1.

5.3.1 *Data Management at Fortran*

This research explores how Data Management (DM) could tackle the data challenges within IT. Data management is however not a new concept for Fortran. The first initiative started as Enterprise Information Management in 2012, it focused on Master Data Management, analytics dashboards and advisory. The scope widened to Data Management in 2014. The bank has a business unit, FDO, which is in charge of Data Management. This unit initially focused on data quality management and extended its responsibilities with providing guidelines for DM for the rest of the bank.

Data Ownership refers to setting accountability on a data source, Data Usership refers to appointing representatives for the users of the data

The implementation of Data Ownership and Data Usership, and Data Management are key strategic initiatives for Fortran to effectively turn data into value according to FDO.

The department has created a Data Management capability framework, presented in Figure 5.3, which provides focus areas for every business line of the bank.

The model is created to support the organisation in their Data Management efforts. The framework and the accompanying dashboard are essential for management to track progress across the organisation. The implementation is essential to roll-out a bank-wide model which in turn can generate value over time.

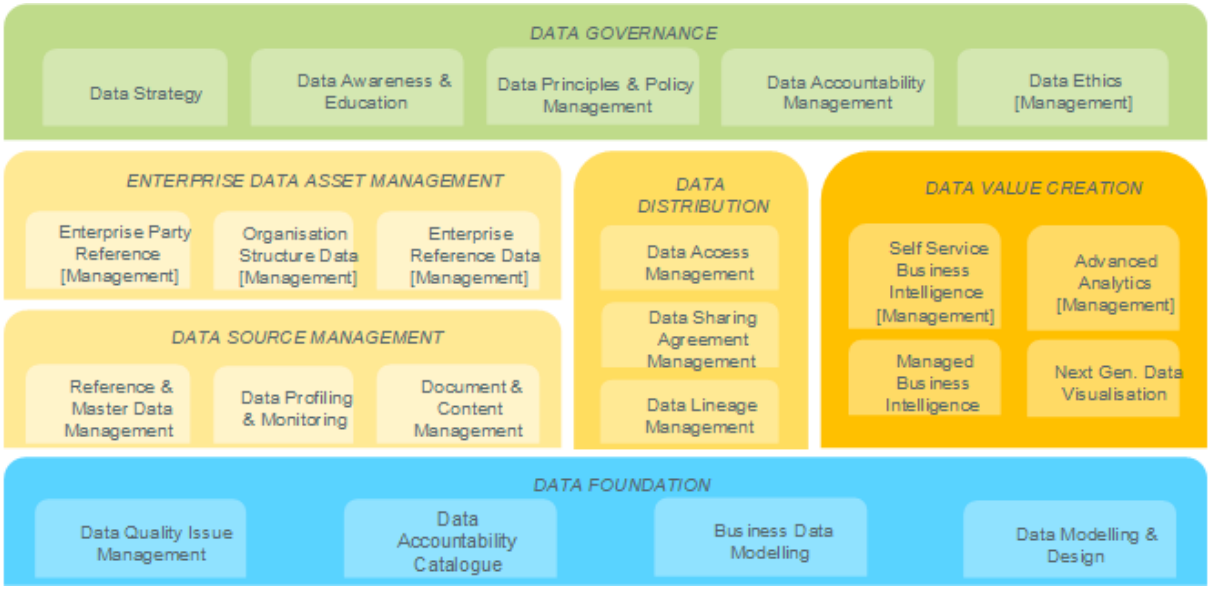

Figure 5.3: Data Management Framework designed by FDO

The model is based on DAMA-DMBOK [17], this model was used before within the bank as a guideline for Data Management. The model is mainly based on three key pillars, namely the DAMA-DMBOK, Fortran's business capability model and Fortran's organisational context.

5.3.2 *Similarities with other frameworks*

The Data Management Framework (DMF) of Fortran shares capabilities with the industry standards. The DMF, displayed in Figure 5.3, can be mapped on the metamodel analysis. Our mapping can be found as well in Figure 5.1. The mapping is done based on the description of the capabilities.

The Fortran Data Management Framework covers most aspects of other Data Management models. At first glance the model does not cover every aspect of the model, but does cover more aspects than the DMBOK model. It provides a capability for stewardship, awareness, data management strategy, policy, and provides special

focus on value creation where this is left in the DMBOK model. The DMBOK key areas that the Fortran model does not cover are Data Architecture and Data Security. During a meeting with the authors of this framework, it was explained that another division within the organisation was focusing on security for the whole company (the CISO department), and that the security aspect is mainly included in the *Data Access Management* capability in the framework. The Data Architecture capability was described as a responsibility of the IT architecture department, and is thus not seen as a data management capability. It can however, be argued that all data related capabilities should be placed in one central model to keep a proper overview. This may help keeping a total overview of the data strategy, instead of transferring the capabilities to separate models.

Other capability areas that the Fortran Data Managemetn Framework does not cover, but are covered by two frameworks or more are: Information Life Cycle Management, Technology Architecture and Organisational Structures.

The terms which can be found in other frameworks sometimes have different names than what is used in the Fortran model. The reason for this is that many of the descriptions in the capabilities use terms that were used before in the organisation. An example is the *Data Accountability Catalogue*, which was and still is used to register datasets that are available in the organisation. It was indicated that the key concepts found in literature may be grouped under different capabilities in the DMF as well.

5.3.3 *Operating Model*

The DMBOK2 framework also describes other aspects that are important to keep in place when implementing a data management framework. The operating model is described as an important aspect. A data management model should be fit for the context it is aimed to be implemented in, and therefore it is necessary to describe how process and people will collaborate. The framework describes several levels of centralisation for Data Management. The most informal level is the decentralised operating model, in which there is no single owner, and responsibilities are spread over different parts of the organisation. The other operating types (network, hybrid, federated, and centralised operating models) increase the level of centralisation by adding one or more Data Management groups within the organisation which share responsibility [17].

We can depict Fortran's operating model as a hybrid model, in which there is a combination of a centralised Data Management department and a shared responsibility within business lines and units. The operating model is adapted here as Figure 5.4.

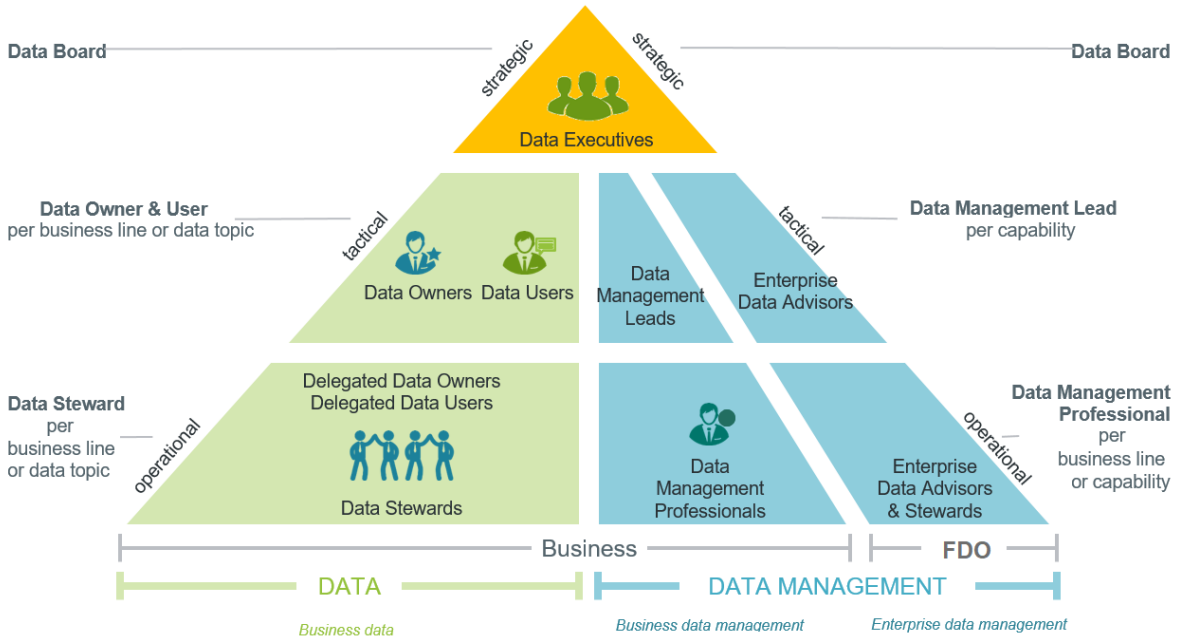

Figure 5.4: The Fortran Data Management Operating model

DMBOK describes a hybrid model as one with a centralised centre of excellence and steering committee. At Fortran, the FDO business unit dictates the guidelines and architecture for the rest of the organisation. The business lines within the bank have dedicated Data Management teams that help business lines with the execution of Data Management.

5.3.4 *Mapping the Fortran model to the Data-Driven model*

In Chapter 4 we presented a capability model for data-driven organisations. This section reflects on this model compared to the Fortran Data Management Framework (DMF).

The two models have different points of view on the use of data within an organisation. The Data-Driven Model (hereafter DDM) includes the capability area data management, which essentially is what the DMF presents. The DDM presents a relationship between organisational aspects on very high level, while the DMF provides a more detailed view on data management capabilities. The DDM presents organisational aspects, which are not tangible, such as organisational culture and management alignment. The DMF however, has capabilities with concrete data management goals. The two models do overlap partially. Both models present governance as one of the key aspects, which include clear policies and stewardship. Another similarity is skilled personnel, besides the technical skills the personnel needs, it is required that personnel is trained to understand the value of data management. The DMF has placed the latter under the capability 'Data Awareness & Education'. Although analytics is not seen as a

separate capability by the DMF, it could be grouped under the value creation capabilities. Infrastructure is not named in the DMF, it could be seen as part of capabilities that are already present in the model. But as the data architecture capability was also missing in the previous analysis in Section 5.3, it might be an opportunity to consider adding a data infrastructure related capability to the framework.

The DMF could be enhanced with capabilities about organisational culture and management alignment. It would be important to make those capabilities more tangible, such that the capability can be made actionable and the organisation could make effort to mature in those areas.

5.4 DATA MANAGEMENT IN IT

Shaykhian, Khairi, and Ziade [49] describe that IT departments aim to choose a Data Management architectural model to help bridge the gap among their organisations, technologies and customers. Such model, in combination with data quality management tools, provides the companies a trusted information foundation to base their analytics on. The authors describe operating models for data management based on two types. With a centralised model the organisation organises and manages enterprise data in a central repository. A federated model on the other hand, does not keep all data in one database, but keeps data in multiple places. It was concluded that centralised models are the best option considering factors such as cost and availability, since all applications consume from the same source. Whereas federated models include a complexity factor that introduces more costs and more problems regarding availability. The federated model could be used as a short-time solution, before moving to the longer-term-strategy with a centralised model.

The research by Thakar et al. [52] followed a 'data management' team in a year that an acquisition with a similar sized company took place. The research did not explore Data Management aspects as defined in Section 5, but explored how Process Mining could address solving data complexity issues that highly dynamic networked and global processes introduce in modern international software businesses. The authors describe that using a solution that can help software projects with discovering and managing important data assets without performing data analysis from scratch [52]. Their solution at the investigated company was able to find relations between applications, services, databases and legacy systems on premise and on cloud systems. The other benefit was that duplicate bad data could be found. Based on this article we see the potential of using Process Mining for IT, as an approach to identify datasets, to identify IT assets and identify relationships in complex application landscapes.

Capilla, Dueñas, and Krikhaar [12] describe Software Configuration Management as a software engineering discipline that addresses practical problems related to the identification, storage, control, definition, relation, usage, and change of the pieces of information. These problems might be prevalent in IT, therefore the importance for IT of this concept may be interesting for further investigation.

Based on these results, we found that the topic of Data Management in IT is underexposed in academic literature. No academic literature could be found on data management frameworks such as DMBOK2. Data Management in IT is also not touched in literature, as far as we found. This research can contribute to filling this gap, by providing one of the first academic writings on Data Management as provided in industry standards and by providing an insight in the industry application of Data Management.

6 RESEARCH METHOD CASE STUDY

The case study provides input to answer research questions 3 - 5.

6.1 GOAL

This case study identifies key initiatives within the IT departments of the bank, use of data within those departments and their challenges with data. We also reflect on how Data Management can play a role in assisting with those challenges. This provides input to improve and adapt the Data Management Framework (Figure 5.3) to suit the needs for the IT business units within the bank.

6.2 INTERVIEW QUESTIONS

This case study is performed with the use of expert interviews. The experts have different roles throughout the IT organisation and are asked about their experiences with data and Data Management, and for the IT scope in particular. The interviews are semi-structured, open questions are used as guidelines. Follow-up questions are asked if the answer describes particular situations seems to provide insight to answer a research question.

The experts are first asked about their position and experience to provide context of their role within IT. The next part of the interview is about uncovering main initiatives within IT in the bank. Each interviewee is asked to list the three most important projects within IT as part of this topic.

The next part of the interview is used to create an understanding if the interviewees have the same definition of data, how data is already being used, what types of data are used and what hinders turning data into value in IT. These questions aim to clarify how IT departments in banks use data already and what the potential value of data could be for IT. The main goal of this part however, is to uncover issues with regards to Data Management (DM), such that we can reflect what DM aspects need attention for IT within the bank. This part of the interview contributes to answer research question 4.

The last part of the interview focuses on Data Management in particular. We ask the experts about their opinion, understanding and frustration on Data Management. We do this to find out if there is need for Data Management in IT, to uncover the current state of the Data Management implementation and which aspects contribute the most

for IT. This part of the interview also contributes to answer research questions 4 and 5.

As a last part we include focused questions about Data Management as found in literature and as defined in the Data Management Framework by Fortran (see Section 5.3). We only ask these focused questions if the topic has not already been touched before in the interview. This part contributes to answer research questions 4 and 5.

All parts provide insight in the current state of Data Management within IT. The different results of the interview are combined and afterwards reflected on the Data Management Framework. This provides answer to research question 5.

6.3 CASE STUDY PARTICIPANTS

The experts that we interviewed have different functions within IT. We spoke with people which represent the different main working areas within IT. As described previously Section 2.2.2, the main activities within IT can be described by the IT Value Chain. We therefore interviewed managers across the different value streams. Managers are interviewed, since they should have a good overview of the current initiatives and priorities of the department. But since the managers view IT from mainly a strategic or tactical perspective, we as well included a number of other employees in the case study to receive their view on an operational level. Fortran IT is going through a number of organisational changes, we therefore also interviewed initiative leads of these changes, to find out what the role of data and data management could be in those initiatives. We mainly interviewed employees from the CTO organisation as they are responsible for providing tools and support for IT teams, including the CIO organisation.

This Table 6.1 provides an overview with the participants.

6.3.1 *Mapping the different viewpoints to the IT Value Chain*

The IT Value Chain is visualised in the top row of Figure 2.2.

The teams that work on development and on operations at Fortran are especially present in the Requirement to Deploy and the Detect to Correct value streams of the IT Value Chain (see Section 2.2.2). The Request to Fulfill stream is partially interpreted as part of the Requirement to Deploy stream at the bank, the responsibilities are also split over development and operations. The Strategy to Portfolio stream is the main responsibility of business stakeholders and is used as a setup for the rest of the streams. This is why our case study at Fortran mainly focuses on the two core IT value streams Requirement to Deploy and Detect to Correct and touch the role of Request to Fulfill.

DEPARTMENT	INTERVIEWEES
Executive Board	- Officer (RP1)
IT Consultancy Office	- Consultant (RP2) - Data Management (RP3) - Consultant IT Strategy & DevOps (RP4)
CTO	- Transition Manager DevOps (RP5) - IT Service Management (RP6) - Incident, Problem & Change Management (RP7) - End-to-End Availability Management (RP8) - Software Lifecycle (RP9) - IT Service Management (RP10) - Senior Software Developer (RP11) - DevOps CICD Developer (RP12) - Software Lifecycle (RP13) - Scrum Master (RP14)
CISO	- Identity & Access Management (RP15)
FDO	- Business Architecture (RP16)

Table 6.1: Interview participants

The interview candidates provide a broad perspective of the IT Value Chain, the interviewees are linked to the value streams as displayed in Table 6.2.

STREAM	INTERVIEWEES
Strategy to Portfolio	RP1, RP11, RP16
Requirement to Deploy	RP1, RP4, RP5, RP13, RP14,
Detect to Correct	RP1, RP4, RP5, RP6, RP7, RP8, RP10
Supporting Capabilities	RP2, RP3, RP9, RP12, RP15

Table 6.2: Interviewees mapped to the IT Value Chain

6.4 INTERVIEW SESSIONS

The interviews are held individually with the researcher, in sessions of generally 60-75 minutes. Due to tight schedules of the interviewees, two interviews are limited to 30 minutes. During one of these shorter interviews we primarily ask about the value of data and the things that do not go well yet, but leave out questions that go in depth about Data Management. For the second one, we prepare eight statements about data which the interviewee responds to, follow-up questions are asked. The statements are based on what we found in previous interviews and

are based on the knowledge area of the interviewee. One interview is conducted in English, the rest are conducted in Dutch. The interviews are voice recorded and short notes are taken by the researcher during the interviews. The interviews are later transcribed literally, these transcription documents are used to create the case results chapter. Passages of the transcribed documents are taken, translated to English, grouped by topic and according to the participant's value stream and changed phrasing for better readability. The notes taken during the interviews are mainly used as a guide for follow-up questions for the interview and are used as a back-up of the voice recording.

