Master Thesis

Data Driven Organizations: A Framework for Assessing Data Maturity in Large Enterprises

Ulrike Stefanie Klein

Subject Area: PMBA, Digital Transformation and Data Science

Supervisor: Dr. Sabrina Kirrane

Date of Submission: June 30$^{\text{th}}$ 2021

WU Execuitve Academy, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

Contents

List of Figures

List of Tables

Abstract

The research work is conducted to find out how large enterprises can transform towards data driven organizations and leverage the ever increasing amount of data. The underlying research question is framed to understand the strategic requirements, organizational structures, and technological requirements needed to become a data driven company. In order to enhance the research a fictive large enterprise named OeTech is in the center of the research's use case. The research is grounded on an integrative literature review. The literature review evaluates indicators and enablers of data driven organizations. Furthermore existing data maturity frameworks are analyzed and discussed. The conclusion shows that available data maturity frameworks do not fit the purpose for large manufacturing enterprises such as OeTech. Therefor, the existing frameworks are synthesized and further developed using the structured case methodological framework. The result is a data maturity assessment for large enterprises (DALE) which is then applied using the example of OeTech. Based on the assessment results a short, mid, and long term roadmap is developed for OeTech to become a data driven organization.

1 Introduction

1.1 Motivation

In 2018, the amount of worldwide data, the Global Datasphere, has reached 33 zettabytes and is predicted to reach 175 zettabytes in 2025. Enterprises will generate more than 60% of the Global Datasphere [35]. For enterprises, these data can unlock huge potentials. The key is to turn high volumes of fast-moving data into meaningful insights [12]. A 2013 study from Bain & Co [31] examined 400 large companies and found that the ones with the most advanced analytics capabilities are outperforming their competitors. According to the research, firms that use advanced data analytics are twice as likely to be in the top quartile of their sectors' financial performance. They are five times more likely than market peers to make quick judgments. They are also three times more likely to carry out choices as planned. Companies that utilize sophisticated analytics are twice as likely to use data regularly when making choices.

Five years after Bain & Co's study, Ulrich [43] compares 16 existing surveys with large enterprises and SMEs around big data and data analytics. The conclusion drawn from this examination shows that companies worldwide see data analytics as an important topic, regardless of their industry sector or company type. For the enterprises observed in these surveys, it is common sense that the potential and impact of data-driven decision-making are high. This is an urgent call for action to operationalize data analytics. A vast majority of organizations already collect data internally, but they fail to analyze it. According to Groggert et al. [16], manufacturing companies only use 5.5% of their available data to apply data analytics. Furthermore, Goekalp et al. [14] highlight that the utilization and installation of available data analytics platforms require significant knowledge and expertise in IT and data science. The lack of knowledge and expertise hinders the adoption of big data technologies and, ultimately, data-driven decision-making. Moreover, enterprises face issues when it comes to managing the value and the quality of the companies data [28].

The potential of data-driven decision-making, as stated by Bain, is not fully exploited and leaves room for optimization. The existing gap between the status quo and the technical possibilities does not stem from missing systems for data analytics or the establishment of basic technical infrastructure but because of the missing strategic alignment when analyzing data. The insights call for structured approaches to leverage data-driven decision-making.

1.2 Problem Statement

As highlighted above, many organizations have not yet set up a data strategy or leverage data analytics to gain insights. OeTech is representative of a manufacturing company. While the amount of data generated during the manufacturing process is increasing year after year, the concept of data utilization is under development. As for many other manufacturing companies, selling data or utilizing data for marketing is not OeTech's core competence. The goal is to deliver high qualitative and highly reliable products to industrial customers. For instance, the management of OeTech is aware of the potential in the data and has dedicated a budget to enable data utilization as part of their digital transformation strategy. However, companies like OeTech might face a challenge now that they are unsure what adjustments are needed for their organizational structure and what kind of infrastructure OeTech needs to invest in. Further, OeTech is not aware of the required employee skill-sets to transform the company into a data-driven organization. Therefore, enterprises such as OeTech need guidelines that guide the company to become a data-driven organization.

1.3 Aim of the Work

With the ever-increasing amount of data and the possibilities of utilizing data, there is a call for action. Therefore this work aims to create a guideline in the form of a roadmap for companies such as OeTech who want to become more data-driven. The first step is to understand the definition and enablers of a data-driven organization. The second step is then to investigate how the data-drivenness of organizations can be measured. The third step is to apply a framework to identify the status quo of OeTech as a representation of a non-data-driven company. Finally, the fourth step is to take the insight from the status quo assessment and develop a roadmap that supports the approaches of large enterprises and SMEs to become data-driven organizations.

1.4 Structure of the Work

Chapter 2 (Theoretical Foundation) describes the use case scenario of OeTech on which this thesis is grounded. Next, we present the overarching and underlying research questions we are answering in the thesis. Finally, we introduce the used methodologies. On the one hand, the integrative literature review and on the other hand the structured case methodological framework.

Chapter 3 (Literature Review) presents indicators as well as enablers for data-driven organizations. We give an overview of existing data maturity frameworks. Furthermore, we compare the dimension and levels assessed in these different frameworks and describe the dominating dimensions in the selected frameworks in detail. This chapter is concluded with a discussion on the fit for purpose of the selected frameworks.

Chapter 4 (Structured Case: Data Maturity Framework) is build upon the literature review. We show how we build the Data Maturity Framework for Large Enterprises (DALE) using the structured case methodological framework. In the first step, we explain the methodology. In the second step, the conceptual frameworks evolve due to iterating research cycles. In the third step, we analyze the conceptual frameworks and present the resulting framework, DALE. Finally, we critically examine DALE's dimensions, complexity, and practical usability.

Chapter 5 (Application and Roadmap) describes the result of the application of DALE using the example of OeTech. Finally, we outline the status quo and conclude with three different roadmaps (short, mid, and long-term) for OeTech to become a data-driven organization.

Chapter 6 (Conclusion and Future Work) produces a critical assessment of the work presented in this thesis and introduces several directions for future work.

2 Theoretical Foundation

This section presents the use case scenario and the research question underlying the analysis.

2.1 Use Case Scenario

OeTech is a fictional representative for a global operating manufacturing company with more than 50000 employees worldwide. The diverse manufacturing locations are located in Europe, South America, and Asia. The produced goods are electronic components for various industries. The company stands for high quality and the industrialization of high-end technology. OeTech understands that the data generated in the manufacturing and other processes have a huge potential for improving the organization's overall productivity. Research into the company's internal human resource database, which contains, for example, the name, job title, and department, shows no employee with the job title *Data Analyst*, *Data Engineer*, *Data Architect*, or *Data Scientist*. Only one employee with the title *Data Manager*.

Screening the organizational chart of OeTech further shows no department for data science or data analytics at any worldwide manufacturing locations. Instead, the *Data Manager* is a function in the headquarters within a department responsible for internal consulting. In the scope of a digital transformation program, OeTech managers conducted two different surveys. The first survey assessed worldwide ongoing and planned projects within the organization with a focus on digitalization. The recent conclusion is that company-wide, around 30 projects out of 160 collected projects are either related to data analytics, data management, or data storage. The second survey assessed, among others, how decisions are taken within OeTech. 73%

stated that the majority of decisions are based on previous experience. Given the fictional exemplary description of OeTech, there is awareness for data utilization. Still, according to the job titles, organizational structure, and executed projects, changes are required to transform companies like OeTech into data-driven companies. Therefore, organizations like OeTech need guidelines. The main use case will show how to go from the illustrated as-is situation described above towards the beginning of a transformation towards a data-driven company.

2.2 Research Questions

As described in the use case scenario, there is a need for guidelines on becoming a data-driven manufacturing company. Derived from this need, the overarching research question for this thesis is:

Which strategic requirements, organizational structures, and technologies are required to create a roadmap towards becoming a data-driven organization?

The research question is further split into three sub-questions.

(1) *How can data-driven companies be quantified under consideration of company culture, organizational structures, and available skills?*

(2) *How can existing frameworks be used to understand the status quo and analyze the gaps towards existing trends?*

(3) *Which initiatives are needed to improve the data-drivenness of an organization?*

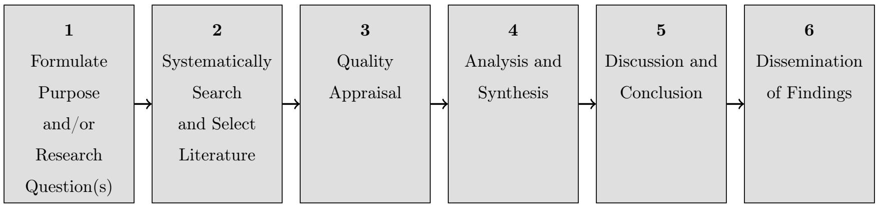

Figure 1: Six steps of the integrative review process [41]

2.3 Methodology

In this thesis, two different methodologies are applied to answer the research questions.

2.3.1 Integrative Literature Review

The integrative literature review is a method for the review of existing literature. Compared to the systematic literature review, the integrative review looks more broadly at a phenomenon of interest than a systematic review. It allows for diverse research containing theoretical and methodological literature to address the review's aim [36]. The integrative literature review has several benefits for the scientific reviewer. Assessing the level of scientific evidence, finding gaps in existing research, and recognizing the need for future study are just a few of the advantages [41]. According to Toronto et al. [41], the integrative literature review consists of six process steps which are presented in Figure 1. In the first step (1), one or more research questions need to be formulated. Based on the research question, the literature is systematically searched, and literature is selected (2). In the third step (3), a quality appraisal of the selected literature is conducted. The quality appraisal of the literature is important for the strictness of the review. The purpose and questions of the review should be dominant to guide the process. In step four (4), the selected literature is analyzed and synthesized. Therefore data are extracted into tables. Based on the research question, similarities and differences are investigated for existing patterns. These patterns are then synthesized. In this step, the focus moves from a facts-based problem towards a conceptual level of knowledge. In step (5), the findings and the meanings of the findings are outlined. The findings are compared and contrasted with existing literature. Furthermore, recommendations and implications for research and practice are made. Finally, the conclusion will summarize the major findings and the contributions to the state of the science. The final step of the integrative literature review is the dissemination, where research results are presented to a scientific audience [33].

2.3.2 Structured Case Methodological Framework

The structured case is a methodological framework developed by Carroll and Swan [6]. It was particularly designed for information systems research to build theory from qualitative data gathered in the field.
The structured case is an overall framework that includes three main elements: the conceptual framework, the research cycle to build theory, and the literature-based scrutiny of the theory building [6]. The structured case methodical framework is visualized in Figure 2.