All passages in the results section are tagged by interviewee, we later sent the interviewees the thesis with their passages highlighted. The interviewees could comment on wording and our interpretation of the results. This way the results have been validated as well.

6.5 MODEL REDESIGN

We then reflect which Data Management aspects could be used to assist the key initiatives. The key capability areas of the Data Management Framework are mapped to what we found in the interviews. We determine how the capability aligns with the processes, point out which capabilities are essential for IT and which capabilities are not in scope of IT. If results do not fit within the capabilities in the Data Management Framework, we determine if it could be included in a data management framework for IT. Results that should not be included in a capability model, are identified as boundary conditions for the implementation of Data Management in IT.

The goal for Fortran is to determine what should be the priority on Data Management within IT. We therefore indicate the importance per capability and determine a degree of maturity.

6.6 VALIDATION

The output of the case study is validated with a six experts that did not participate in the case study before. We chose two experts that are in a role in which they have overview of the DevOps & CICD transformation. These expert interviews are mainly used to validate the data challenges within IT. The validation rounds with three data consultants from the organisation wide Data Management departments are used to validate a redesigned data management framework for IT and the data management roadmap for IT. The interviewee from the IT Consultancy Office has expertise in both. The experts interviewed for the validation phase can be found in Table 6.3. The interviews lasted about 1 hour each. Three of the interviews were voice recorded and short notes were taken during each interview by the researcher. These documents are used as a reference to adapt what we proposed before and as a reference for constructing the validation chapter.

KEY	BUSINESS UNIT	ROLE
RC1	IT Consultancy Office	Data Management Consultant IT
RC2	CTO	DevOps & CICD Expert
RC3	CTO	DevOps Expert
RC4	FDO	Corporate Data Consultant
RC5	FDO	Corporate Data Consultant
RC6	FDO	Corporate Data Consultant

Table 6.3: Validation participants

7

CASE RESULTS

We first describe the opportunities and data challenges that come with the change initiatives within IT. The sections after describe the results grouped by the value stream. The chapter concludes with a discussion in which we provide identified data challenges.

7.1 CHANGE INITIATIVES

This section describes the results about the main projects that are executed within IT, namely the transition to a DevOps way of working and the implementation of Continuous Integration and Continuous Delivery. We also describe the introduction of Public Cloud and the changed relation with vendors, these first two projects impact the other two projects.

7.1.1 *DevOps*

One interviewee explained that responsibilities shift for the teams with DevOps, the teams are enabled by the bank to deal with those responsibilities. Teams need to be able to do activities themselves after the transition. They for example need to be capable of checking whether an application is up and running, they need to have the right toolsets and need the right data to be able to get grip on those responsibilities. They also need to be able to register incidents in the right format, and they need to be capable to solve the incident.

Instead of 'throwing issues over the wall' to IT Operations, the departments are integrated. The integrated teams are responsible for development, as well as operations. This change makes feedback loops a lot smaller since the teams have to solve their own problems. This change also decreases the distance from the business and shortens the development cycle, which in return gives the teams better grip on what they are creating. It is expected that this end-to-end responsibility will increase the quality of the code.

An interviewee indicated that there are teams that might not be able to handle the new responsibilities yet. Some teams make advanced delivery pipelines, but some are not able to cope with the new responsibilities yet. A reason for this might be that they would prefer to stick their traditional way of working, or just because they do not have the skills yet.

The teams need more skills because of the changed team composition and changed duties. An interviewee indicated that not every

team is ready for this yet and they are not eager to change those responsibilities.

Another interviewee in the Detect to Correct value stream indicated that a downside of the DevOps transition is that the responsibility comes down to a small team. A team may regularly make the same mistake when registering incidents, it could be the case that an incident may have happened before in another team, but it cannot be related to each other incidents because of the poor data quality of the registration. For this reason, higher-level relationships between an incident and the technological causes of similar incidents cannot be created.

There are different interests between the development and operations departments. While development wants to release new functionality as fast as possible, operations actually would not want anything to change, to ensure that the production environment can keep running stably. The data from operations comes from the production environment, while the data from development comes from the development process. Quality of code comes together with running performance and the value for the customer. Insight in this data needs to be created.

There is an extend of data teams should look at, to measure how they are performing. At a strategic level, data is used to look at the velocity of the output of the teams, but also at the composition of the teams (what skills and experience team members have), and technical debt (if they are working on legacy software).

7.1.2 *Continuous Integration and Continuous Delivery*

STANDARDISING Standardised processes and standardised tooling are provided to the IT team to connect the different parts of the IT Value Chain. This way data about the software development and monitoring process can automatically be extracted from the tooling.

Once this is in place, it is possible to retrieve metrics from the tools. This way teams can for example gain insight in their performance, but it is also possible to get a better grip on the change initiatives if teams do their administration via standard tooling.

Other benefits of the standardisation include:
For teams

1. The teams do not have to maintain their tools themselves
2. The tools always have the latest updates
3. It is cheaper (because it is mainly open source)
4. Easier to retrieve metrics

For managers

1. More grip on control

2. More cost-efficient
3. Easier to manage IT landscape

A lot of teams did not implement a standard tooling pipeline yet, therefore metrics such as code quality and lead time are not generated automatically. An interviewee thinks that the manually given data of those teams is therefore not always reliable. The two interviewed DevOps engineers explained that many errors are made in manual data. This is why there is now focus on automating the value chain and retrieving metrics from standard tooling pipeline.

Pipelines are chains of tools and are different pipelines for different technologies. For example, Java or Cloud.

A reason however, is that the teams do not implement a standard pipeline because they say they are not like the rest of the teams and need tools that are not provided in the standard. A lot of teams resist this standardisation and come up with excuses, such as "I have an Apple computer, which does not support it".

The Enablement Team of the bank is pushing to standardisation, while the DevOps teams are pushing to customisation. The challenge here is to find an optimum in between, to provide the teams with enough flexibility, but still have enough benefits from the standardisation.

A DevOps CICD engineer also explained that not everyone should perform CICD. He advised not to do it, if it is not necessary, it could be too much work for the value it can deliver.

METRICS The engineer, who is responsible for providing teams with metrics, explained that the bank is already quite mature in retrieving metrics from the available tooling. Instead of relying on what someone has reported manually about what happened, data can be retrieved from the tooling. The engineer explained that many requests to retrieve data from tooling can be fulfilled already. However explained that there are lots of data that can be used, but is not used yet. It might therefore be advisable to register what data can be retrieved and point out where the data can be retrieved from.

The same DevOps CICD developer explained that much data has been gathered in the application designated for raw data storage (Splunk). His team that gathers the metrics of the teams know a lot about what is available in the data on the platform and know a lot about how things happen in other teams. But he explained that many teams do not know that it is available, which is a waste. He advised that the metadata should be shared, which explain what data sources are available in the data storage and what can be done with the data sources. People also do not know how Splunk works, so that needs to be taught for teams to be able to act on their data.

7.1.3 *Public Cloud Transition*

An interviewee explained that all activities that are performed on the public cloud is done by the teams within the bank itself, not by external parties.

The introduction of Public Cloud opens up opportunities to generate more data about the infrastructure. The teams get insight in the hosting data, which was handled before by the main vendor. The data and extracted information can in turn be used to get insight in the performance of the public cloud environment.

The public cloud can provide data such as:

1. How well the application is running;
2. Where the application is running;
3. What the application costs to run;
4. If an application is being used or not;
5. How much resources are used.

The cloud solutions are designed for giving IT flexibility, while on-premise services are quite cumbersome. Scaling up infrastructure is easy on the public cloud. Teams can manage this infrastructure easily with the use of infra-as-code. It is wanted that this can be done automatically in the future.

Infrastructure as Code is used to standardise configurations to deliver consistent and stable environments at scale on the public cloud [27]

An interviewee explained that another benefit of the cloud is that events on the running applications can automatically be extracted from the environment. In turn, actions automatically be performed to fix the issues reported in those events with the use of this data.

Another interviewee indicated that a large part of the operational layer of IT will be removed, maintenance that was done for on-premise servers will be handled by the public cloud services. Since the DevOps teams will gain the responsibility, it becomes clearer where the costs go.

The switch to public cloud also has impact on service processes. An interviewee explained that people are now used to call the helpdesk if a problem occurs with their application. But the infrastructure is not managed on premise anymore, and thus the helpdesk cannot provide the client with help.

7.1.4 *Off-shoring and Vendor Relations*

An interviewee explained that the bank has always outsourced their *Service Management* process to IT vendors. The main vendor is responsible for the process for 12-13 years already. It has the responsibility to let all the bank's IT vendors work together.

An interviewee indicated that because operational tasks are the responsibility of the vendor, they are also the owner of the data. Since the data could not be accessed, no operational data was used. Only questions on tactical level were asked, but none on operational level. The ownership of the data is transferred from the vendor to the teams with the DevOps transition.

Some critical components are still maintained by the vendors, these parts will be brought to the teams as well. In these cases it is important that the correct monitoring systems are placed.

Teams have to manage their own IT operations with the transition to DevOps. An interviewee mentioned that some teams have difficulty to deal the new responsibilities. Instead of the mentality “Someone else will do it", the teams now will have to do it themselves. The teams need more maturity to manage the new responsibilities, such that the applications can keep running. This is why there is focus on a maturity model, which is used to identify if a team is able to run an application, to monitor the application and to do continuous integration and delivery. This way responsibility can be given to teams in a controlled manner.

7.2 STRATEGY TO PORTFOLIO

The first value stream in the IT Value chain is Strategy to Portfolio (Figure 2.2). In this phase the wishes, requirements, limitations and everything in that perspective is determined before IT services are being developed.

The IT landscape needs to be designed to make sure that IT adds value to the organisation. Processes need to be described, for example descriptions of data usage need to be made and it needs to be made sure that the organisation chart is clearly defined.

There are numerous types of data in the Strategy to Portfolio value stream. For example designs (how the solution will look like eventually), requirements (constraints the solution will need to comply with) and architecture (dependencies, standards, and principles). These types of data describe what the IT landscape looks like. The wishes from a business perspective are logged in the Strategy to Portfolio stream, the architecture of applications is also created in this stream.

7.2.1 *Challenges*

7.2.1.1 *Data Quality*

An interviewee explained that data feels close to technology and it would therefore be logical that ‘at least IT will have their data in order’ within the bank. This however could be improved, the initiative lead by the central Data Management department to improve data quality

did not yield much-expected reports in IT over the years, the given process was not followed. He indicated that just a few data quality issues were reported.

7.2.1.2 *IT Landscape design*

A DevOps consultant indicated that it is not always clear what the exact IT organisation looks like. He explained that it is for example unknown how many IT teams there truly are. There are multiple sources of truth, different people provide different answers. An interviewee indicated that the designated point for registration contains a lot of teams with only one member, this is an example of a data quality issue that is not solved. He further explained that data is however important, else decisions are made based on assumptions. There should be one tool that is the single source of truth, people should keep the data up-to-date there. At the moment the data needs to be entered in multiple places, which is confusing. Timeliness of the data is also important, there is no real-time overview, a source of a couple months ago is already outdated.

7.2.1.3 *Inconsistent Data Models*

A business architect explained that it is often unknown what the relation between an interface and an application is, and therefore unknown what the data is that passes through the applications. It is important to know these relationships, to explain to a regulator which interfaces are connected to an application for example. Being able to connect data sources across the IT value chain with each other (*Traceability*) is also important to explain how money is spent in IT (Cost Insight). The data is essential to be able to identify how IT costs could be reduced (Cost Management).

The architect indicated however, that it is a pitfall that the model which describes the relationships between IT assets will become too large and cannot be maintained. This happened once before, a vendor created an information framework for the bank. The model looked perfect at glance, but it would only work if the whole organisation would implement the model literally. It turned out that the terminology could not be accepted in the organisation and that it did not fully fit in all departments' contexts.

7.2.1.4 *Descriptions of IT Products*

An interviewee, that works at a department for product planning and design, explained the data needed from others is not always fit for the purpose. He indicated that people create data about IT products (such as applications and software) from their limited point of view, and therefore a data source which is said to be a description of an IT product, might be different from a data source about the same IT

product, but created by someone with a different perspective within the organisation. Another problem is that it is not clear how trustworthy and how up-to-date the data source is. Also, it is hard to find the person who can explain which regulations apply to a certain application which is being worked on. It can be the case that if someone searches for data, there could be a lot of different explanations about what the product actually is, and who is responsible for the product.

He explained that there are too many versions of a data source is because there is no control over who may and who may not provide a description of an IT product. A person is allowed to describe a product from another department, this description is seen as reliable as the description by the actual owner of the product. For this reason, it is not clear for the user what the correct description would be.

The challenge is to provide controls about the creation of descriptions of IT products, as well as provide clear metadata of the data source.

The interviewee advised to let people only provide descriptions of their own products, and let someone who needs to create a higher level design of an IT product, bring those specialised descriptions together, instead of creating an own description. This also needs a change in culture, such that this division of responsibilities is accepted.

There is an initiative within the bank to standardise definitions of IT products and their composition, such that a common understanding within IT would be created. This is done with the use of a common data model.

7.2.1.5 *Design tooling*

A lot of data is created in the design process, but this happens in a lot of different ways. Those designs are created with tools such as Office Powerpoint but also with Excel, or a sketch is created with a design tool, for example. An interviewee explained that it is often unclear how these sources relate to each other and a central point of view. This indicates that there is not enough standardisation in this process.

7.3 REQUIREMENT TO DEPLOY

The emphasis of this stream lies in the development of IT services.

7.3.1 *Value of Data*

Data about IT performance can be extracted from the tooling pipeline that CICD offers. These metrics can give more insight into IT performance since planning can be linked with development metrics, which also can be linked with operational metrics. These metrics are given

back to the DevOps teams and IT management. Combining these data sources could provide novel insights to drive decision making.

7.3.1.1 *Value for teams*

An interviewee indicated that teams can steer based on facts with the metrics. Teams can use it to get insight into how they spend their time, this way they can identify their problems and gain insight on how they can improve their way of working.

Teams receive numerous insights about their development process. They, for example, get to see how good the quality of their code is, how much work they have performed during a period or how long it took to repair an issue.

A DevOps transition manager found it important that not only managers get to see data to get a grip on the overall IT performance, but also the teams get to see their metrics themselves. This data is valuable for the teams such that they see the importance of the work they do, instead of that a manager tells them that something should happen.

A DevOps CICD engineer explained that the correlation between factors is the most interesting to see for the teams, to identify improvements in their process. Another engineer added that it is most interesting for teams to measure their performance over time. Else you are measuring a point in time, which does not say much, as it cannot be told if you have improved or what the cause of an issue might be.

7.3.1.2 *Value for Management*

Managers want to steer on the data that comes from the CICD pipelines, to gain insight into how teams perform. The managers want to know why some teams perform better than others, data gives insight into this. An interviewee explained a use case of the data with the example: “If a team does not deploy anything the whole year, then you can start wondering what the team did throughout the past year."

Another use case of IT Value chain monitoring for managers is to compare the performance of teams with each other, to recognise why one team performs better than another. Low performance does however not necessarily has to be caused by the quality of the developers but can have other influences, such as personal situations or distractions in a team.

An interviewee from IT Service Management explained that an operations employee might be cheaper than a software developer, but if it is known that a recurring issue is being fixed on regular basis by an operations employee, it might be more cost-effective in the long term to fix the root cause of the problem by a software developer.

7.3.1.3 *Cost insight*

The metrics from the CICD pipeline provide data about every stage of the development process. This way it can become clear what a change has cost and what a similar change will cost in the future. Instead of steering on portfolio- and project plans, there will directly be an overview of how much a feature will cost. There were no total image of IT development and IT operations costs, by combining these with DevOps and CICD you do have that overview, and you can see from the beginning to the end what an application has cost and a similar application will cost.

7.3.2 *Challenges*

IT LANDSCAPE REGISTRATION An interviewee indicated that there are many different lists about what the IT landscape looks like. For example how many IT teams there are or how many applications are available in the IT landscape. This is confusing since it is difficult to find out what truth is. He explained that the detection of the IT landscape could be automated. A downside with automation however, is that employees whose task it is to maintain those lists will become unnecessary.

DEVELOPMENT REGISTRATION Two interviewees explained that some things need to be registered about IT applications, such as which change has been made or who have been provided access to an application. Another interviewee indicated that the data which requires manual input can be very valuable. Analysis and insights enable to improve strategically, as well as operational. He explained that this is generally known in the organisation. The problem, however, is that if the data that is entered manually is bad, the output is also bad. It requires a lot of effort to fix the data. The tendency is to fix these problems afterwards, but the interviewee told that this enables people to keep entering the wrong input. The interviewee thought that laziness is not the cause, but it is a capacity problem. The priority of work should be looked at.

One interviewee indicated that when it is unknown what has changed, it might be unknown if security checks are passed. In turn, it cannot be explained with full certainty to the regulatory supervisor if all security checks are performed.

The interviewee indicated that it however is questionable if the teams see the importance of the registration of processes and administration. This is why the DevOps program also gives the teams awareness about this topic.