Conceptual Framework. The starting point of the structured case is the development of a conceptual framework that is grounded on literature. The initial conceptual framework depicts the current understanding of the researcher. The conceptual framework is further formed throughout the course of the research. At the end of each research cycle, it is critically examined and updated to incorporate new information and insights regarding the research theme concerns. It then acts as the

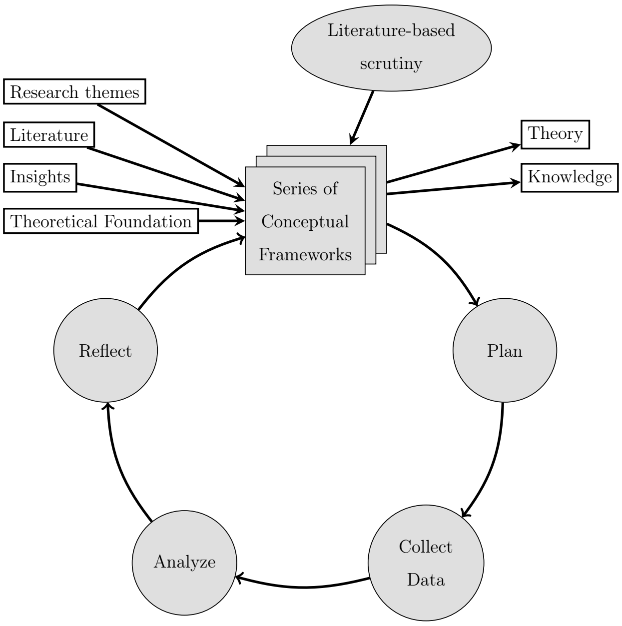

Figure 2: Structured case methodological framework [6]

foundation for a subsequent research cycle that will deepen the comprehension of the research theme. As a result, the conceptual framework is made up of iterative improvements of the conceptual framework.

Research Cycle and Theory Building. The structured-case research cycle is divided into four stages. In the following, the four stages are described as inclusive and separate. Nevertheless, the stages are fluid, and the movement through the research cycle does not follow a defined, sequential pattern.

Plan. The investigation of the research is planned. Planning steps include the selection of cases, organization, and informants, such as possible interview partners. Further, the methods to collect, record, process, and analyze data are planned. The planning stage serves as guidance within the research cycle because qualitative research must be responsive and cannot prescribe the research activities.

Collect Data. The initially designed plan guides the data collection. In qualitative research data, collection and analysis might have strong overlaps. During the data collection, the research analyzes and examines the results and can react to opportunities or changes in the plan. For instance, the researcher can add additional questions in an interview in the course of the survey.

Analyze. The analysis is an iterative, ongoing activity. The goal is to get a deep understanding of the collected data to generate knowledge and theory.

Reflect. Within the reflection phase, the research reflects upon the previous steps. The planning, the way how we collected data, and the results of the analyzes. These gained insights are then incorporated into the next iteration of the conceptual framework. The reflection phase marks the end of one research cycle.

The theory is built from multiple cases in the structured case to enrich and revise the conceptual framework sequentially. The methodology is closely linked to practice. The fieldwork causes theory development which leads to further research in the field. The result is that the theory developed reflects the actions, problems, and issues of practitioners.

Literature-based Scrutiny of the Theory Built. In theory, the development of the conceptual framework never finishes; in practice, the framework is finished once the researcher realizes that no further insights or knowledge can be derived in another research cycle. After the final conceptual framework is completed, the built theory needs to be compared and contrasted with a broad range of existing literature. Within the scrutiny, the researcher looks out for agreements and conflicts with the existing literature. Finally, all findings need to be explained and interpreted. This final element in the structured case may lead to a critical reevaluation of the results or a re-examination of the data with new findings that raise the built theory to a more abstract level and increase the applicability of the theory to further contexts. Expanding the existing literature and reconciling it with conflicting literature indicates the end of the research process.

3 Literature Review

According to *google trends*, the search term *data-driven organization* only became known to the Internet in 2005, when the trend curve immediately peaked [42]. The curve flattened to almost zero in 2012. Since 2013 *data-driven organization* has gained an increasing search frequency. We are investigating an expression that is not much older than 15 years; it presents an occasion to dive deeper into the understanding of data-driven organizations.

3.1 Data-driven Organizations

When looking for examples of data-driven organizations, the Internet mainly points out the gang of four: Google, Amazon, Facebook, and Apple, commonly referred to as GAFA. Their revenue relies on the usage of data. They utilize data to make their products and services even better. Especially Google's and Facebook's core competence lies within the evaluation of data. Google, for instance, made up to 70.9% of its revenue by advertisement in 2019 [39]. Advertisers provide Google advertisements with a list of keywords relevant to a product, service, or company. Once a Google user searches the Internet using one or more of these keywords, the advertisement appears in the sidebar. Advertisers are charged by Google each time a user clicks on their ad. Users are then directed towards the advertiser's site. The better the algorithm works that selects significant advertisements for the user, the higher is the benefit for Google [10].

For organizations such as OeTech, the core competence is manufacturing. Yet, the motivation for OeTech is high to advance towards becoming a data-driven organization.

In the next subsection, indicators that define a data-driven organization are investigated. For example, it is shown how data-driven companies handle their data.

3.1.1 Indicators for Data-driven Organizations

In the first step, we look at definitions of a data-driven organization as described in the literature. Therefore, we analyzed five different contributions published between 2006 and 2018: Davenport et al. [9], La Valle et al. [20], Patil et al. [30], Anderson [2], O'Neal [27] and Berndtsson et al. [3] investigated data-driven organizations. The definitions and indicators they outline are split into four aspects:

(i) *Data handling:* how data-driven organizations handle data.

(ii) *Data usage:* how data-driven organizations make use of their data.

(iii) *Focus area:* what data-driven organizations focus on.

(iv) *Application area:* in which areas of the organizations data-driven companies establish data-driven processes.

An overview of each aspect is presented below. In addition, a summary of the qualitative indicators is depicted in Table 1.

Data Handling [9], [30], [2], [27] Davenport [9] describes a data-driven company as one that constantly churns its data. Patil et al. [30] highlight that data-driven enterprises acquire, process, and leverage their data. Anderson [2], a data-driven organization acquires data by continuously testing. O'Neal [27] points out that data-driven companies value, manage and protect their data.