Another interviewee however indicated that for example granted access rights are already registered in the application's code, which could be used instead. The registration process would thus be an

unnecessary and time-consuming process. Two interviewees also explained that changes in code need to be described twice at the moment, once when the code is being created, and afterwards in another tool for administrative purposes. The data is available in applications but is not used yet. For some applications, teams are already helped with the automatic administration of changes in code.

It is however problematic if the quality of the registration is not done correctly. An interviewee gave several examples:

1. If it is unknown that new incidents arise because new code has been released, then you will not know what should be done differently next time;
2. Incidents remain unsolved longer;
3. Applications are less available;
4. Unsatisfied customers;
5. Wrong use of resources;
6. Decreased effectivity of the organisation.

This is why it is necessary to tackle data quality issues with registration.

7.4 DETECT TO CORRECT

This last value stream of the IT Value chain focuses on the monitoring and maintenance of the running IT services. Data is used for various purposes in this value stream. Such as incident resolving or the monitoring of running software. The automation of the IT Value chain was one of the most named opportunities for data usage. The interviewees mentioned numerous benefits of creating an automated pipeline.

7.4.1 *Value of data*

7.4.1.1 *Incident and Problem Resolving*

The department Incident, Problem and Change (IPC) Management is responsible to act in crises, in which an incident should be solved.

One interviewee explained that incidents are created when something goes wrong with a system, for example, if a running application breaks in the production environment. Those incidents are registered and solved manually, based on error reports. Two interviewees explained that most incidents are reported via phone calls by customers, the call centre of the bank then creates the incident manually. This cannot be automatised yet.

The interviewee described that during a crisis situation, something happens that did not happen before. At the moment a lot of experts

come together and explain their vision on what happened. Facts must arise, and assumptions can be verified or discarded, to find the root cause of the crisis. Machine data could be used to objectify the situation, such that the current state of the systems can be compared with the stable situation just before the crisis situation started.

Although manually reported incidents cannot be resolved fully automatically, most of error handling of systems could be automated, such that systems generate errors automatically, which can automatically create incident reports, which in turn can be sent to solving parties, who could solve it automatically with their systems. This would benefit departments like the Service Management since many of the system's errors can be solved with simple automatic actions.

7.4.1.2 *Anomaly detection and automatic remediation*

Anomaly detection could be used to prevent problems with the systems. At the moment, with event management, a system creates an error after something went wrong. A combination of incident data can be used for Problem Management, which focuses on tackling the root cause of problems. The IPC department needs a clear overview of the IT landscape and its data components, to quickly identify the root cause of the incident. An interviewee of IPC believed that analytics on machine data could be used to extract facts about hardware and applications. Tools to interpret the data will help determine the deeper technical cause of an incident. The machine data could be correlated with the incidents that are reported manually, which might link types of incidents to what happened in the system at the same moment. That level of analytics is not performed yet in the organisation.

Once the software is in production, it is interesting to see how it is performing. Data from monitoring tools can be used to detect anomalies in the performance. Three interviewees explained that exception-based information would be the most interesting since normal behaviour is not worth looking at. First, a baseline of normal system situation should be created based on the data. As a next step, the system could provide real-time anomaly detection (e.g. the CPU load increased with 10% in the last hour and might increase even more). This could be used to prevent an incident since the cause of the problem is detected before the systems break. Some teams already make use of such information, they have proven to be successful in preventing a problem, or to solve it quickly. Data about the systems could also provide insight if the Service Level Agreement requirements will be met over the long-term.

Some applications have peak periods in which the load is very high, one of the payment applications for example. If something goes wrong, you want to know what the reason is. The cost of downtime can be very high in some cases, it may lead to reputational damage as well as to direct losses.

An interviewee explained that it would be necessary to bring data together from multiple layers of IT to get the full picture. This could be used to get an end-to-end overview of what changes could have led to the incident. He saw an opportunity for automatic remediation on anomaly detection, in which problems are solved automatically. Because in the end, you want the developer to be able to create new functionality, so more automation the better. Another example of the value of an integrated IT pipeline is synthetic monitoring. Which can be used to provides a link with how an end-user experiences the performance of an application. If it is known how many users are using an application, priorities can be set on which issue should be fixed first, an application with 1000 users might have a higher priority than one with 10 users.

VIEW OF A DEVELOPER One of the goals of monitoring the IT value chain is to be able to predict the impact of a change in code. A software engineer thinks it is unfeasible to determine the true impact of a change, no matter how many people look at it. The engineer explained that there are too many variables in play that it is almost impossible to determine.

7.4.1.3 *Performance Management*

Once the software is deployed, data can be generated about the performance of how it runs on the infrastructure, as well as about how users interact with the application. Metrics that were mentioned in the interviews included:

1. Response time of a website
2. Availability of a website
3. Number of users of a website

Different metrics are used as well for *Performance Management*, for example, how many times a system crashes, how long it takes to get it up and running again, time to release or how many releases are done.

OTHER VALUE GENERATING ASPECTS

1. Monitoring more than just IT incidents, also monitor financial transaction data. The data is there but is left unused since it is handled as purely IT data.
2. Linking system performance to customer retention.
3. Linking interface layout to customer retention. (e.g. Android users click more often to a next screen than is done on PC)

7.4.2 *Challenges*

7.4.2.1 *Incident registration*

Another interviewee explained that there is a lot more human work needed to solve incidents after the incident is reported. The department responsible for solving incidents wants to know what the root cause of the problem was, which people have been contacted and what went wrong. In many cases, this goes well, but in many other cases people fill in dots to skip through the form.

Two interviewees both explained that the cause of that engineers do not fill in the registration form might be because they do not find it interesting or do not see the value of doing it for themselves. An interviewee indicated that the engineers feel it is overhead, they could help someone else instead of spending time on the registration of the incident. Teams are not aware of these possibilities. This leads to decreased Data Quality since they do not put in the effort. The interviewee explained that if there is no insight into how much time a team spends on solving incidents and a major part of the time is spent on solving issues, it might be because the code is not good enough. Which may result in incidents every deployment. It also decreases the velocity of development, in practice, the teams are mostly fixing issues. The management of IPC, who uses the data, on the other hand, could improve the understanding by making better use of the data and by showing the results back to the organisation according to the interviewee.

He continued to explain that it would be useful for the IPC team to perform data analytics on the quality of the incident related to team metrics, to gain insight on what factors may influence incorrect settlement of an incident. It could be the education or work instruction for example. This data analytics cannot be performed yet because the data is not available, the interpretation is still difficult and the skills to work with the tooling is not present in the IPC team.

If the data quality of the data is not in order for IPC, it is harder to pinpoint where a change was made that could be the cause of the incident. If the IT landscape looks different than thought, you are looking at the wrong systems and will cost money because it takes longer to find what you are looking for.

It was also explained that people who have skills to create something useful with the data from incidents and the tool pipeline are very scarce. Not only technical skills are needed to transform the data, but those people also need to have an understanding of the solution that is wanted by the business.

An interviewee from the IPC team noted that one of the biggest challenges in working with the data is to create a connection with the people who have the problem. Operational data always has a connection with the organisation, you have to speak with teams to

relate what happened in practice. This does not happen enough yet within IT. The challenge here is to understand each other since the languages spoken are different.

The bridge between the engineers and the business is hard to make. The adoption of the tools and processes that are offered to the engineers to do their work better is hard. The interviewee indicated that it goes too slow at the bank. It is not about the tools itself, but people still work as they did 10 years ago sometimes. The adoption has to do with curiosity to do it better and faster of the employees, but might as well be because of the focus on the craftsmanship of individuals.

Other issues with data are that people do not follow the given process and therefore are not aware that they are using obsolete data.

7.4.2.2 *Data Awareness*

An interviewee from the Service Management department indicated that it is insufficiently known what Data Sources there are. He also indicated that raising awareness about data is quite abstract. It might work to show the value of data as a way to make people aware of the possibilities with it, besides raising awareness of the availability of data. It was also indicated that it is often unknown who the owner of the data is. An interviewee indicated that this is for example the case with Service Management data.

Another interviewee indicated there need to be guidelines for such administration. If it is unclear what data is registered where, then people start to input all different sorts of entries in the application. This leads to inconsistency in the data about the IT landscape. The interviewee explained that if you cannot give clear guidelines, then you cannot expect the organisation to do the right thing. These guidelines are needed since the organisation does not want to create central teams that keep the registration for everyone but want to provide self-servicing.

7.4.2.3 *Common Source Management*

One interviewee was largely involved in enabling the IT value chain with data integration. He explained that there is a product administration, in which the organisation records all products that are delivered to the business parties. The relations between application components are registered, who have used them, who is the technical administrator and the operational operator. The rest of the organisation is enabled with tools to perform Service Management.

It was however noted by numerous interviewees that often multiple versions of a data source exist. An interviewee described that this may be because the introduction of a new initiative is often combined with an initiative to gather data, while the data is already available somewhere. He also mentioned that a Common Source is not used

because it does not fully comply with the needs of the user, and thus an own source is created instead. The importance of a central data source was stressed, the quality of the data should be improved in this source, not in clones. Empowering the organisation to connect to these Common Sources is a work in progress for IT.

7.4.2.4 *Data Distribution*

An interviewee indicated that their department makes use of a Service Management tool to distribute data, but it is not clear why it should happen. No document describes what is done with the data once shared, if it is mandatory to do so, and what it will be used for. Sharing data in a central place is sometimes felt as just creating a large data lake, but the goal of what to do with it is unclear. It is not monitored if the data in a data lake is useful and is thus worth the costs. An interviewee mentioned that there should be a club that should drive these types of insights.

REGISTRATION OF INFRASTRUCTURE The IT landscape has become very complex according to an interviewee of the Service Management department. He explained that there is about 50 years of automation behind it. An example is a CRM system, which has gained 100 interfaces over the years. The systems emerged and lack constraints imposed by prior work, this makes it unclear. Managing data can be complex, often there is no overview of the links between systems. It is therefore a large effort to find out which systems behind an interface the data goes to. Another interviewee of the IT Service Management department explained that many applications have dependencies on other applications. If a change is made in one application, something might break in another. Therefore approval of a change is necessary for the dependent applications. If applications are more disconnected from each other, then approval might not be necessary anymore.

Another interviewee from the same department indicated that the common language between business and IT is also important in the registration of infrastructure to know which applications are in practice available in the IT landscape. (See Section 7.6.1). He explained that it for example is important that the status of applications is known for the business. Statuses include 'concept', 'allocated', 'in use', 'to be decommissioned' and 'decommissioned'. It is very important that this administration reflects the true situation, as it imposes risks concerning management (the business thinks an application is allocated, while it is already in production) or security (an application is running in the infrastructure, while the administration tells it is decommissioned).

This administration is done manually, there is an initiative to be able to discover the status of an application based on what is running on the infrastructure. This applies the other way around as well, a

ticket for the teams that manage the infrastructure can be created automatically if it is required to change the state of an application.

One of the uses of the system which is used for the registration of IT assets, components and relations to applications, is to compare what has been ordered and what has been delivered by the vendor. After a cleanup and automation with the use of new tools, the register's quality had improved. This used to be bad since there were too many loosely coupled Access databases.

7.5 SUPPORTING CAPABILITIES

7.5.1 *Grip on Change*

7.5.1.1 *Defining processes*

According to an interviewee, DevOps can be seen as the next step after Agile, it fundamentally changes how teams work. The Agile transformation introduced more flexibility and freedom for teams. The interviewee indicated that this leads to a sort of 'Free for All', which lead to the introduction of many divergent processes across teams. The Apollo program, and the DevOps transition, in particular, is now also used to introduce more standardisation again. Another interviewee explained that teams should start to describe their process, which includes the resources and information the team needs to execute their processes. An example is backlog management. Control mechanisms are defined to manage risks, to get grip on the risks, the teams are expected to register several fields about their way of working. For example, the tools they are using, their team composition or what the responsibilities of the team are. The teams generally feel it is extra work, and are not eager to fill out those lists regularly. There is no discipline to do administration manually, an interviewee indicated that it might happen twice when you ask the teams for it, but then it stops. The interviewee explained that this is why there is a focus on the automation of this process. But some things will still need manual input from the teams.

7.5.1.2 *Measuring the transformation*

Measuring the adoption of change initiatives is important. An interviewee described that the bank is good at identifying new things, but spreading it out such that everyone is benefiting from it, is a journey of itself. If the adoption is not measured, people may not be adopting it. The result is that the organisation is doing better and better services are created, but nobody uses the services.

There are no metrics yet to gain insight into the Agile transformation. A team within Fortran is working on this. The steps which are taken right now is to define which management information can be retrieved

from which tools and to give a standard definition of a set of KPIs. These KPIs are monitored to get grip on DevOps. An interviewee described that the DevOps Enablement team also created a list of requirements for teams going in the DevOps transition. The checklist contains data about the maturity level of the teams.

A DevOps CICD developer explained that the bank is very progressive in sense of implementing CICD, but is less progressive in the feedback. He thought it is sometimes unclear why projects such as CICD are done. It is unclear if a team will perform more releases of their product than before CICD. Automation of the pipeline will help retrieve this insight. He explained that it is wanted to measure transformations in the organisation. It should for example be measured if the teams actually work fast with CICD and DevOps. This is not known if it is not measured. The data creates a vision of what you want to build. He indicated that 450 teams and feelings about their performance do not work to measure how well the organisation is performing. Thus other metrics are needed, which could be retrieved from the CICD pipeline.

7.5.1.3 *Measuring productivity*

Improving IT is often about productivity, lower time to market and other metrics on velocity. An interviewee indicated that some sort of proxies of productivity were introduced. Much of it is based on Story Points, which is not uniform since each team has a different estimate on how much effort a piece of work will cost. Indicators regarding productivity would be really interesting to understand IT even more according to the interviewee.

Story Points are numeric estimates on how much a piece of work will cost [4]

There used to be metrics to measure productivity, for example with the use of function points, but it did not work out.

7.5.1.4 *Measuring value of IT*

An interviewee explained that there is a focus on making processes more efficient, which should result in a more efficient and cheaper IT factory. At the moment there is not enough insight into the costs of IT (besides the hourly wages of employees). More could be done to measure the costs, but also more could be done to measure the added value of IT. But two interviewees explained that at the moment it is very hard to measure what the added value of IT actually is. Timing is one of the difficulty factors, the inputs come at some point, the benefits come a bit later, which could be three years later according to a manager. With Agile there is a focus on slicing things down. *Minimum viable products* and quick releases shorten the time of delivering, which shortens the time to measure value.

The interviewee explained that people have to start thinking about the outcomes, and about the adoption of the solutions as well. If people

think about these aspects, the deliverable becomes stronger, because you have thought about it. This reinforces end to end responsibility, sometimes someone can deliver something that is asked for, but in the end does not know if it is going to work.

7.5.2 *Identity & Access Management*

The Identity & Access Management (IAM) within the bank is responsible for providing digital identities for customers as well as employees of the bank. Other responsibilities of the department are to provide access rights for those two groups and to provide credentials for IT systems.

Data is a very important asset for IAM, they store who has accessed a system when, how long they had access and when they returned the access token.

ACCESS TOKENS An interviewee of the department explained that data for IAM could also be used to find out how long tokens are used. Sometimes an employee requests an access token for a day, but only uses it for 15 minutes. This data can be found as machine data in the systems but is not used at the moment.

IDENTITY AND ACCESS MANAGEMENT An interviewee found that the usage of the data-sharing platform for IT is quite low. Instead of sharing raw data, interpretations of the IT data could be shared with the rest of the organisation. Several sources of the IAM department could be placed in this data platform, such as Digital Identities, an overview of access rights, a list of vaults (which contain passwords) or access tokens.

7.6 OTHER IT ASPECTS

7.6.1 *Data Model*

One of the goals of CICD is to connect tools within IT with each other such that there is a pipeline of tools that can work automatically. One of the IT challenges however is that the data models of the different tools are not compatible, according to numerous interviewees. Data models describe which data elements the tools deliver. This needs to be solved first, before value can be created. An interviewee explained that teams within the organisation already work with the data models for a long time already. The connection between the models was however not created before. He explained that it might be because the teams just did not see the opportunity to create the connection, or that the tools at the time were not mature enough to support it.

Another issue with the data model is that departments within the IT Value chain describe applications and data components using different terminology.

An interviewee explained that traceability goes from application to infrastructure, or from user story up to development. This is only possible if data elements can be linked with each other, which can only be done if everyone registers their data elements in the same way. Two interviewees from the Service Management department explained that different teams may have a different definition of data elements and a different view of what the term *application* means, this might be because they have another vision of how it should be used. It is a challenge to bring the definition of lots of different teams together. One of the interviewees explained that this is not the case anymore for infrastructure. If you ask what a server is, then you will get a consistent answer. He also contributed the relatively low maturity of conceptual thinking on application level as one of the reasons that there is difficulty with traceability.

A business architect explained that a lot of metadata is needed to accompany a data source. It has to be agreed on what the definition of an application, a service or a product is. There is a mismatch in what the business and IT mean with certain terms, for example a business product may for the business mean an 'Insurance-linked mortgage', while others within IT might see a piece of software used by the business as a product. It is important that the differences are made insightful.

The bank provides a tool in which all available applications should be registered. The bank also provides a lexicon in which the definitions of data elements need to be registered.

One of the interviewees explained that the business wants to know which applications are most important, and which are not. They for example might want to decommission an application. Everyone should be talking about the same model in order to be able to actually phase out the whole application.

A DevOps Engineer also explained that it is needed to classify what a component, or type is. You want to identify those components with numbers, such that developers can order what they need for their applications in a catalogue.

The problem however is that because of the different professional languages across departments, it is very difficult to let the different parties within the organisation talk the same language about IT. This is reflected in the registration lists of applications which are already available within the bank. One list has very different contents than the other list, which make the lists incompatible.

On one hand an interviewee indicated that it is not advisable to create one data model that describes the IT landscape. This interviewee indicated that dialects should be allowed, it is more important to understand each other, instead of forcing everyone to speak the same language. While another interviewee found this common model and language a necessity, since the bank is making progress to include

business and IT in one team. This starts with the DevOps movement, and could later be extended to concepts such as *BizDevOps*.

7.6.2 *Cleaning the IT Landscape*

An interviewee described that IT is built on a wrong basis, there are many factors which limit a smooth transition. Besides the DevOps movement it was noticed that the current process management does not fit an Agile way of working. This is why the program managers of Apollo need to know what constitutes a good team composition and how processes should be structured.

7.6.3 *Language gap*

An interviewee indicated that there is still a language gap between a business owner and the IT teams within a same department. A good translation is needed for the business owner in order to understand how the department performs. This starts with awareness and education.

7.6.4 *Lack of Skilled Personnel*

It was mentioned that the country does not have enough Software Developers for what is needed by its companies. There is a cost-association with outsourcing, but another reason companies have to outsource their work is because they are competing over the same pool of workforce in the country. This is why a large part of the bank's software development is outsourced to India, and many developers who work in the country come from India.

7.6.5 *Culture*

7.6.5.1 *Decision Making*

Numerous interviewees explained that people determine what should happen before looking at data, they decide on what they already know. Reasons could for example be that teams are not aware of what the value of data is or do not have the maturity to use the data. Other reasons include that people do not know what to do with data, it might feel confronting, or it might also be seen as boring to just make decisions based on data.

Another manager later explained the data could also be used to challenge their biases with the use of data.

7.6.5.2 *Trust in Data*

On one side there is a lot of enthusiasm to do more with data, but on the other side there is an opposing force. One of the interviewees said this is organisational culture. He thought that if people see the value then they start trusting it. One of the other reasons named was that it is too easy to blame an external factor why the data should not be correct. For example: *"The data model does not align with what I need, so I cannot compare these two sources."*. The quality of the data is an important factor not to trust the data. Especially if the data has not been used that much, the quality would not be that good yet. If people see that the data is used, people are more diligent to ensuring the quality of the data.

An interviewee described that another reason is that the source did not contain a certain field. A solution is to start using Common Sources instead of creating new ones (see Section 7.4.2.3), as well as includes a cultural change, in which people learn to make decisions on data that is somewhat less complete.

7.6.5.3 *Trust in usage of data*

As described previously, there is a desire to gain data from all stages of the IT value chain, which includes performance tracking of teams. Trust in the usage of the performance data plays an important role.

The developers might be wary on what happens with the data that is collected about them, they might fear consequences if it is noticed by their managers that they perform less than their peers or as other teams. A DevOps CICD engineer explained that development teams can see the monitoring of performance through two lenses. Some see it purely as content for the managers, such that they can point teams out that under-perform. But on the other side the teams do understand that with DevOps principles more responsibility is gained, with this data they can gain this responsibility. The performance metrics have to be made attractive, else there is no adoption.

An interviewee indicated that as a consequence of the lack of trust the developers might even manipulate data if they know it could negatively impact them. He and the DevOps CICD engineer both explained that trust should come from both sides. The teams should have the right to have an opinion on performance measurement, and management should listen to their concerns. A transparent culture about how data is used is very important. The engineer also suggested that feedback should be positive too. The data should not be used to punish, but more for coaching the teams to tackle problems.

Another interviewee also indicated that people do not trust the solutions of others. An example is from the Retail department, they got presented a model which predicted when a customer would leave the bank. They already used one, which was a lot less accurate, but

they did not trust the new model, since they were accustomed to using their own.

7.6.5.4 *Showcasing the value of data*

An important factor in the effective transformation of data into value is to show people why they need to put effort in making the data available in good shape. An interviewee explained that once people find out their data is used, the discipline to follow up on it can be seen. He found it natural that when you find out nobody is using your document, then you just skip through a form to get done with it.

It is for a manager sometimes hard to identify data quality issues, since data is looked at from a very high level perspective, which makes it hard to see if data is right or wrong.

7.7 DISCUSSION

7.7.1 *Data Challenges*

Based on on the case study results we distinguished 33 situations in which problems with data play a main role, we call these situations *data challenges*. We listed those challenges in Appendix B. For each of those challenges the given cause as described by interviewees is given. We indicate what the solution might be, we provide this recommendation based on our insights from the case study, or if available based on the interviewee's answer. As a last part we provide which capability from the Data Management Framework could assist tackling the issue. We mapped the challenges according to the value streams they are most relevant to. The mapping is presented here as Figure 7.1. Challenges that are listed under *IT Supporting* describe challenges which span multiple value streams.

7.7.2 *The causes and opportunities of the data challenges*

Based on the challenges and further results from the case study we examined what the overarching reasons for better Data Management in IT would be and how the causes to the challenges are related to each other.

We found that IT is pressured to become more efficient and cost effective. IT is however struggling with a number of issues with regard to their IT landscape. The costs of IT are not insightful (19, 20, 21), the integration of IT services is difficult, data from different systems cannot be correlated and a lot of manual work has to be done.

The numbers between brackets refer to the challenges as described in Appendix B

We found that it should be a main priority of IT to be capable to relate most of the data sources within IT with each other, such that there is a complete and reliable overview of the IT Landscape

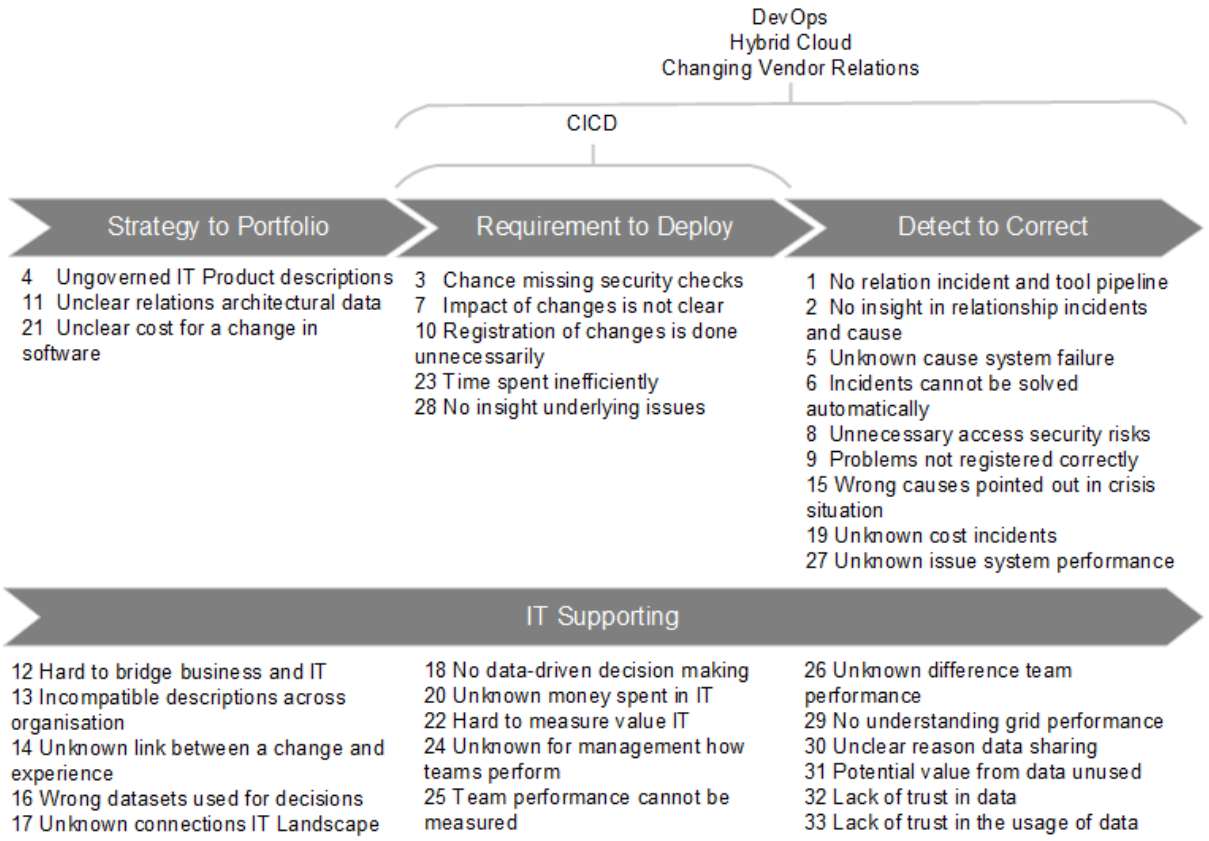

Figure 7.1: The data challenges mapped to the IT Value Chain

(infrastructure, applications and DevOps teams) and of the processes in the IT value chain. We present this concept as *Traceability*.

In the validation rounds (Chapter 9) we found that traceability could also be identified with the term *Auditable DevOps* and is related to *Data Lineage*. The latter however focuses on tracing the mutations done on a data source, while our definition of traceability focuses on the ability to create relationships between different data sources.

If traceability is not addressed, IT misses value generating opportunities, such as automatic remediation of incidents or finding a common denominator in recurring problems during development. Other risks include: Wrong priorities are set (23); Unnecessary work is done; Unnecessary security risks (8); Unknown relation between changes or system performance and customer experience (14); Unknown system behaviour before an incident (5); Wrong causes are pointed out during incidents (15); Organisational change projects cannot be measured effectively (24).

The creation of value with data is also subject to numerous challenges, such as inability to solve incidents automatically (6), inability to measure performance (25, 26 27), potential cases to turn data into value are unused (31) and a lack of trust in how data is used to create value (33).

A number of causes hinder the traceability of the IT landscape. We found that the challenges as listed in this section share higher level causes, as such we found two main challenges that need to be tackled. We present these relations and main challenges as a diagram, which is presented in Figure 7.2

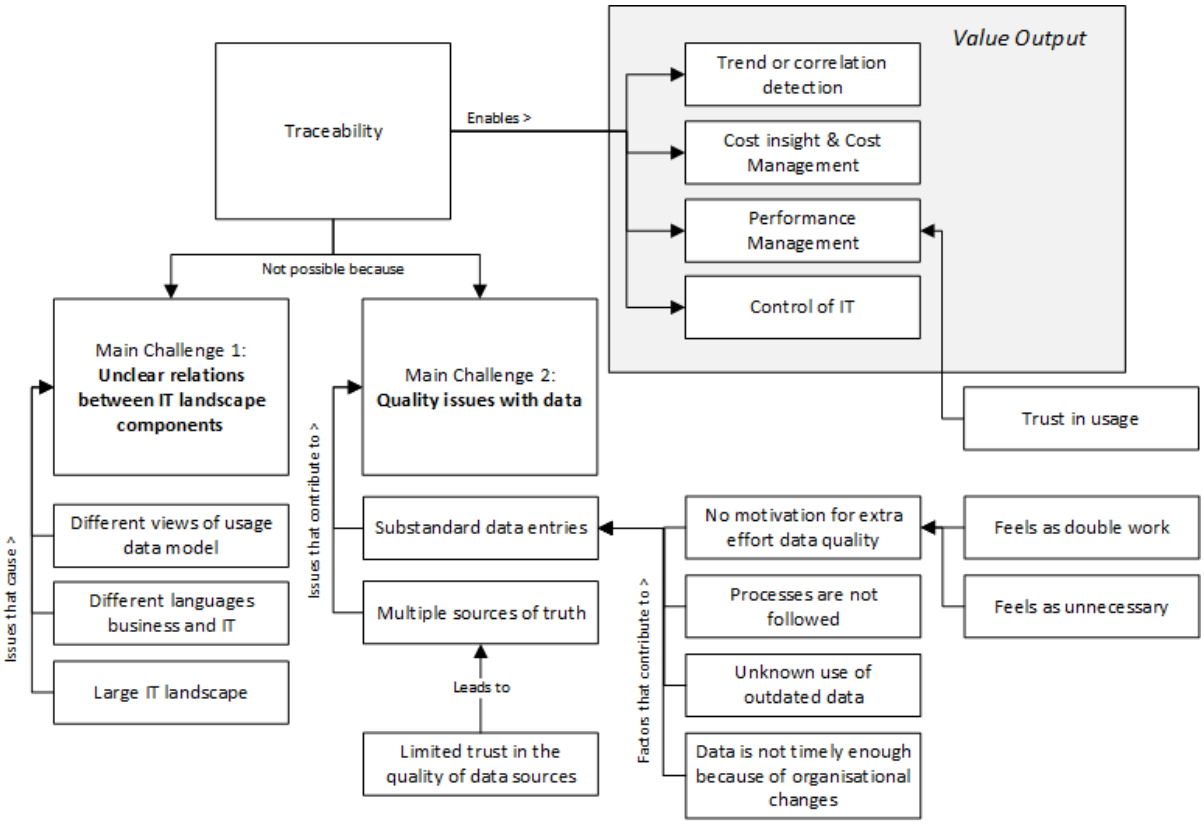

Figure 7.2: Relationships between data challenges

7.7.2.1 *Main Challenge 1: Unclear relations between IT landscape components*

Data models are architectural models which describe how the actual IT landscape looks like in sense of IT products and teams. A data model for example describes which applications have which functions, what data elements they provide, how applications are correlated to each other or how the applications are composed in smaller application components. The data models which are available are however not good at the moment. Challenges: 1, 4, 5, 7, 12, 13, 16, 17, 27, 29.

It is because of the following reasons: Different views of the usage of IT components in the data model and different languages between business and IT.

7.7.2.2 *Main Challenge 2: Quality issues with data*

There are numerous problems with regard to the registration of data. Manual registration needs to be done in various places. Forms for changes in code need to filled out, as well as for incidents. The registration of the IT Landscape also is done manually and so is the registration of given access to systems. Challenges: 2, 3, 4, 9, 10, 11, 17, 28, 30, 32.

The following problems occur with the manual registration: There is no trust in the data; The registration process is not followed correctly; It is felt as redundant work; The organisation changes too fast to be able to keep the registration up-to-date; Registration feels like overhead; People are not aware of the use of outdated data; The quality of the data is substandard and thus not usable, which is because of a lack of trust in the data, because of a lack of standardised format and because there is no ownership on the data.

7.7.3 *Data Types*

In one sense data is used to make the development and monitoring process of IT more efficient and better. This type of data is mostly machine data, which could be large log files, or individual data entries, this type of data is useful when used when aggregated. This type of data can be used to drive the integration of applications. We classify this type of data as *Transactional Data*.

On the other hand data is also used to get grip on the IT landscape. Data is used to explain how IT assets are available and how they relate to each other. These types of data are used by multiple business processes and we therefore categorise this type of data as *IT Landscape Data*.

During our case study we found a lot of different data sources within IT, we have listed them in Appendix C. We also categorised two data sources as Metadata. As these sources describe other datasets. We did not find Reference Data types during our case study.

8

DATA MANAGEMENT OPPORTUNITIES AND MODEL REDESIGN

8.1 ADDED IT ENABLING CAPABILITIES

We found that the Data Challenges that arise within IT (Section 7.7) cannot be covered with solely the Data Management Framework (DMF) as described in Section 2. Our proposed capabilities reflect what IT needs to be able to tackle their data challenges and to get in control of data. It could be argued that the capabilities would not fit in a Data Management Framework, that is why we present them as IT enabling capabilities next to the framework.

8.1.1 *Automatic Data Generation*

Data sources that are created manually are subject to quality issues (see Section 7.7.2.2). Some data sources could, however, be extracted from tooling automatically, such that no manual action is needed. The lists will probably contain fewer errors, will provide a complete overview and will therefore probably be trusted more. This capability will, therefore, reduce the need for Data Management processes.

This capability could assist in providing automatically generated IT Landscape Data. An example is that the status of running infrastructure could be automatically detected. This way, the IT infrastructure designs do not have to be kept updated manually but can be extracted automatically. This machine data is indisputable, as opposed to the manually created lists.

8.1.2 *Data Value Presentation*

The creation of a *reliable* data source requires effort by employees. However, it is often not clear why effort should be made to maintain and manage data assets.

This is supported by Tadhg and Sammon [51] which also observed the difficulty that organisations have in clearly communicating and deriving value from specific data initiatives.

It is therefore important that the value of the data is shown to the organisation. It was found that it is highly likely that more effort will be made to create and maintain a source that is reliable if it is known that the data is valuable for others.

We distinguish this capability from the Data Management capability Data Awareness & Education, as that capability focuses on raising

awareness on the process of data management, instead of presenting the value it can offer when it is used by a 'Data User'.