Data Usage [9], [20], [30], [2], [3] In 2005 Davenport described a data-driven organization as one that uses data to solve business problems [9]. In 2011 La Valle et al. [20] extended this aspect by pointing

Table 1: Summary of indicators for data-driven organizations

Author	Data Handling	Data Usage	Focus Area	Application Range
Davenport 2005 [9]	churning constantly	solving business problems	building right culture, hiring right people	multiple
LaValle et al. 2011 [20]	-	creating strategies, guiding day-to-day operation	-	widest possible
Patil et al. 2013 [30]	acquiring, processing, leveraging	creating efficient iterations, developing new products	-	-
Anderson 2015 [2]	testing continuously	choosing future options.	improving continuously, learning from failures	-
O'Neal 2017 [27]	valuing, managing, protecting.	-	investing and embodying in data-drivenness	-
Berndtsson et al. 2018 [3]	-	testing, experimenting, outweighing opinions	accepting and learning from failures	-

out that data-driven companies use their data to create strategies and to guide their day-to-day operations. Patil et al. [30] highlight that data are used to create efficient iterations and to develop new products. Anderson [2] mentions that data-driven enterprises use data in their decision-making process to choose among future options. Finally, Berndtsson et al. [3] specify that data-driven companies use their data to test, experiment, and ultimately outweigh opinions not based on data.

Focus Area [9], [2], [27], [3] Davenport [9] highlights that data-driven organizations focus on building the right culture and hiring the right people to become data-driven organizations. Improving continuously and learning from failures is the focus area that Anderson [2] highlights, which is similar to Berndtsson et al. [3]. Berndtsson et al. also describe accepting and learning from failure as an indicator. O'Neal [27] point out that data-driven organizations focus on investing and embodying data-drivenness.

Application Range [9], [20] The application range is rather limited and generic described aspect. For example, Davenport [9] states that being data-driven applies to multiple business problems. In contrast, La Valle et. [20] point out that a data-driven organization utilizes data in the widest range possible.

Over time the observation is made that in 2005 the approach towards data usage is superficial, with Davenport [9] being very generic when saying that data-driven companies use data to solve business problems. However, over time the explanation of how data-driven companies use data gained more depth. In 2018 Berndtsson et al., [3] for instance, highlighted that data-driven organizations use data to test, experiment, and outweigh opinions.

All five contributions present different aspects of data-driven organiza-

Table 2: Summary of enablers towards a data-driven organization

Author	Strategy	Organization	Technology	Data Management
Davenport 2005 [9]	right focus	culture	right technology	-
Anderson 2015 [2]	decision making	culture, data leadership, organization, people	-	-
Pentek et al. 2018 [32]	-	people, roles, responsibilities	data applications, data architecture	processes and methods
Berndtsson et al. 2018 [3]	decision process	organization, management	tools	data
Sejahtera et al. 2018 [37]	perceived benefits	people, organizational support	system support	-
Andersen et al. 2018 [1]	-	managers, organization	-	-

tions. There is no unified definition available that explains a data-driven organization.

3.1.2 Enablers for Data-driven Organizations

For companies such as OeTech, it is important to understand what actions will enable them to become data-driven organizations. From the literature, five publications were selected which describe these enablers: Davenport [9], Anderson [2], Pentek et al. [32], Berndtsson et al. [3], Sejahtera et al. [37] and Andersen et al. [1].

In this thesis, the enablers are structured and clustered into four different dimensions. These dimensions are *Strategy, Organization, Technology*, and *Data Management*. For instance, in the dimension *Organization* Andersen et al., [1], state that managers need to commit to becoming a data-driven company to enable a data-driven organization. Table 2 summarizes the enablers for each of the four dimensions.

Strategy [9], [2], [3], [37] In 2005 Davenport [9] highlighted that to become a data-driven enterprise, the organization must set the right focus. Meaning that resource-intensive efforts need to be directed and that several functions together need to be combined to serve an overarching strategy.

Anderson [2] highlighted 10 years later that organizations need to incorporate data in their decision-making progress. The key is to be question and decision-focused, rather than data-focused. It is important to utilize the available and relevant data and not to rely on intuition alone.

This is similar to the statement of Berndtsson et al., [3] who also highlights the decision-making process as a facilitator. In detail, they state that the leaders in an organization should not have an instinct-based veto over insights generated from data. Establishing a data-driven culture represents a major shift in how decisions are made in most businesses.

Sejahtera et al. [37] point out that focusing on the perceived benefits of data utilization is an enabling factor towards becoming a data-driven organization.

Organization [9], [2], [32], [3], [37], [1] All five publications highlight the dimension *of Organization* as an enabling factor.

In 2005 Davenport [9] highlights the right people as a key enabler, meaning that organizations should hire the best analytical people. Further, Davenport [9] states that it is crucial to building a company-wide culture that respects measuring, testing, and evaluating quantitative evidence.

Anderson [2] also highlights an organization's culture as an enabling factor while focusing more on accessibility and data sharing. In detail, Anderson [2] states that an organization needs to provide broad access to data. A culture needs to establish where data is shared with staff outside the core analytics organization and business units, teams, and individuals. Furthermore, Anderson [2] points out data literacy and data leadership as enablers. Decision-makers and managers of an organization need to be data literate. A data-driven organization requires a strong top-down data leadership, where the leadership needs to promote a data-driven culture.

Pentek et al. [32] point out people, roles, and their responsibilities as enablers. To achieve efficient data management and uniform data use across the company, responsible managers must establish the skills and organization.

Berndtsson et al. [3] describe management and the organization itself as enablers. The senior management must be actively involved in the development of a data-driven cultural plan. Furthermore, regardless of where advanced analytics is located inside the business, the IT unit must adjust its focus to provide simple access to data for all workers.