Data User refers to a person who represents the people who use a data source to turn it into value

We advise creating a dedicated page on the intranet of the organisation in which value cases are presented. The page could contain a structured list with crafted *white papers* based on the value cases. The value could also be presented with the use of business intelligence dashboards, links to those interactive reports could be provided on the page. This way, anyone in the organisation can gain inspiration on what is possible with data in general, but can also see what data they contributed to is used for.

Tadhg and Sammon [51] created a 'Data Value Map' which functions as a template that can be filled in by Data Owners and Data Users to create a shared understanding of the value as well as the Data Governance approach. Although it shares similar model areas with the DMF, it could provide lessons in which the value-generating process is spelt out for all data stakeholders.

8.1.3 *Data Value Tracking*

Data Management is performed to enable the creation of value with data. Data Management requires resources; this process is, however, not worth the effort if data is not used, or if the value of the data does not outweigh the effort that is put in.

In order to become a *data-driven* IT organisation, it should be known how data is being used and what the value is. We provide the capability Data Value Tracking as a tool to measure the adoption of data and to show if the data management process is worth the investment.

This is especially relevant for IT since it was explained that the registration process is felt as overhead or that the reason registration process is not understood. As opposed to the capability *Data Value Presentation*, which is used to convince the organisation of the value of the registration process, we provide this capability to measure if the registration process is actually necessary.

8.2 PRIORITIES DATA MANAGEMENT

Based on the challenges described in priorities for Data Management in IT. Figure 8.1 displays the capabilities with our given level of priority.

Data Management can give guidance to tackle these challenges. The next sections explain the priorities.

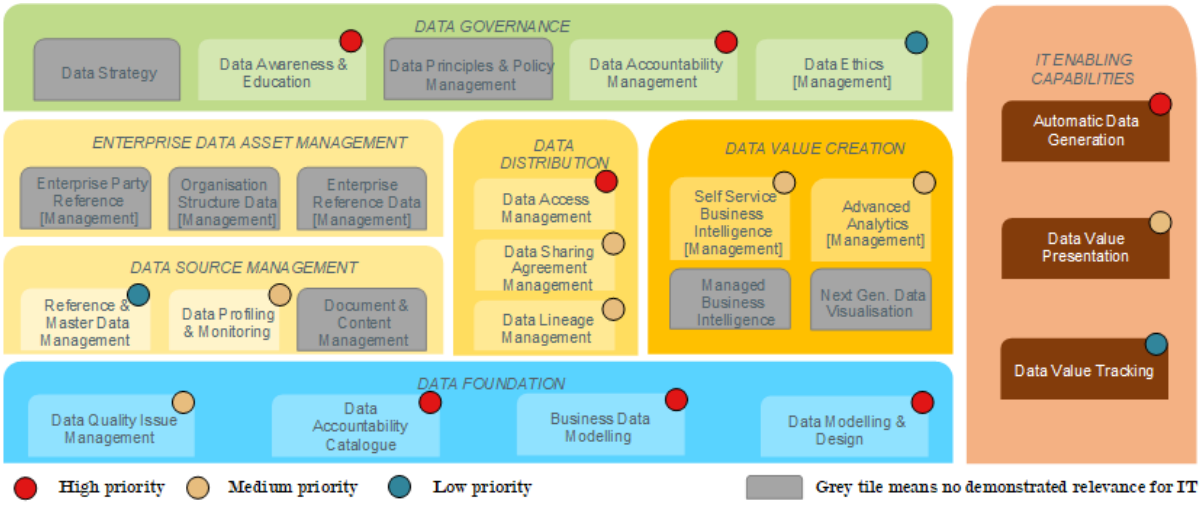

Figure 8.1: Priorities Data Management IT

8.2.1 *Tackling Main Challenge 1: Unclear relations between IT landscape components*

One part of Data Management focuses on the alignment of the organisation. The *Business Data Modelling* and *Data Modelling & Design* capabilities are complementary, but the combination has two goals for IT.

The first goal is to align IT definitions with business definitions, such that other levels of the organisation can communicate their needs to IT and use data from IT. Employees have to register the descriptions of the different data when a data source is going to be added to the Data Accountability Catalogue. This approach allows different 'dialects' of data descriptions. This means that the different interpretations of the data are registered. Business Data Modelling should tackle the challenge that business and IT speak different languages.
This should assist in tackling challenges 12, 13 and 29.

The second goal is to align IT data definitions, such that everyone in IT is talking about the same data model. A uniform data model, on the other hand, leaves no room for dialects. DevOps has a large influence on the need to create a common understanding. The two worlds with different terminology in their data models come together in one model. A common data model is necessary to be able to relate the different technological building bricks of infrastructure and applications with each other. At the moment it can not be described in a uniform way how IT products are composed out of smaller components.

The goal is to provide a common understanding on how IT works. A data model is described as follows according to DAMA International [17]: "A data model describes an organization's data as the organization understands it, or as the organization wants it to be. ... Data models are the main medium used to communicate data require-

ments from business to IT and within IT from analysts, modelers, and architects, to database designers and developers.".

A clear data model is necessary for traceability in the IT landscape. This is needed to perform analysis throughout the whole IT value chain (see Section 7.7.2.1). A data model should contribute to link data sources, such that every error or feature can be traced across the value chain.

The capability *Business Data Modelling* is defined by FDO as the modelling on a semantic level. While the capability *Data Modelling & Design* focuses on the modelling on logical and physical level. Business Data Modelling should be performed first, before a logical and physical data model can be created.

The data model enables that data elements across the IT value can be linked together. This leads to the traceability of the IT landscape. This traceability is needed for challenges 1, 5, 7, 14, 15, 19, 20, 21 and 27.

8.2.2 *Tackling Main Challenge 2: Quality issues with data*

8.2.2.1 *Creating Awareness and providing education*

We have noticed that Data Management is not a widely known concept within IT yet. The IT organisation should first be taught that it is important to manage data assets, which is required to tackle their data challenges. Currently, the level of awareness of the topic is generally low in IT.

With 6000 employees, from which a part works from vendor location, it is a challenge to let everyone join in on the Data Management processes. Organisation-wide campaigns for data management, especially data quality, were organised. But it might be a good idea to bring focus to a campaign for IT. We have heard numerous times that awareness to maintain data works best if the value is shown. We suggest to work out a number of cases from beginning to end, in which the process of Data Management is handled in IT. It should be explained what the value of Data Management was, instead of explaining what the Data Management or Data Quality process is. It would be easier for employees to imagine if their data could need a similar process.

We see the need to raise more awareness in the organisation about the data quality issues that arise (9, 28), to make people aware of the data management process for better data governance (16, 17, 26), but also to show the value of what their data can offer (18).

8.2.2.2 *Onboarding sources and assigning accountability*

Data Accountability Catalogue, Data Accountability Management and Data Quality Issue Management are the foundation for the next steps

in the management of data in IT. Automatic Data Generation should be prioritised as well.

DATA ACCOUNTABILITY CATALOGUE The Data Accountability Catalogue is a known concept in the organisation, but IT has only provided limited data sources. This is why the focus should be to fill the list with data sources that are available within IT. It is also advisable to think of interpretations of the transactional data, which might be useful to know for others in the organisation.

We especially see priority to create a single source of truth for Master type data sources. These sources are used throughout the IT organisation to get a timely and reliable overview. It should be known that the appointed Common Sources are the source that can be relied on.

Transactional data types also need a Common Source, such that the data can be aggregated at one place, to prevent the spread of loosely coupled datasets with the same data. It should, however, be known who is the owner of the data, such that the data can be shared to interested users and such that data quality issues can be solved.

It should be verified which of the sources listed in Appendix C could be included in the Data Accountability Catalogue.

We suggest focussing on gathering as much used data sources within IT and grouping the resulting list by department and data type. This could be done by sending out a survey to the product owners and department leads within the organisation. Another approach is to identify the data within tools used within IT, together with the product owners of these tools. Data could also be found with the help of managers within IT since they might have an idea of how the data in their department could be used to innovate. It might be a good idea to conduct focused interviews.

We think it is important to ask strategic-level employees as well as operational-level employees about data sources, as the latter may be working with data sources that are not directly known to strategic-level employees.

A Data Accountability Catalogue could be needed for incident data (2), for IT product data (4), for IT landscape data (17) and team characteristics (26).

DATA ACCOUNTABILITY MANAGEMENT The registration of Common Sources goes hand in hand with the administration of accountability of the sources. Once Common Sources are identified, it is important to assign ownership.

FDO has defined a step-by-step approach which should be used for the onboarding of Data Owners and Data Users. We advise using this as a guide in IT as well. The approach consists out of activities in which the Data Owners register their data sources in the Data

Accountability Catalogue, in which they register the definitions of the data, in which they make the data available, and eventually come to an agreement with the Data Users on how the data is structured and how it is shared with each other. The Data Users should make clear what their data needs are and what the quality of the data should be. The step-by-step approach also provides activities in which data is shared. Once this is all in place, data quality issues can be raised by the Data Users, such that the Data Owners can act on it.

It is also important to educate the people that are assigned to those roles, such that they can perform responsibilities that come with their role. Data Owners, for example, represent the interests of the data creators and Data Users represent the Data Consumers within the organisation. The people that are assigned to those roles have to have meetings with those who they represent and be able to act in the Data Management process.

The challenges are similar to the Data Accountability Catalogue.

AUTOMATIC DATA GENERATION We see the opportunity to make better use of data that is already available in internal applications within the bank, in order to replace manually created IT Landscape Data sources. We also see the opportunity to generate data using modern tooling, which, for example, can identify how the IT Landscape looks like.

We suggest starting identifying data sources that are manually composed but are already available in another application elsewhere in the bank. An example is the registration of members of a team (source C.1.1). A department might keep an own list of their employees and where they work, while the data may already be registered in an application at the HR department. The data should be retrieved and maintained there instead of using a manually composed list.

We also suggest determining which IT Landscape Data sources could be automatically created with the use of modern tooling. We see the potential for automation for the discovery of: Running infrastructure, applications running on infrastructure, team composition and team members from the vendor.

There is also a potential to extract some data sources which we listed as transactional data from systems. The registration of changes in development and the registration of given access rights to applications could, for example, be automatically retrieved from the tooling that is used by the DevOps teams.

The Data Management team from the IT Consultancy Office could take the lead in this initiative, and discuss which sources could be automatically retrieved with stakeholders within IT. This capability could be combined with the search for Common Sources within the organisation as described previously.

This capability can assist in tackling challenges: 3, 8, 10, 17, 20 and 27.

8.2.2.3 *Data Distribution and Quality Monitoring*

Once there are users for the data, the IT organisation then needs to share the data with each other, this is done via Data Distribution. While data might already be shared across the IT organisation, it is not administrated yet. As defined in the Data Management process defined by FDO, the data source should first be listed in the Data Accountability Catalogue, and the Data Owner and User should agree on the conditions for sharing before data sharing is started in the Data Management process. The Data Management team should follow this process.

DATA ETHICS Data Ethics is not a very prominent capability for IT, but it is important for some use cases. Data Ethics should be thought of when the data is going to be shared, as well when the data is actually being handled. The importance of Data Ethics is, for example, shown in the use case in which managers get insight into the performance of teams or individuals. Unethical decisions based on the data could have negative impact on how employees feel and work in the organisation. Another case for Data Ethics with IT data could, for example, be with the usage of customer data, with the goal to create a digital customer profile based on click actions or to correlate the customer's actions with system performance. This may have an impact on privacy concerns of the customers. We advise to include ethical thinking in the data-sharing agreement process.

DATA ACCESS MANAGEMENT Data should be shared according to the standards set by the FDO department.

We, however, see that data which manually handled is often subject to quality issues. The data created by applications is indisputable, in contrast to the manually defined administrations. Required data can often be extracted automatically from another application, while the creation of data sources happens manually in various cases. This automation step eliminates redundant work and removes issues that arise with manual work. An automation step also enables the consumption of more timely data, as it is possible to create real-time exports of datasets.

It is desired to integrate applications, but the integration should be designed such that additional applications could be integrated easily as well. Creating point-to-point integrations between applications is not desired, as it introduces a lot of dependencies. The data should be retrieved from designated hubs, such that applications can share data directly via APIs with each other.

This capability should be used to enable the organisations with integrating applications with the use of APIs such that data can automatically be retrieved. The use of APIs is important for transactional data sources. These sets are already machine-generated; API integration takes unnecessary manual actions away.

The data-sharing platform provides tools to share data across the organisation. The data sources should be hooked into this platform, such that there is a central place the data is supplied from. There is a predefined process to onboard data sources on the data-sharing platform, as well as a process for consumption.

While *Automatic Data Generation* focuses on the automatic creation of data sources, this capability focuses on the integration of data sources that are already available.

DATA QUALITY ISSUE MANAGEMENT AND DATA PROFILING MONITORING Data Quality issues are rarely reported within IT, but issues with the quality of data came up during our interviews. The lack of issue reporting can also be explained by the limited number of Common Sources registered. Once ownership is assigned, it is important to make IT aware that issues are fixed in the Common Source. Data Awareness & Education should provide this awareness.

It is especially important for Master Data sources to be of good quality. This means that the data quality process should be in good order. We also found data quality issues with incidents and problems. We advise to make people in the Incident, Problem & Change department responsible for communicating the data quality issues back to the people that created the entry in the system for the issue or problem.

Data Quality Issue Management can help tackle challenges: 2, 9, 16, 17 and 26.

8.2.3 *Value Output*

ADVANCED ANALYTICS Advanced analytics could be used for value creation in the IT organisation. Correlations might be found between different processes across the IT value chain. We advise performing advanced analytics by centre of excellence (CoE) departments across IT.

This capability can assist in tackling challenges: 5, 6, 14, 15, 22, 23, 24 and 26.

SELF-SERVICE BI Also, give teams tools to track their performance. This could be done with self-service business intelligence, in which trends can be visualised. The most value can be extracted from data if it is shown over time, such that trends are visible and potential correlations between different datasets can be created.

This capability can assist in tackling challenges: 22, 23, 24.

8.3 DATA MANAGEMENT RECOMMENDATIONS

We pointed out the priorities for each capability that needs action. In this section, we provide advice for the Data Management department at the IT Consultancy Office such that Data Management can be rolled out within IT.

CONTINUE CREATING THE DATA MODEL FOR IT The initiative to create a uniform data model within IT is already started. We advise keeping actively participating in the initiative.

CREATE AWARENESS CAMPAIGNS WITHIN IT AND FIND COMMON SOURCES Work out or present a case within IT in which the data quality process has shown the value. Actively search for Common Sources together with stakeholders within IT. Also think of interpretations of data that could be interesting for others. Management might, for example, be interested in the average mean-time to recover of all DevOps teams.

Data sources are already being used within IT; else there would be no use case for it to exist. These users could already be approached to represent the users of the data source.

Then assign Owners and Users. We have listed potential sources in Appendix C, determine if these could be used. Start with the Master data sources as those could benefit the most from the quality governance process. Then hook up Common Sources and Owners/ Users and continue with the step-by-step Data Management approach of FDO.

1. Suggestion: Point out a Common Source for IT products (challenge 4)
2. Suggestion: Create awareness of quality issues at problem registration (challenge 9)

ENABLE AUTOMATIC DATA GENERATION Look for sources that could benefit from Automatic Data Generation. And go into discussion with stakeholders that create the source or could help retrieve the data automatically. Once a source is found, create a project plan for data creation with automation.

1. Suggestion: Retrieve access rights from tooling within Identity & Access management (challenge 8)
2. Suggestion: Retrieve changelogs from the tools the developers use to register changes instead of supplying an extra tool for the registration (challenge 10)

Also, look for sources that could benefit from API integration.
Suggestion: HR database in Identity & Access management

SET UP A DATA VALUE PRESENTATION PLATFORM Create a project plan to promote Data Value Presentation.
Suggestion: For a first proof of concept version a page on the intranet which is controlled by the IT Consultancy Office. If it has potential, it might be possible to expand it such to provide more functionality, such as interactive demo's and let employees upload their cases themselves. Also, promote the page via the awareness campaigns.

DATA VALUE TRACKING Periodically evaluate the value of the Data Management process by quantifying the value that it delivers. If the effort of maintaining a data source does not outweigh the benefits, it might be advisable to stop.

VALUE CREATION Provide brainstorm sessions with different stakeholders across the value chain on use cases in which advanced analytics can be used with the IT Value chain data. Create a project plan with the stakeholders and set-up a team to implement advanced analytics projects.

1. Suggestion: Correlation analysis incident and IT value chain data (challenge 2, 6, 15)
2. Suggestion: Correlation between system failure and data from monitored infrastructure (challenge 5)
3. Suggestion: Correlation analysis with customer experience and changes in software (challenge 14)

Also allocate resources within the Centre of Excellence of the CTO business unit to educate the DevOps teams on the usage of tools that enable them to perform analysis on the data they generate with the CICD tooling.

8.4 ADDRESSING THE REDESIGN OF THE DATA MANAGEMENT FRAMEWORK

The Data Management Framework is an extensive model; we explained how it could be used within IT. The model is well constructed and provides capabilities that enabling solving the key challenges that arise with data in the organisation. Therefore, we do not see the need for a largely redesigned model. We, however, see the need for three strategic capabilities which could be used for IT (Section 8.1). We do not suggest to provide a separate Data Management Framework for IT; we, however, find it important to indicate the focus areas to spark a discussion about Data Management needs. We did not find a demonstrated value during our case study within IT for eight capabilities. These might be however valuable in the future if Data Management has matured in the organisation, this is why we present a model in which these capabilities are greyed out. We also added the 'IT enabling capabilities' for visibility. The Data Management Framework for IT would look like presented in Figure 8.2.

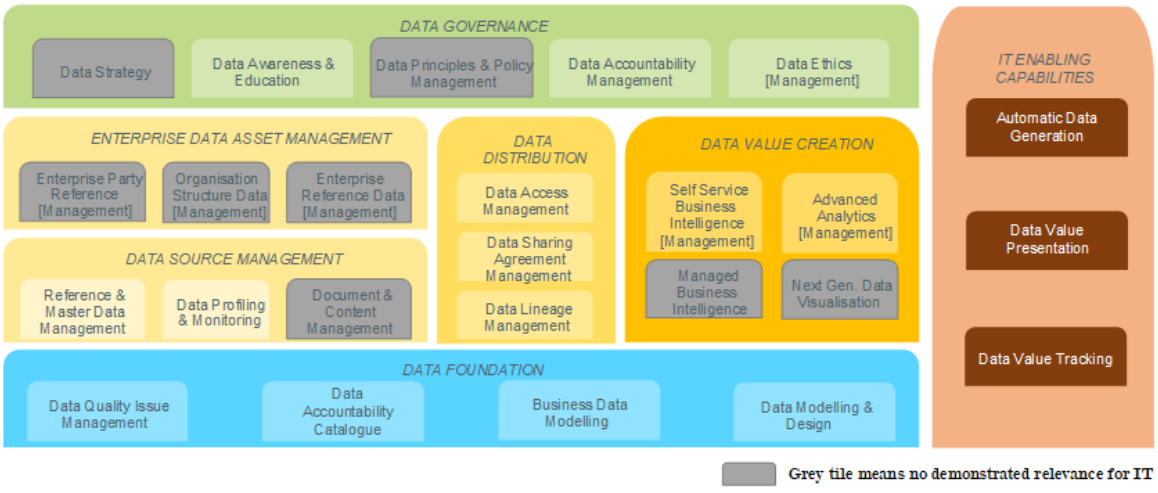

Figure 8.2: Data Management Framework with added strategic capabilities for IT

9

VALIDATION

We presented key findings of this study to six IT experts that did not participate in the case study before. The validation interviews were based around a slide-show presentation in which the experts could comment on the main challenges we defined, our recommendations for a Data Management approach in IT, and our proposal for added capabilities that would be useful for IT. The validation results were used to improve the draft versions of those deliverables and were used as input for the discussion in Chapter 10. The validation results are summarised in this chapter.

Interviewees generally acknowledged and recognised the data challenges that we presented. We defined the *Traceability* of the IT landscape as one of the main goals of data within IT. An interviewee indicated that traceable sounds more like a means than a goal. In his view, the goal would be more control and efficiency of the IT landscape. Three interviewees also recognised that the capability Data Lineage also enables some a type of 'traceability'. While Data Lineage mainly describes how a single data source is transformed through a flow of processes [2], we describe traceability as mapping the relationships between the different data sources. Data Lineage focuses on the data life cycle, while traceability describes a situation in which data sources are referenced to each other.

During the interviews we proposed four other focus areas that could be placed next to the Data Management Framework: *Data Standardisation*, *Point-to-Point Data Integration & Automation*, *Metrics Creation* and *Data Value Presentation*. *Data Standardisation* was formulated as a capability that can be used to define standards for tooling in which data is registered, in which format a data source should be published and which fields the data source should contain. The interviewee, however, indicated that it is more important to define a source for the Data Accountability Catalogue.

The *Point-to-Point Data Integration & Automation* capability had in addition to our definition of *Automatic Data Generation* the goal to integrate applications directly with each other, instead of using manually created datasets. Two interviewees indicated that the IT organisation is moving away from point-to-point integration, since focusing on API platforms instead removes a 'spaghetti' of relations between applications. They explained that these APIs are facilitated in an existing data-sharing platform. This platform is part of the Data Accessibility capability in the Data Management Framework.

The goal of *Metrics Creation* was to enable the organisation with the creation of use cases by quantifying data. It described the need for IT to use quantified data to get in control and improve. This capability did, however, not provide a practical approach and could not be defined clearly.

An interviewee indicated that the proposed capability *Data Value Presentation* was partially present in Advanced Analytics, which also presents what the value of Advanced Analytics could be. Our definition of Data Value Presentation presents the need for visibility for all types of data value-generating processes.

The interviewees generally agreed with the Data Management priorities, but some were changed. Two interviewees also indicated that it is important that it described who is responsible for the execution of a Data Management capability, such that theory is translated into practice. We furthermore included the interviewees' suggestions in our advice for Data Management for IT.

Another point we validated is whether data accountability might be needed for data in the raw data storage. Two interviewees, however, indicated that accountability should be created at the source of the data, not in this storage. Therefore we discarded this idea.

The interviewees furthermore provided suggestions to edit how definitions and challenges were worded. Interviewees also provided additions to the results we presented. We used this to improve Chapter 8.

10 DISCUSSION AND CONCLUSION

10.1 DISCUSSING DATA-DRIVEN IT

In Chapter 4 we presented what constitutes capabilities for large data-driven organisations. We found that the capabilities presented in the chapter also come back within Fortran IT.

The Fortran Data Management Framework (DMF) does not focus on organisational aspects on culture and people, while they are important aspects for a data-driven organisation. The Data-Driven Capability Model, which was created as part of the literature review (Figure 4.3), provides the dimensions: organisational culture, skilled personnel and management, but also touches the importance of technological infrastructure.

10.1.1 *Management*

In Chapter 4 we found that the main challenges of building a data-oriented culture are managerial and cultural focused, whereas the technical focus plays a minor role [38]. We can endorse Kontio, Marttila-Kontio, and Hotti [38] that the main challenge is not technical, but also not mainly managerial. The interviewees in a management position explained that a data-driven culture should come from both sides. Managers saw the need to create a data-driven strategy, but it was indicated that in the end employees on all levels of the organisation have to come up with use cases for the available data. The people who work with data should be willing to put effort in the data facilities that are provided to them, in order to be able to drive Data Management and value creation with data. Enabling this culture cannot be solely done by management.

We found that people are not always comfortable working with data (Challenge 18 in Appendix B). Teaching people to work with it is one part to tackle it [42], but we found that showing the value of data such that people challenge their intuition with data is also important. This could be one of the responsibilities of management.

Fortran leads their data strategy via their FDO department, provides leadership via a designated *Data Board* of *Data Executives* and has pointed individual leaders throughout all parts of the organisation (see Figure 5.4 just like Kim and Gardner [35] suggested).

10.1.2 *Culture*

We previously described that data should be democratised. We see that this principle is embraced very well in Fortran. Instead of creating siloed datasets for own purposes, employees have to move to shared data sources to solve organisation-wide data quality problems. The democratisation is enabled with organisation-wide Data Management sharing capabilities within Fortran. Just like Patil [46] teams at Fortran are given access to data and to self-service business intelligence tools to create their own analysis and dashboards.

A data-driven *Culture* depends on how employees use data in their way of working. We especially see a shift in mindset for the software development teams because of the transition to DevOps and the introduction to CICD. The teams are handed data about the full chain of development and operating processes, which can be used to track performance. The teams can decide on data, which was not insightful before. As explained in Section 7.6.5.3 some teams are eager to become a *data-driven* team, but some are still sceptical on how the data is used. This is also a challenge for *Management* to give the teams trust that their data is used to improve the organisation, instead of being used for controlling purposes.

Trust in the data itself is also a *Cultural* aspect that needs to be improved in the organisation. However we think this cannot be imposed by rules, but the quality and availability of the data should be improved instead. Data Management enables this improvement.

10.1.3 *Infrastructure*

We found that infrastructure is an important component in the enabling value creation with data. CICD and Public Cloud are two principles that change how the infrastructure is configured and used. Modern tools that come with these changes make it easier to integrate applications for continuous data flows and make it easier to extract data about the running infrastructure as well. Fortran IT is cleaning up their infrastructure in the landscape, to assists in making IT data more accessible. Better data quality might lead to more trust [48], Fortran is tackling these quality issues with Data Management (Data Quality Issue Management and Data Profiling & Monitoring). Their framework also provides capabilities to improve traceability of data sources with the use of uniform data models (Business Data Modelling, Data Modelling & Design), such that the end-user can see what the source of the data was [35].

10.1.4 *Skilled Personnel*

Marchland and Peppard [42] addressed the lack of understanding of the value of data, we recognised this as well at Fortran. The Data Management capability Awareness & Education focuses on the importance of data and data management, but could also be used to train employees to frame questions and interpret their results like Marchland and Peppard suggest. The lack of trained skilled staff [35] is mainly tackled by Fortran by outsourcing a lot of work in IT to vendors.

10.1.5 *Analytics*

Data is used for analytics within IT, but also to connect the tooling in CICD. The Data Management Framework assists analytics creation with Business Intelligence and Advanced Analytics. A clear map of which data is collected [35] is done with the Data Accountability Catalogue. This approach is value-driven, initiatives to consolidate data should have a business case [22]. This is why Data Users are registered in the Data Accountability Catalogue as well.

10.2 ADDRESSING THE RESEARCH QUESTIONS

This study aimed at investigating how data challenges in the IT organisation of the large financial institution Fortran could be tackled with the use of a Data Management Framework. Secondly, the research also aimed to find out what the next steps for Data Management within the IT organisation of Fortran would be in order to get grip on data assets.

It has the following main research question:

What constitutes a usable capability model for Data Management in an internal IT organisation in a financial institution like Fortran?

We used five sub-questions to answer this question. A literature review and a case study with 16 interviews with IT experts was performed to answer these questions.

1. What are key capabilities that support large data-driven organisations?

We first investigated what it means for large organisations to be data-driven. We found that data-driven focuses on the use of data in highly organisational specific processes, but a basis for a data-driven organisation at scale can be defined with related concepts such as *Business Intelligence and Analytics*, *Data-driven decision making* and *Big Data*.

These concepts are highly related and are described as direct enablers of each other. In this sense, Business Intelligence & Analytics effectuate Data-Driven Decision Making and the former continues to transform based on technological advancements. Their theoretical basis is alike. Literature provides us with capability and maturity models about these topics, which turned out to share similar key capability areas. The transition from a company with data to a data-driven company is enabled by more than technology alone. Other factors, such as organisational culture, skilled people, management buy-in, analytics solutions and data management are influential for successful adoption of a data-driven way of working.

With the use of those capabilities and with literature we defined a Data-Driven Organisation and present it as a unified capability model.

2. What is the current state and research agenda of Data Management Capability models?

Fortran focuses on tackling organisation-wide data challenges with the use of Data Management. Academic literature did not yield results on the topic, while we found numerous industry standards. We explained the key concepts of the topic and introduced how their guiding framework relates to other industry standards. The Fortran Data Management Framework covers most aspects of other data management models, most of which are based on the DAMA-DMBOK2 model. A DMBOK2 key area that the Fortran model does not cover is Data Security, which was later explained as a responsibility for the IT security department. While security is an important aspect, we did not find proof in our case study that the way this is handled should receive attention to be changed and thus propose to leave it out of the Fortran model as it is.

We found that the topic of Data Management in IT is underexposed in academic literature. No academic literature could be found on data management frameworks such as DMBOK2. Data Management in IT is also not touched in literature, as far as we found. This research can contribute to filling this gap, by providing one of the first academic writings on Data Management as described in industry standards and by providing an insight into the industry application of Data Management.

3. What are data challenges of the IT organisation at Fortran?

During the case study we found that the main intention of IT is to make the IT landscape traceable to perform automatic remediation, get insight in the costs of IT, to be able to solve incidents quickly and to be able to find correlations between data across the IT Value chain. This goal is however hindered by two main problems: (1) Relationships between infrastructure, applications and teams are not clear, (2) and there are quality issues with the data.

We found that IT is pressured to become more efficient and cost-effective. IT is however struggling with issues concerning their IT landscape. This is why the main priority of IT is to make the IT landscape traceable. If this is not done, IT misses value-generating opportunities.

There are however challenges that hinder the traceability of the IT landscape. We identified 33 challenges which occur across the different IT Value Chain streams, these can be found in Appendix B.

4. How can Data Management contribute to the IT organisation at Fortran?

We performed a reflection on how Data Management can assist each of those challenges. We indicated the Data Management capabilities that need attention such that the challenges can be tackled. The first main challenge (1) should be tackled by creating a common data model for IT. The second main challenge (2) can be tackled with the use of Data Management capabilities that focus on agreeing on a common source and controlled sharing of these sources across the organisation.

We provided advice for the Data Management department for IT such that they can prioritise the work that needs to implement Data Management within the IT organisation.

5. What would be a suitable capability framework to support data challenges in IT organisation in financial institutions?

We used a Data Management Framework from the organisation-wide Data Management department as a guide for the capabilities. The framework is well designed and fits most of the needs for IT, we therefore do not propose a new design of the model, but suggest a prioritised model with added focus areas for IT. We found, however, that some challenges could not be solved with the Data Management Framework as presented. We therefore present three more capabilities, outside of the Data Management Framework, that should enable IT to tackle their data challenges. We showed the importance of automating the composition of data as opposed to the creation of manually created sources with *Automatic Data Generation*. We described the need to make it clearly visible how data sources are used to create value, by introducing the capability *Data Value Presentation*. And we described the need to measure the value of data and the Data Management initiatives with *Data Value Tracking*. We furthermore indicated which capabilities on the organisation-wide Data Management Framework do and do not need action for IT.

10.3 APPLICATION IN OTHER CONTEXTS

We think the lessons presented in this thesis can be useful for other organisations than Fortran, as other financial institutions are also

working on creating better governance of their data [54]. The ING Bank in the Netherlands for example, is struggling with similar issues, according to a report on their journey to become data-driven with the use of Data Management [15]. The authors of the report describe that almost none of their systems shared a similar data model when they started with Data Management. Main issues explained conflicting definitions across the organisation leading to an inability to share data, the use of outdated data and an unclear IT landscape. The ING tackles these issues mainly with the use of an organisation-wide data lake, a data catalog, definition of glossary terms, data lineage and support for ING specific metadata. The ING establishes, in contrast with Fortran, a single data-exchange language for the whole bank (named *ING Esperanto*). An interviewee in the validation round of our case study explained that Fortran has deliberately chosen not to impose a common language to the whole organisation. It was explained that due to the different interests of the different departments, the common language for the whole bank has a high chance of not being accepted. The report, however, illustrates that ING is tackling similar issues as Fortran with the use of Data Management and thus it is likely that our findings could be relevant for their Data Management approach.

The findings of this thesis could even be valuable for organisations outside the financial services industry since we depicted the IT department as one that mainly focuses on software engineering. Other organisations are looking for ways to increase their grip on data assets within an IT organisation as well; this is a concept that is not only applicable to the financial services industry.

10.4 FINDINGS AND CONTRIBUTIONS

10.4.1 *Key findings*

1. Being in control of data needs a shift in mindset

Data Management is a well-thought approach to get grip on data assets, but it needs to be embraced by the organisation. Setting up processes to control data is one thing, but people have to be willing to put the effort into the process as is expected from them. The process is not useful if no effort is done to create data of good quality, if data quality issues are not raised or if people are not willing to use a recognised data source. In the end, the process depends on the effort that is done, therefore an effective implementation of Data Management needs a shift in mindset.

2. Standardisation is an important part of controlling IT data assets

Reaching a common understanding is an important step for the effective use of data within IT. The large size of the organisation and many

different kinds of activities make room for different ways of working and different terminology. Standardisation of terms and definitions is needed within IT, it is needed to be able to relate IT assets together. Agreeing on key IT definitions is necessary to bridge terminology the gap between business and IT. Standardised publishing formats are needed to be able to relate datasets with each other. Standard CICD tooling should be used for performance management. Agreeing on common data sources is key to recognised good quality data.

3. Responsibility for data assets is key to adoption

We found that the most important aspect concerning IT data assets is that someone is accountable for the data. Someone should be in charge of deciding what the data should look like, how quality issues are resolved and how it relates to other data sets. Nothing changes if data is just used by people in the organisation without any governance. Setting responsibility is the first and most important step in improving data quality, which translates to better adoption of data in the long term.

4. DevOps and CICD lead to more IT control, Data Management enables control of data

Control on IT is very important for financial institutions. Within the IT organisation of Fortran there is a strong need to get in control on performance, on organisational change, on money expenditure and on what IT assets are available. The large size of IT makes it a challenge to manage all teams, applications and infrastructure. The change initiatives enable more control. The change to DevOps provides more control on IT teams. The implementation of CICD leads to more control over the development pipeline. Data is an important asset to enable the needed insights throughout the organisation; Data Management provides better control of the data available within IT.

5. Traceability is key to value creation within IT

With the large size of the IT organisation different ways of working emerge, which leads to different data definition and source formats. In combination with Data Quality issues it is hard to connect data sources with each other. The ability to connect IT data sources, *traceability*, can translate into value, such as, increased control and the enablement of advanced analytics. Once traceability is in place it is possible to get a reliable overview of the IT landscape, but it is also possible to perform correlation analysis or performance management because there is an end-to-end overview of the software development process.