Sejahtera et al. [37] focus on people and organizational support. Like Berndtsson et al. highlight, Sejahtera et al. also state that the top management needs to support a culture of collaboration. Stated brief is further than technical skills, good working attitudes, and the right people are enablers as well.

For Andersen et al., [1] the key enablers are managers and the organizational structure. A data-driven organization requires bilingual managers that speak both: machine and human. The organizational design needs to be embedded in the rules and protocols for interaction rather than in a fixed structure.

Technology [9], [32], [3], [37] In 2005 Davenport [9] highlighted using the right technology as an enabler, meaning that data-driven organization investigate the latest statistical algorithms and decision science approaches and push the frontiers of IT.

13 years later, Pentek et al. [32] highlight data applications and data architecture as key enablers. Organizations need to define the conceptual data model, specify which data is stored in which application, and describe how data flows between applications. Further software components supporting data management activities need to be defined.

In the same year, as Pentek et al. [32], Berndtsson et al. [3] outline, employees need to be allowed to use any tool to develop a dashboard related to their daily work.

Sejahtera et al. [37] briefly outline tools and system support as key enablers for a data-driven enterprise but do not specify in more detail.

Data Management [3], [32] Among the sources presented, Pentek et al. [32] and Berndtsson et al. [3] are the authors who highlight an enabling factor in the dimension of data management.

Berntsson et al. [3] describe that strong data governance and access to good quality data are mandatory to enable a data-driven organization.

Pentek et al. [32] explicitly highlight processes and methods as an enabling factor for a data-driven company. Procedures and standards for maintaining and utilising data in a consistent and efficient manner must be established.

We conclude that different aspects enable data-driven organizations. Based on the existing literature, we can summarize that data-driven organizations require the right strategy, the right organization, the right technology, and the right data management.

3.2 Data Maturity

The previous subsections have outlined the qualitative indicators and the enablers of data-driven organizations. In this subsection, we are looking at frameworks that can quantify the data-drivenness of an organization.

3.2.1 Data Maturity Frameworks

To quantify the maturity of an organization in terms of data-drivenness several maturity frameworks evolved. Table 3 presents five selected data maturity frameworks. As pointed out by their authors, the common goals of all five frameworks are to provide guidance and a framework for the organizations that perform these assessments. Additionally, Halper et al. [17] highlight that their framework can be used as a benchmark to compare the organization's big data deployment to their peers. Enterprises can directly apply the assessment to an industrial context as it is a free-of-charge self-service. It

Table 3: Summary of selected data maturity frameworks

Author	Institute	Name	Goal	Origin	Readiness
Halper et al. 2013 [17]	TDWI Research	TDWI Big Data Maturity Model Guide	guidance, benchmark, framework	business consulting	survey, directly applicable
Comuzzi et al. 2016 [8]	-	Big Data Maturity Model (BDMM)	guidance, framework	applied research	research paper, not directly applicable
Termer et al. 2018 [40]	Bitkom	Digital Analytics & Optimization Maturity Index (DAOMI)	status quo assessment, guidance, framework	digital association	online survey, directly applicable
CMMI Institute 2019 [7]	CMMI Institute	Data Management Maturity Model (DMM)	guidance, framework	business consulting	survey, guided assessment
Gentsch 2019 [13]	-	framework and maturity model	guidance, framework	business consulting	guideline, not directly applicable

requires a full online registration to start the survey, and finally, Transforming Data with Intelligence (TDWI) Research will automatically provide the result and possible actions. The data maturity framework of Comuzzi et al. [8] originates in academic research. The "Big Data Maturity Model" results from a scientific comparison of five different data maturity assessments that the authors combined into one. The paper explains each stage of maturity in each dimension, but it is not directly applicable to an industrial context. It needs to be transferred from a framework to a survey before organizations can use it for self-assessment. The result is not provided, and actions need to be independently derived. The assessments of Halper et al. [17], Capability Maturity Model Integration (CMMI) Institute [7], and Gentsch [13] originate in business consulting. The CMMI Institute [7] offers the maturity assessment in the form of a guided assessment which requires an appointment to be scheduled. After this assessment, an action plan for the organization is, provided. It is also possible to purchase the more than 200 pages framework from their website for 100 USD, which allows for an organization to assess itself. Gentsch's [13] maturity assessment can, similar to Comuzzi et al. [8], not be directly applied to the industry. Gentsch [13] provides in his book the stages of each maturity model, including an explanation for which the assessment can be derived. The assessment of Termer et al. [40] originates from a digital association. It is a comprehensive self-service assessment that companies can apply directly. It is accessible via a website, but it is only available in the German language. In regards to ease of use, the available assessment offers the least hurdles. Yet, the DAOMI (Digital Analytics & Optimization Maturity Index) focuses on data analytics and optimization in regards to customer communicating. It assesses practices in the area of Sales and Marketing.

The selected frameworks disclosed limitations in the operational readiness for large enterprises, which are further discussed at the end of this chapter.

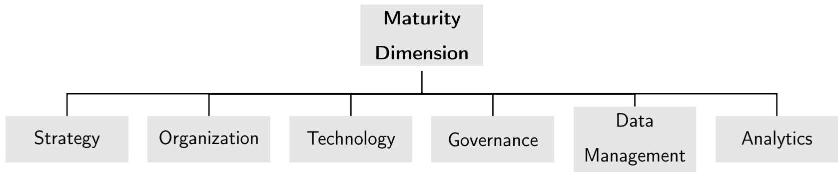

Figure 3: Dimensions of data maturity frameworks

3.2.2 Data Maturity Levels

Each of the data maturity models highlighted above evaluates different dimensions of a data-driven organization. For instance, Comuzzi et al. [8] focus on strategic alignment, governance, organization, information technology, and data. Each of the dimensions described in the data maturity frameworks of the different authors is synthesized. Figure 3 shows the identified dimensions. The dimensions partly match with the four main enablers (organization, technology, strategy, and data management) for data-driven organizations as describes in Section 3.1.2.

Each data maturity framework specifies different levels of maturity. The five maturity dimensions vary in the number of stages in between which they distinguish. There is no industry standard in regards to the number of stages nor the naming of each stage. Starting with the smallest number of stages is Gentsch [13]. Gentsch differs between non algorithmic enterprise, semi algorithmic enterprise and automated enterprise. The CMMI institute [7] assesses in regards to five stages: performed, managed, defined, measured, and optimized. Halper et al. [17] use six stages: nascent, preadoption, early adoption, corporate adoption, and visionary. Comuzzi et al. [8] also use six stages: nonexistent, initial, repeatable, defined, managed, and optimized. Figure 4 depicts how we can map these five assessments with different stages. In contrast to the other four assessments, CMMI Institute [7] has no stage

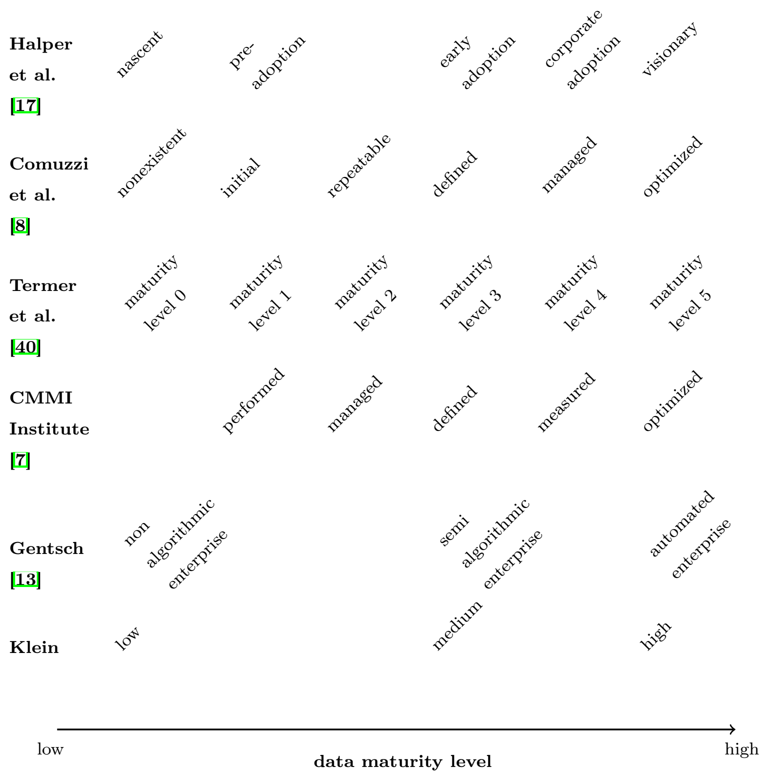

Figure 4: Mapping of data maturity levels

that describes a nonexistent data maturity. Comuzzi et al. [8] and CMMI institute [7] both use the term *"managed"* but mean different stages of maturity. Gentsch [13] uses, compared to the other sources, fewer stages. Halper et al. [17] stretch the stages from *preadoption* to *early adoption*, whereas Comuzzi et al. [8] and Termer et al. [40] split this level into three stages.

Table 4: Comparison for the dimension strategy

Author	Level		
	low	**medium**	**high**
Comuzzi et al. 2016	big data not considered in corporate strategy	big data strategy included in corporate strategy	big data is economical requirement and strategic imperative
Termer et al. 2018	no big data strategy available	big data strategy implicitly included in corporate strategy	big data strategy anchored in overall strategy
Gentsch 2019	no big data strategy available	rudimentary big data strategy	decided big data strategy

3.2.3 Data Maturity Dimensions

With the six dimensions we have previously clustered, this section investigates how each author describes or assesses the maturity level. We investigate the lowest level, a medium level, and the highest level of maturity in each dimension for a comprehensive understanding. The level can be identified in the mapping in Figure 4 under *Klein*.

3.2.3.1 Strategy

The dimension strategy evaluates the extent to which data is considered to define an organization's strategy.

Comuzzi et al. [8] assess two different subdimensions when investigating strategic alignment: *strategy* and *process*. In the subdimension *strategy*, the lowest level of maturity, named nonexistent, refers to an organization where the implications of big data are not considered. The medium level of maturity, named defined, applies to companies where the corporate strategy includes big data vision and strategy. Big data is being used to measure strategy fulfillment. The highest level of maturity, named optimized, is valid for enterprises where big data is an economic requirement and a strategic imperative. Comuzzi et al. [8] rate a company with no progress in using big data tools the lowest in the subdimension process. A medium level is achieved by organizations that use big data in most operational and decision-making processes. Further, big data-related key performance indicators and service level agreement with the IT functions are defined to homogenize big data across functions and departments. The highest level of maturity, optimized, refers to organizations where big data is incorporated in the enterprise-wide continuous improvement process.