10.4.2 *Contributions*

In this research we demonstrated the practical application of Data Management. The contributions are:

1. **practice**: We have uncovered common bottlenecks that arise with data in an IT organisation and have shown the potential for Data Management to tackle these bottlenecks.
2. **practice**: We provided prioritised advise which can be used as a guide to enrol Data Management within the IT department of a large financial institution.
3. **academic**: We demonstrated the need for the discipline Data Management by providing an academic thesis that gives novel insight in the practical interpretation and application of an industry recognised Data Management model.
4. **academic**: We provided focus areas with clear practical use that could improve an industry used Data Management Framework.

10.5 LIMITATIONS, FUTURE WORK, AND RECOMMENDATIONS

This research has been conducted in a single financial institution. We can imagine that the results of this study are interesting for other companies, but due to the large influence of organisational specific context the proposed Data Management Framework could not be fully applicable. Future work could be done to verify the applicability of the Data Management approach of Fortran in other contexts.

The interviewees presented a broad image of the work in IT, they could particular tell about their field of work. This however makes that some challenges are based on a single interview. More work should be done within the organisation to validate if the challenges are valid.

The scope of the research is focused on the business units CISO, CIO and CTO (see Section 2.2). Our case study participants mainly covered the CTO department, since this department assists the other two. Further investigation should be done with stakeholders within the CISO and CIO business units to validate the findings of this report.

This thesis introduces the role of data within DevOps. We pointed out the introduction of DevOps and CICD within IT, but as the focus of this research is on data, we did not provide an in-depth research what problems the concepts solve. More research could be done to explain this, what data is needed for it and how data could be used for a successful implementation of DevOps and CICD.

Although the role of Artificial Intelligence (AI) was only covered partially as an approach to perform Advanced Analytics, we see it as a promising work of field that can benefit from Data Management and can assist in automation of Data Management. As with any type of

analytics, Machine Learning models need good quality data in order to produce a reliable and precise result [13], good Data Governance also leads to less time spent on cleansing and preparing the needed data. On the other hand it could be interesting to *use* AI as an accelerator within our proposed capability Automatic Data Generation, in order to automatically compose data assets. Another promising trend in technology is Robotic Process Automation (RPA). It is a software based solution that can automate processes which involve routine tasks that used to be done by humans [1]. RPA might be a solution to perform Automatic Data Generation. Future research could be done to investigate the possibilities of Artificial Intelligence and Robotic Process Automation for Data Management.

We recommend that Fortran uses our suggestions as a guide to create their own Data Management Roadmap for IT. We have provided practical data challenges which could be resolved, potentially with the solutions we suggested. We also recommend FDO to use this thesis as an input to improve their Data Management strategy, as they were specifically looking for links between their theoretical model and the practical application of it in the business lines.

BIBLIOGRAPHY

[1] S. Aguirre and A. Rodriguez. "Automation of a business process using robotic process automation (RPA): A case study." In: *Communications in Computer and Information Science*. Vol. 742. Springer Verlag, 2017, pp. 65–71. ISBN: 9783319669625. DOI: 10.1007/978-3-319-66963-2.

[2] M. Allen and D. Cervo. "Data Quality Management." In: *Multi-Domain Master Data Management*. Elsevier, 2015, pp. 131–160. DOI: 10.1016/B978-0-12-800835-5.00009-9.

[3] Amazon Web Services. *What is DevOps?* URL: https://aws.amazon.com/devops/what-is-devops/ (visited on 06/28/2019).

[4] Atlassian. *Secrets to agile estimation and story points*. URL: https://www.atlassian.com/agile/project-management/estimation (visited on 07/02/2019).

[5] Atlassian. *What is DevOps?* URL: https://www.atlassian.com/devops (visited on 06/28/2019).

[6] A.W.A. Boot. "The Future of Banking: From Scale and Scope Economies to Fintech." In: *European Economy - Banks, Regulation and the Real Sector* 3.2 (2017), pp. 77–95.

[7] H. Braun. "Evaluation of Big Data Maturity Models - A Benchmarking Study to Support Big Data Maturity Assessment in Organizations." MA thesis. Tampere University of Technology, 2015. URL: https://dspace.cc.tut.fi/dpub/bitstream/handle/123456789/23016/braun.pdf.

[8] E. Brynjolfsson, L.M. Hitt, and H.H. Kim. "Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance?" In: *SSRN Electronic Journal* (Apr. 2011). DOI: 10.2139/ssrn.1819486. URL: http://www.ssrn.com/abstract=1819486.

[9] E. Brynjolfsson and K. McElheran. "The Rapid Adoption of Data-Driven Decision-Making." In: *American Economic Review* 106.5 (May 2016), pp. 133–139. DOI: 10.1257/aer.p20161016. URL: http://pubs.aeaweb.org/doi/10.1257/aer.p20161016.

[10] R. Buitelaar. "Building the Data-Driven Organization: a Maturity Model and Assessment." MA thesis. Leiden University, 2018. URL: https://theses.liacs.nl/pdf/2017-2018-BuitelaarRuben.pdf.

[11] CMMI Institute. *Data Management Maturity (DMM)*. URL: https://cmmiinstitute.com/dmm (visited on 04/16/2019).

[12] R. Capilla, J.C. Dueñas, and R. Krikhaar. "Managing Software Development Information in Global Configuration Management Activities." In: *Systems Engineering* 15.3 (2012), pp. 241–254. DOI: 10.1002/sys.20205.

[13] B.K. Chan. *Why Data Governance is important to Artificial Intelligence?* URL: medium.com/taming-artificial-intelligence/why-data-governance-is-important-toartificial-intelligence-fff3169a99c (visited on 08/16/2019).

[14] H. Chen, R.H.L. Chiang, and V.C. Storey. "Business intelligence and analytics: From big data to big impact." In: *MIS Quarterly: Management Information Systems* 36.4 (2012), pp. 1165–1188.

[15] M. Chessell, F. Scheepers, M. Strelchuk, R. Van Der Starre, S. Dobrin, and D. Hernandez. *The Journey Continues From Data Lake to Data-Driven Organization*. Tech. rep. 2018. URL: http://www.redbooks.ibm.com/redpapers/pdfs/redp5486.pdf.

[16] R. Cosic, G. Shanks, and S.B. Maynard. "A business analytics capability framework." In: *Australasian Journal of Information Systems* 19.0 (Sept. 2015). DOI: 10.3127/ajis.v19i0.1150. URL: http://journal.acs.org.au/index.php/ajis/article/view/1150.

[17] DAMA International. *DAMA-DMBOK: Data Management Body of Knowledge*. Ed. by Deborah Henderson, Susan Earley, and Laura Sebastian-Coleman. Second edition. Basking Ridge, New Jersey: Technics Publications, 2017. ISBN: 978-1634622349.

[18] Data Crossroads. *Data Management maturity models: a comparative analysis.* URL: https://datacrossroads.nl/2018/12/16/data-management-maturity-models-a-comparative-analysis/ (visited on 04/08/2019).

[19] Data Crossroads. *Exploring data management metamodels: DAMA-DMBOK 2 vs DCAM*. URL: https://datacrossroads.nl/2018/12/02/data-management-metamodels-damadmbok2-dcam/ (visited on 04/11/2019).

[20] T.H. Davenport. "Competing on analytics." In: *Harvard Business Review* 84.1 (2006), pp. 98–107.

[21] Y. Demchenko, P. Grosso, C. De Laat, and P. Membrey. "Addressing big data issues in Scientific Data Infrastructure." In: *2013 International Conference on Collaboration Technologies and Systems (CTS)*. IEEE, May 2013, pp. 48–55. ISBN: 978-1-4673-6404-1. DOI: 10.1109/CTS.2013.6567203. URL: http://ieeexplore.ieee.org/document/6567203/.

[22] A. Díaz, K. Rowshankish, and T. Saleh. "Why data culture matters." In: *McKinsey Quarterly* 3 (2018), pp. 36–53.

[23] EY and Nimbus Ninety. *Becoming an analytics-driven organization to create value*. Tech. rep. EY, 2015. URL: https://www.ey.com/Publication/vwLUAssets/EY-global-becoming-an-analytics-driven-organization/%24FILE/ey-global-becoming-an-analytics-driven-organization.pdf.

[24] B. El-Darwiche, V. Koch, D. Meer, R.T. Shehadi, and W. Tohme. "Big data maturity: An action plan for policymakers and executives." In: *The Global Information Technology Report 2014* World Economic Forum, Geneva (2014), pp. 43–51.

[25] A. Fabijan, P. Dmitriev, H. Holmström Olsson, and J. Bosch. *The Evolution of Continuous Experimentation in Software Product Development From Data to a Data-driven Organization at Scale*. Tech. rep. 2017. URL: https://exp-platform.com/Documents/2017-05%20ICSE2017_EvolutionOfExP.pdf.

[26] N. Forsgren and M. Kersten. "DevOps metrics." In: *Communications of the ACM* 61.4 (Mar. 2018), pp. 44–48. DOI: 10.1145/3159169. URL: http://dl.acm.org/citation.cfm?doid=3200906.3159169.

[27] S. Guckenheimer. *What is Infrastructure as Code?* URL: https://docs.microsoft.com/en-us/azure/devops/learn/what-is-infrastructure-as-code (visited on 06/10/2019).

[28] F. Halper and K. Krishnan. *TDWI Big data maturity model guide: Interpreting your assessment score*. Tech. rep. TDWI research, 2013. URL: https://tdwi.org/~/media/3BF039A2F7E1464B8290D8A9880FEC22.pdfma.

[29] J.S. Horsburgh, S.L. Reeder, A.S. Jones, and J. Meline. "Open source software for visualization and quality control of continuous hydrologic and water quality sensor data." In: *Environmental Modelling & Software* 70 (Aug. 2015), pp. 32–44. DOI: 10.1016/j.envsoft.2015.04.002. URL: https://linkinghub.elsevier.com/retrieve/pii/S1364815215001115.

[30] IBM. *Definition Software Development*. URL: https://researcher.watson.ibm.com/researcher/view_group.php?id=5227 (visited on 05/31/2019).

[31] IBM. *What is software development?* URL: https://www.ibm.com/topics/software-development (visited on 05/31/2019).

[32] IDC. *CSC Big Data Maturity Tool: Business Value, Drivers, and Challenges*. 2013. URL: http://csc.bigdatamaturity.com/.

[33] A. Josey. *The Open Group IT4IT™ Reference Architecture, Version 2.0*. Tech. rep. Berkshire, UK: The Open Group, 2015.

[34] J.H. Keppels. "Qualitative Measurement of BI Maturity in a SME ICT Organisation." MA thesis. University of Twente, 2018. URL: https://essay.utwente.nl/75889/.

[35] H. Kim and E. Gardner. *The science of winning in financial services: Competing on analytics Opportunities to unlock the power of data*. Tech. rep. EY, 2015. URL: https://www.ey.com/Publication/vwLUAssets/EY-the-science-of-winning-in-financial-services/$FILE/EY-the-science-of-winning-in-financial-services.pdf.

[36] D. Kiron, P. Prentice Kirk, and R. Boucher Ferguson. *The Analytics Mandate*. 2014. URL: https://sloanreview.mit.edu/projects/analytics-mandate/.

[37] B. Kitchenham. *Procedures for Performing Systematic Reviews*. Tech. rep. Keele: Keele University, 2004.

[38] M. Kontio, M. Marttila-Kontio, and V. Hotti. "Data Assessment Model for Strategic Management." In: *Communications of the IBIMA* (2015). DOI: 10.5171/2015.561648. URL: http://www.ibimapublishing.com/journals/CIBIMA/cibima.html.

[39] D. Laney. *3D Data Management: Controlling Data Volume, Velocity and Variety*. Tech. rep. 2001. URL: https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf.

[40] K. Liu, G. Pinto, and Y.D. Liu. "Data-Oriented Characterization of Application-Level Energy Optimization." In: Springer, Berlin, Heidelberg, 2015, pp. 316–331. ISBN: 978-3-662-46675-9. URL: http://link.springer.com/10.1007/978-3-662-46675-9_21.

[41] A. Mackenzie. "The Fintech Revolution." In: *London Business School Review* (2015). DOI: 10.1111/2057-1615.12059.

[42] D.A. Marchland and J. Peppard. "Why IT Fumbles Analytics." In: *Harvard Business Review* (Jan. 2013).

[43] A. McAfee and E. Brynjolfsson. "Big Data: The Management Revolution." In: *Harvard Business Review* 90.10 (2012), pp. 60–66. URL: http://tarjomefa.com/wp-content/uploads/2017/04/6539-English-TarjomeFa-1.pdf.

[44] T. Mettler and P. Rohner. "Situational maturity models as instrumental artifacts for organizational design." In: *Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology - DESRIST '09*. New York, New York, USA: ACM Press, 2009, p. 1. ISBN: 9781605584089. DOI: 10.1145/1555619.1555649. URL: http://portal.acm.org/citation.cfm?doid=1555619.1555649.

[45] NewVantage Partners LLC. *Big Data and AI Executive Survey 2019: Executive Summary of Findings*. Tech. rep. 2019.

[46] D.J. Patil. *Building Data Science Teams*. Tech. rep. Oreilly Radar, 2011. URL: www.asterdata.com.

[47] F. Provost and T. Fawcett. "Data Science and its Relationship to Big Data and Data-Driven Decision Making." In: *Big Data* 1.1 (Mar. 2013), pp. 51–59. DOI: 10.1089/big.2013.1508. URL: http://www.liebertpub.com/doi/10.1089/big.2013.1508.

[48] G. Rejikumar, A. Aswathy Asokan, and V.R. Sreedharan. "Impact of data-driven decision-making in Lean Six Sigma: an empirical analysis." In: *Total Quality Management & Business Excellence* (Jan. 2018), pp. 1–18. DOI: 10.1080/14783363.2018.1426452.

[49] G. Shaykhian, M.A. Khairi, and J. Ziade. "Factors Influence Data Management Model Selections: IT Expert Testimonies." In: *ASEE Annual Conference and Exposition*. Seattle, Washington, June 2015, pp. 26.759.1 –26.759.11. ISBN: 978-0-692-50180-1. DOI: 10.18260/p.24096. URL: https://peer.asee.org/24096.

[50] D. Stahl, T. Martensson, and J. Bosch. "Continuous practices and devops: beyond the buzz, what does it all mean?" In: *2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. IEEE, Sept. 2017, pp. 440–448. ISBN: 978-1-5386-2141-7. DOI: 10.1109/SEAA.2017.8114695. URL: http://ieeexplore.ieee.org/document/8114695/.

[51] N. Tadhg and D. Sammon. "The Data Value Map: A framework for developing shared understanding on data initiatives." In: *25th European Conference on Information Systems (ECIS)*. Guimaraes, Portugal, 2017, pp. 1439–1452. URL: http://aisel.aisnet.org/ecis2017_rp/93.

[52] T. Thakar, T. Tsultrim, L. Stapleton, and L. Doyle. "Enterprise Level Integration of Ontology Engineering and Process Mining for Management of Complex Data and Processes to improve Decision System." In: *IFAC-PapersOnLine* 51.30 (2018), pp. 762–767. DOI: 10.1016/j.ifacol.2018.11.200. URL: https://linkinghub.elsevier.com/retrieve/pii/S2405896318328659.

[53] The Economist. *The fintech revolution*. Tech. rep. 2015. URL: http://www.economist.com/node/21650546/.

[54] S. Traulsen and M. Tröbs. "Implementing Data Governance within a Financial Institution." In: *GI-Jahrestagung*. Berlin, 2011, pp. 195–210. ISBN: 9783885792864. URL: http://informatik2011.de/519.html.

[55] S.F. Wamba, A. Gunasekaran, S. Akter, S.J. Ren, R. Dubey, and S.J. Childe. "Big data analytics and firm performance: Effects of dynamic capabilities." In: *Journal of Business Research* 70 (Jan. 2017), pp. 356–365. DOI: 10.1016/J.JBUSRES.2016.08.009. URL: https://www.sciencedirect.com/science/article/pii/S0148296316304969.

[56] R.J. Wieringa. *Design Science Methodology for Information Systems and Software Engineering*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2014. ISBN: 978-3-662-43838-1. DOI: 10.1007/978-3-662-43839-8. URL: http://link.springer.com/10.1007/978-3-662-43839-8.

[57] J.F. Wolfswinkel, E. Furtmueller, and C.P.M. Wilderom. "Using grounded theory as a method for rigorously reviewing literature." In: *European Journal of Information Systems* 22.1 (Jan. 2013), pp. 45–55. DOI: 10.1057/ejis.2011.51.

[58] H. Yu and J. Foster. "Towards information governance of data value chains: Balancing the value and risks of data within a financial services company." In: *Communications in Computer and Information Science*. 2017. ISBN: 9783319626970. DOI: 10.1007/978-3-319-62698-7.