Termer et al. [40] examine the dimension strategy and split it into three subdimensions: *strategy*, *performance management*, and *data value*. In the subdimension *strategy* the lowest level of maturity applied to organizations where no digital analytics and optimization strategy (big data strategy) is available. A medium level of maturity refers to organizations where the aspects of big data are implicitly included in corporate strategy, but cannot be recognized explicitly as such. This is only possible by breaking down the overall strategy on sub-areas or lower company levels. The highest level of maturity is achieved by organizations where the big data strategy is anchored in the overarching strategy. The second sub-dimension assesses *performance management*. An organization has a low maturity level when performance management does not exist. Medium level is achieved when a regular measurement in the form of a process exists. The highest level applies to organizations that have regular and proactive measurements incorporated in their improvement cycle. In the third sub-dimension, *data value* is assessed. Organizations, where there is no existing value add from data, have a low maturity level. When an organization has a systematic optimization of selected sub-areas and individual data-driven processes are introduced, a medium maturity level. The highest level is achieved by enterprises that have a huge amount of data-driven innovation.

Gentsch [13] names the strategic relevant dimension *strategic decision*. Gentsch's framework distinguishes three levels: non-algorithmic enterprise, semi-automated enterprise, and automated enterprise. The lowest level, non-algorithmic enterprise, applies to companies where data is not crucial to success and analytics is part of the IT department. The medium level, semi-algorithmic enterprise, is achieved by organizations that regard data as business relevant and where analytics is part of the IT department and specialist department. The highest level, automated enterprise, applies to an organization where data is the value driver and brings competitive advantages. The analytics department is part of the specialist' department supported by IT.

An overview of how the authors assess the dimension strategy is depicted in Table 4. The biggest difference can be seen for the medium level, where the authors differ between a big data strategy included in the corporate strategy [8], a big data strategy which is implicitly included [40] and a rudimentary strategy [13]. The assessment criteria for a medium level vary significantly in regards to the expected maturity.

Table 5: Comparison for the dimension governance

Author	Level		
	low	**medium**	**high**
Halper et al. 2013	no data governance	data governance structure defined	-
Comuzzi et al. 2016	no data governance	data governance structure defined	data governance anchored in enterprise level governance
CMMI Institute 2019	data governance functions for at least one project	data governance structure defined	data governance structure communicated to peer industry

3.2.3.2 Governance

The dimension governance evaluates to which extend data is governed in the organization.

Halper et al. [17] name, in their assessment, the lowest level in the dimension data governance, *nascent*. A low level applies to organizations that have no data governance. The medium level, *corporate adoption*, is achieved by organizations where there is a data governance with a well defined data strategy, data management is in place and where a steering committee oversees the progress of data. Halper et al. [17] do not explicitly state in their framework how the highest level of maturity is achieved in terms of data governance.

Comuzzi et al. [8] assess in their framework the dimension data governance, similar to Halper et al. [17]. The lowest level in the framework of Comuzzi et al., named *nonexistent*, applies to enterprises where no data governance is in place. The medium level is achieved by a company that has a complete specification of a big data analytics governance structure, that is further integrated into the enterprise governance structure. The highest level, *optimized*, is achieved by an organization where the data governance is anchored in the enterprise-level governance.

CMMI Institute [7] investigates the dimension governance management. The lowest level in the framework is named *performed*. In contrast to the other frameworks there is no level 0 or level nonexistent. The level *performed* applies to a company that has implemented governance functions for at least one project. The medium level, *defined*, applies to enterprises in which an organization-wide data governance structure and a roll-out plan with executive sponsorship are established. The highest level of maturity, *optimized*, applies to a company that communicates its data governance structure to the peer industry.

An overview of how the authors assess the dimension of governance is depicted in Table 5. The medium level is assessed consistently by the authors whereas major differences occur for the high level.

3.2.3.3 Organization

The maturity dimension organization evaluates the extent to which people and the organization's culture can be data-driven.

Halper et al. [17] assess in their framework the dimension organization. A low level applies to organizations with low awareness of big data and data value. A medium level is seen when the organization gets excited about the prospects of big data and more people start to come

Table 6: Comparison for the dimension organization

Author	**Level**		
	low	**medium**	**high**
Halper et al. 2013	low awareness for big data or data value	increased interest in big data	analytics seen as competitive weapon
Comuzzi et al. 2016	no awareness for big data; big data not part of org. values	awareness for big data utilization; positive and proactive attitude towards big data across organization	empowerment to experiment with big data; data-driven decision making defines organizational culture and leadership
Termer et al. 2018	no collaboration; lone some fighters; data silos	partial collaboration; static team structure; data access upon request	open and regular collaboration; dynamic team structure; free access to all data
Gentsch 2019	No CDO; no data scientist;limited analy talents	typically no CDO; analytically oriented staff	CDO; data-driven mindset

on board. In addition, a team has been formed to begin planning and strategizing for a larger big data scope. The highest level of maturity is achieved when analytics is seen as a competitive weapon and when enterprises look continuously for opportunities to leverage data.

Comuzzi et al. [8] - similar to Halper et al. [17] - assess the dimension organization in their assessment. Comuzzi et al. [8] split the dimension into two sub-dimensions, *people* and *culture*. In the sub-dimension *people*, the lowest level of maturity applies to organizations where the staff lacks awareness for big data. According to Comuzzi et al. [8], a medium level is achieved when the staff understands how big data can improve operational and decision-making processes and are fully engaged in using related tools. The highest level of maturity is achieved when the staff feels empowered to experiment with big data tools. In the sub-dimension *culture*, the lowest level of maturity applies to enterprises where big data is not part of the organizational values. A medium level of maturity regarding culture refers to organizations where the attitude towards big data is positive and proactive across the organization. The highest level is achieved when data-supported operations and evidence-based decisions are at the core of organizational culture and leadership.