Part I

APPENDIX

A

DESCRIPTIONS DATA MANAGEMENT FRAMEWORKS

KNOWLEDGE AREA	DESCRIPTION
Data Strategy	is a capability that enables the organisation to define, maintain and drive execution of strategic direction setting on turning data into value.
Data Awareness & Education	ensures that the necessary data and Data Management awareness, knowledge and skills are embedded within the organisation in a timely and sustainable fashion
Data Principles & Policy Management	is a capability that enables the organisation to define, maintain and drive execution of guiding principles and subsequently policies on data management.
Data Accountability Management	is a Data Governance capability that enables the organisation to embed key accountabilities and responsibilities for data assets within Fortran in order to effectively manage and steer data on strategic, tactical and operational level
Data Ethics Management	is a capability that enables the organisation to define, maintain and drive a common moral compass for right or wrong data handling.
Enterprise Party Reference Management	is a capability that provides a single authoritative enterprise view for shared party master data across the organisation, therefore promoting consistent (re)use across systems and processes
Organisation Structure Data Management	is a capability that provides a single authoritative source and maintenance for acknowledged enterprise organisation structures, therefore promoting consistent (re)use across systems and processes

Enterprise Reference Data Management	is a capability that provides a single authoritative source and maintenance for acknowledged enterprise reference values across the organisation, therefore promoting consistent (re)use across systems and processes
Reference & Master Data Management	is a capability that enables the capturing, storage, access, and use of data and information in individual data sources is aligned whenever data is shared. This helps to avoid incorrect use of data, reduce risks associated with data redundancy, ensure higher quality, and reduce the costs of data integration.
Data Profiling and Monitoring	is a Data Source Management capability that enables active monitoring of critical data through profiling, data quality checks and dash boarding, ensuring that data quality incidents and issues are identified in a timely manner
Document & Content Management	is a capability that enables the capturing, storage, access, and use of data and information stored outside relational databases. Its focus is on maintaining the integrity of and enabling access to documents and other unstructured or semi-structured information.
Data Access Management	is a capability that empowers everyone within the bank to access data in a secure, agile and fast way. This is being enabled by our data distribution platform – the one stop shop for finding and accessing data
Data Sharing Agreement Management	is a capability that enables the organisation to govern and mutually agree on the sharing of data between Data Owner and Data User regarding accountabilities, intended use, conditions, service level agreements and metrics
Data Lineage Management	is a capability that enables the capturing, management and visualisation of the origins, movement, characteristics and transformations of data as it moves through the organisation across various systems, processes and people
Self Service BI Management	is a capability that empowers the organisation to interact with (big) data, discover patterns and insights and present these insights in a comprehensible way

Advanced Analytics Management	is a Data Value Creation capability that enables (semi-) autonomous examination of data using sophisticated techniques and tools to discover deeper insights, make predictions, generate recommendations and make business decisions
Managed Business Intelligence	is a capability that provides the ability to provide accurate data, reports and dashboards to support fact-based decision making and report to (external) stakeholders via a dedicated service organisation.
Next Gen. Data Visualisation	is a capability that enables the organisation to visualise data using interactive methods, multi-dimension views and animation in order to present vast amounts of heterogeneous data in a timely, relevant and comprehensible fashion – including storyboarding.
Data Quality Issue Management	is a Data Foundation capability that provides a standardised process for managing and monitoring issues throughout the entire issue lifecycle with the ultimate goal of permanently solving issues at the source with sustainable solutions.
Data Accountability Catalogue	is the overview of common data sets each with a specific business context, and maintained in a unique (common) source, with a dedicated data owner
Business Data Modelling	enables the organisation to describe, capture and share knowledge about data in a business meaningful manner. This knowledge can be described from a business specific perspective and is captured in terms, definition, relations and collections
Data Modelling & Design	enables the organisation to describe, capture and share information about data on a logical and physical level

Table A.1: Descriptions of the Fortran Data Management Framework

KNOWLEDGE AREA	DESCRIPTION
Data Governance	provides direction and oversight for data management by establishing a system of decision rights over data that accounts for the needs of the enterprise.
Data Architecture	defines the blueprint for managing data assets by aligning with organizational strategy to establish strategic data requirements and designs to meet these requirements.
Data Modeling and Design	is the process of discovering, analyzing, representing, and communicating data requirements in a precise form called the data model.
Data Storage and Operations	includes the design, implementation, and support of stored data to maximize its value. Operations provide support throughout the data lifecycle from planning for to disposal of data.
Data Security	ensures that data privacy and confidentiality are maintained, that data is not breached, and that data is accessed appropriately.
Data Integration and Interoperability	includes processes related to the movement and consolidation of data within and between data stores, applications, and organizations.
Document and Content Management	includes planning, implementation, and control activities used to manage the lifecycle of data and information found in a range of unstructured media, especially documents needed to support legal and regulatory compliance requirements.
Reference and Master Data	includes ongoing reconciliation and maintenance of core critical shared data to enable consistent use across systems of the most accurate, timely, and relevant version of truth about essential business entities.
Data Warehousing and Business Intelligence	includes the planning, implementation, and control processes to manage decision support data and to enable knowledge workers to get value from data via analysis and reporting.
Metadata	includes planning, implementation, and control activities to enable access to high quality, integrated Metadata, including definitions, models, data flows, and other information critical to understanding data and the systems through which it is created, maintained, and accessed.
Data Quality	includes the planning and implementation of quality management techniques to measure, assess, and improve the fitness of data for use within an organization.

Table A.2: Descriptions of the DAMA-DMBOK capabilities [17]

B

DATA CHALLENGES IN IT

ABBREVIATION	TERM
P	Problem
C	Cause
S	Solution
DM	Data Management Capabilities

Table B.1: Legend for Data Challenges

1	
P	It is not always possible to create useful insights from the data about incidents and the tool pipeline.
C	(1) Data cannot be linked with each other because of data model mismatch. (2) There is not enough staff which can interpret technology as well as understand business needs.
S	(1) Redesign data model. (2) Train staff, hire staff or offshore work.
DM	Business Data Modelling, Data Modelling & Design

2	
P	It is difficult to create high level relationships between incidents and technological issues and it is not possible to correlate machine data with manually reported incidents.
C	(1) Not possible to trace incident back to an earlier stage in the IT value chain. (2) Incidents are incorrectly registered.
S	(1) Perform analytics on data from IT Value Chain tooling combined with incident reports. (2) Improve Data Governance for incident data.
DM	Data Quality Issue, Data Accountability Catalogue, Data Accountability

3	
P	There is a chance that not all security checks are performed.
C	(1) It has not been manually registered what has been changed in the software. (2) It has not been manually registered who has been provided access to an application.
S	Educate about and show the value of this registration and/or retrieve data automatically from the IT value chain
DM	Data Value Presentation, Automatic Data Generation

4	
P	It is not always known which description of an IT product is correct.
C	(1) There are too many different versions (2) People without the right expertise create those descriptions. (3) It is unknown if a data source is outdated. It is unknown if a data source is trustworthy.
S	Appoint a Common Source and introduce accountability. Create controls on who may create or edit a data source about IT products.
DM	Data Accountability, Data Accountability Catalogue

5	
P	It is not always known what leads to a breaking system and it is too hard to prevent system failure.
C	(1) It is unknown which IT components relate to other IT components. (2) It is unknown how components behave before a system will break down. (3) Not enough knowledge about system performance.
S	(1) Model the components throughout IT Value Chain. (2) Perform predictive analytics on data from monitored infrastructure and applications.
DM	Business Data Modelling, Advanced Analytics, Metrics Creation

6	
P	Incidents cannot be solved automatically
C	(1) They are reported manually and still need to in the future. (2) Not all incident registrations are of good quality
S	(1) There might be an opportunity for performing machine learning on the combination of incident reports and the machine data. (2) Show the value of the usage of the data to increase effort in the quality of the data.
DM	(1) Advanced Analytics (2) Data Value Presentation

7	
P	The impact of changes on IT applications is not always clear, it may break other applications
C	There is no complete overview of the links between systems.
S	(1) Model the relationships between systems on semantic and logical level (2) Disconnect applications to remove dependencies.
DM	Business Data Modelling

8	
P	Unnecessary access security risks might be introduced
C	Tokens could be longer valid than is needed.
S	Consume application data to check active usage of access tokens.
DM	Automatic Data Generation

9	
P	Not all problems are registered correctly.
C	People skip through forms which leads to data quality issues.
S	Create awareness of data quality issues and show the value, or create mandatory fields in forms which enforce data quality.
DM	Data Awareness & Education, Data Quality Issue, Data Value Presentation

10	
P	Registration of changes is done unnecessarily, since the data is already available elsewhere.
C	Code is not linked with registration of changes.
S	Retrieve change log from the tools the developers use to register changes instead of supplying an extra tool for the registration.
DM	Automatic Data Generation

11	
P	It is not always clear how architectural data sources relate to each other.
C	The file formats are incompatible and there is no agreed file format standard.
S	Agree on a standardised data format and model relations.
DM	Data Modelling & Design

12	
P	It is sometimes too hard to create connections with people who experience problems.
C	(1) Not enough is done to make a connection with people. (2) Different languages are spoken between business and IT
S	Create a common data model in which the business and IT terms are aligned.
DM	Business Data Modelling

13	
P	Descriptions of applications across the organisation are incompatible, but is necessary since business and IT move to integration.
C	(1) Different terminology is used. (2) They have a different vision on how it is used (3) Mismatch between Business and IT
S	Create a common data model in which the business and IT terms are aligned.
DM	Business Data Modelling

14	
P	It is unknown if a change in the software is linked to changing customer experience.
C	(1) Not possible to trace data at the end of the value chain back to development. (2) Data sources are not linked with each other.
S	Create data model on physical level, create link with monitoring data with changes made in the software and provide advanced analytics to find trends in the data.
DM	Business Data Modelling, Data Modelling & Design, Advanced Analytics

15	
P	Wrong causes are sometimes pointed out during incidents and incorrect assumptions are made during crisis situations.
C	The IT landscape looks different than thought.
S	Model the relationships between IT landscape components and use machine data to objectify the situation.
DM	Business Data Modelling, Data Modelling & Design, Advanced Analytics

16	
P	Wrong datasets are used for driving decisions.
C	Decisions are made based on outdated data, people unawarely use obsolete data, people do not follow the process
S	Introduce better Data Governance.
DM	Data Awareness & Education, Data Accountability, Data Accountability Catalogue, Data Quality Issue Management

17	
P	It cannot be completely explained to the regulator which interfaces are connected to an application and data about it cannot always be trusted.
C	There are too many different lists about how the IT landscape looks like.
S	Automate the detection of the IT Landscape, Improve Registration process
DM	Automatic Data Generation, Data Awareness & Education, Data Accountability, Data Accountability Catalogue, Data Quality Issue Management

18	
P	A share of decisions are made based on assumptions, which are not always correct and wrong decisions are made because no data is used.
C	People decide on what they already know, they do not know what to do with data, it is felt as boring, data might feel confronting.
S	Create a data culture by showing the value of the data, such that people challenge their intuition with data and integrate data in decision making process.
DM	Data Awareness & Education, Data Value Presentation

19	
P	It is hard to determine what incidents actually cost
C	No traceability of the IT Value chain makes it difficult to determine which IT components are affected during an incident and it is hard to measure the impact besides direct losses.
S	Increase traceability across IT Value chain by creating a common data model.
DM	Business Data Modelling, Data Modelling & Design

20	
P	It is not completely known how money is spent in IT exactly and knowledge about how costs can be reduced could improve.
C	The IT landscape is not clear
S	Automate IT Landscape data generation and create a common data model such that relations in the IT Landscape become clearer.
DM	Automatic Data Generation, Business Data Modelling

21	
P	It is unclear what a change in software will cost or has cost exactly.
C	(1) It cannot be measured, it is put on one heap in portfolio- and project plans. (2) No total image of IT development and IT operations.
S	Bring development and operations together in DevOps. Make IT landscape traceable with common data model.
DM	Business Data Modelling, Data Modelling & Design

22	
P	It is hard to measure the value of IT is.
C	IT value is also intangible, value comes after a long time.
S	Create shorter delivery cycles, introduce metrics and use analytics to measure value.
DM	Self-Service Business Intelligence, Advanced Analytics

23	
P	Teams do not always spend their time efficiently.
C	They are not certain know how their time is spent.
S	Track data from CICD pipeline and provide the teams analytics tools.
DM	Self-Service Business Intelligence, Advanced Analytics

24	
P	It is hard for management to determine how teams perform.
C	There is no insight in the performance data of the teams.
S	Track data from CICD pipeline, set KPIs and perform analytics.
DM	Self-Service Business Intelligence, Advanced Analytics

25	
P	Team performance cannot always be measured.
C	Teams did not implement a standard tooling pipeline, teams do not want to implement a standard tooling pipeline, manually given data is not reliable, there are too many tools used which to be able to extract metrics about teams.
S	Show the value of the data that can be retrieved from the standard tooling pipeline, such that the teams are willing to implement it.
DM	Value Visibilty

26	
P	It is hard to determine why some teams perform better than others.
C	(1) The time spent on which types of work is unknown. (2) Team characteristics are unknown (composition, skills, technical debt, etc.).
S	(1) Track data from CICD pipeline and perform analytics. (2) Improve data management on team characteristics data.
DM	Advanced Analytics, Data Accountability Catalogue, Data Awareness & Education, Data Accountability, Data Quality Issue Management

27	
P	It is hard to determine what the underlying issue of low system performance is
C	The time spent on which types of work is unknown.
S	(1) Create traceability throughout IT Value Chain to detect trends. (2) Show importance of registering changes or automatically retrieve this type of data.
DM	Business Data Modelling, Data Modelling & Design, Data Value Presentation, Automatic Data Generation

28	
P	It is hard to determine what underlying issue within IT teams are.
C	Incidents are not registered, registration is felt as overhead, there is a lack of awareness of the value of incident data.
S	Raise awareness of the value of the data and the need for better quality data.
DM	Awareness & Education, Data Value Presentation

29	
P	Business owners in of IT departments sometimes have difficulty to understand how the departments performs.
C	(1) Business and IT speak a different language. (2) Business owners do not have enough knowledge about IT.
S	(1) Create a clear Data Model and provide descriptions for data definitions. (2) Educate the business about IT data.
DM	Business Data Modelling, Awareness & Education

30	
P	It is not always clear why data should be shared with each other.
C	It is unclear what happens with data once it is shared.
S	Make the usage of the data visible.
DM	Data Value Presentation

31	
P	Potential cases to turn data into value are left unused.
C	(1) Raw IT data is not useful and too much to share. (2) The usage of the data sharing platform by IT is quite low. (3) Operational data could not be accessed before DevOps.
S	(1) Create exits in data which might be useful (2) Make use of data sharing and ownership capabilities within the organisation.
DM	Data Distribution, Data Accountability Management, Data Accountability Catalogue

32	
P	There is a lack of trust in data.
C	(1) External factors are blamed for it. (2) The quality of the data is not in order (3) Data does not contain all wanted fields.
S	Make use of a Common Source, make use of Data Management Capabilities to increase quality of the data, ensure quality of data by presenting how it is used.
DM	Data Accountability Catalogue, Data Quality Management, Data Accountability Management, Data Distribution, Data Value Presentation

33	
P	There is a lack of trust in the usage of team performance data.
C	Developers fear consequences when their performance is monitored.
S	Managers need to consider how they will use the data ethically and it needs to be explained how the data is used.
DM	Data Ethics, Data Value Presentation

C

IDENTIFIED DATA SOURCES WITHIN IT

C.1 IT LANDSCAPE DATA

C.1.1 *DevOps teams*

1. Registration teams in department (Master data)
2. Registration members in team (Master data)
3. Registration skills members in team
4. Registration responsibilities of team (Master data)
5. Registration members from vendor which came from IT Operations

C.1.2 *Application and Application Components*

1. Registration application (metadata about application) (Master data)
2. Registration status of application
3. Registration dependencies applications and application components (Master data)
4. Registration configuration of application
5. Registration ownership of applications and application components (which team is responsible)
6. Registration usership of applications and application components
7. Relations between producing and consuming applications on the organisational data sharing hub

C.1.3 *Infrastructure*

1. Registration of infrastructure (Master data)
2. Registration of infrastructure product descriptions
3. Registration ownership of infrastructure (which team is responsible)

C.1.4 *IT Products*

1. Designs of IT products
2. Requirements of IT products
3. Architecture of IT products

C.2 TRANSACTIONAL DATA

C.2.1 *IT Value Chain data*

C.2.1.1 *Development data*

1. Registration changes development
2. Productivity based team data (story points)

C.2.1.2 *Identity & Access Management*

1. Registration of access rights
2. Registration of digital identities
3. Registration of password vaults

C.2.1.3 *Incident & Problem management*

1. Registration incidents
2. Registration problems

C.2.1.4 *Strategic initiative data*

1. Registration of checklist DevOps transition for teams (based on maturity model)
2. Registration of checklist CICD tooling pipeline for teams

C.2.2 *Machine data*

1. Log data development tooling
2. Log data operational tooling
3. Log data monitoring infrastructure (event management)
4. Data from synthetic monitoring
5. Performance management data (of running infrastructure)
6. Costs of running infrastructure

7. Registration software licences in use
8. CICD metrics of DevOps teams, e.g. about:
    a) Deployments
    b) Changes
    c) Errors
    d) Availability

C.3 METADATA

C.3.1 *Strategic alignment data*

C.3.1.1 *IT Definitions*

1. Registration data definitions and relations
2. Registration IT definitions