Termer et al. [40] focus in their framework on organization, people, and staff. Within these dimensions several sub dimensions are investigated. Three of them are outlined here: *collaboration, team structure and data democracy*. In the sub dimension *collaboration* the lowest level of maturity applies to organizations with total isolation of data. A medium level is achieved by organizations where data and the knowledge gained from it are regularly and openly exchanged. Collaboration works partially. There is no knowledge or consideration of common goals. The highest level of maturity applies to organizations with open and regular sharing of data and derived insights. Collaboration works across departments and common goals are followed. A low maturity level in the sub-dimension *team structure* means that an organization has lonesome fighters and that different people do the same task repetitively. A medium-level refers to enterprises where the specialists departments have contacts who act as power users and multipliers. Knowledge is exchanged, but resources are tied to one department. The highest maturity is achieved when dynamic team structures exist and knowledge and resources are available at the right time at the right place. In the sub dimension *data democracy*, the low maturity applies to organizations with no access to data but many data silos. The medium level refers to a company where employees can request data access, where there are few silos and a good overview of the existing data landscape is available. The highest level is achieved when there is free access to all relevant data across departments.

Gentsch [13] follows a similar approach as Termer et al. [40]. Gentsch [13] also investigates in his framework people and organization, considering 3 levels. In a non-algorithmic Enterprise, at the lowest level, there is no Chief Data Officer (CDO), no data scientist, and limited analytics talents. In an enterprise with medium maturity, there is typically no CDO, but it has analytically oriented staff. Finally, an enterprise with a high level of maturity usually has a CDO and a data-driven mindset.

An overview of how the authors assess the dimension organization is depicted in Table 6. This dimension shows huge variances in the assessment criteria. The authors have different focuses concerning an organization. Termer et al. [40] for instance assess collaboration and team structure while Gentsch [13] assesses among others the existing roles. There is no consistency throughout the frameworks when assessing the maturity of the organization.

3.2.3.4 Technology

The maturity dimension of technology evaluates to which extend technology is used to handle and utilize data.

Halper et al. [17] focus in their assessment on the dimension *infrastruc-*

Table 7: Comparison for the dimension technology

Author	**Level**		
	low	**medium**	**high**
Halper et al. 2013	no awareness for differentiation between data and infrastructure	no unified architecture or ecosystem	coherent deployed data analytics infrastructure
Comuzzi et al. 2016	fragmented business applications; no awareness for big data integration	big data analytics tool deployed	full spectrum of big data technology exploited
Termer et al. 2018	no data management tools available; future usage not planned	data management platform available and partially integrated	data management platform available; ability to unite internal and external data as single point of truth
CMMI Institute 2019	target data architecture aligns business requirements with the implemented data store for at least one project	target architecture collaboratively developed and jointly approved	prediction models evaluated against architectural changes and adjusted as needed

ture. An organization with low maturity often misunderstands the need to differentiate between infrastructure and data which ultimately leads to the short-lived success of the organization's big data journey. An organization with medium maturity has different kinds of big data technologies available. The concept of a single architecture or ecosystem is still in its infancy, and these technologies have yet to be operationalized. The highest level of maturity, *visionary*, is achieved when the company has deployed a coherent analytics infrastructure that is fully operational and can be applied in critical aspects of the business.

Comuzzi et al. [8] assess the dimension information technology which is split into the sub dimensions *information management* and *IT infrastructure*. In the sub-dimension *information management* the lowest level of maturity, nonexistent, refers to organizations where information is not formally organized and there is no relationship between information structure and big data tools. In a medium mature organization, the IT function and other business functions decide which data should be acquired and stored. Data are collected in an enterprise with high maturity, and information structures and enterprise architecture are periodically reviewed to assess limitations. In the sub-dimension *IT infrastructure*, the low level applies when business applications are fragmented and there is no awareness of how the organization can integrate big data. The medium level refers to companies, where big data analytics tools are deployed at the production level, installed, and maintained at the enterprise level either on-premise or in the cloud. In an *optimized organization*, the full spectrum of big data technology is exploited as part of the enterprise IT infrastructure.

Termer et al. [40] focus in their framework on the dimension of technology and processes. Outlined here are the sub-dimension *data management tools*. A low level of maturity applies when no data management tools are available and future usage is not planned. A medium level is achieved when a data management platform is available and partially integrated and first data are centrally aggregated. The highest level applies when a data management platform is available and can unite internal and external data as one single point of truth.

CMMI institute [7] assess to which extend platforms and architecture are used to handle and utilize data. Depicted here are the sub-dimensions *architectural approach, architectural standards*, and *data management platform*. The lowest level, *performed*, in the sub-dimension *architectural approach* refers to companies where a goal data architecture connects business needs with the established data store for at least one project. The goal architecture is cooperatively created and jointly authorized by business units, IT, and data governance in a medium-maturity company. Prediction models are tested against architectural changes and changed as needed in a *optimal organization*. The stage *performed* applies to the sub-dimension *architectural standards* when data architecture standards are established and followed for at least one project. When metrics for monitoring and regulating architectural standards, as well as compliance with them, are developed and implemented, medium applies. A mature organization investigates new data technologies and processes for possible adoption and establishes acceptable new standards for implementation. The sub-dimension *data management* applies to companies who have defined data management platforms and components for at least one project. Critical data components for which the platform is an authoritative source, trustworthy source, or record system are documented in an organization with a medium maturity level. Based on statistical performance data and causal analysis, a mature organization constantly enhances the platform.

An overview of how the authors assess the dimension of technology is depicted in Table 7. This dimension shows huge variances in the assessment criteria. The authors have different focuses concerning technology. Especially the highest level shows that there is no unified understanding of what an organization with a high maturity needs to accomplish with its used technology.

3.2.3.5 Data Management

The dimension data management evaluates the extent to which data is managed in the organization.

Halper et al. [17] focus on the dimension of data management. According to the concept, an organization with a low level of data maturity has gathered data as files in various formats with no name conventions and poorly specified storage structures. The medium level applies to businesses that gather data in a variety of formats. Most likely these organizations have division or enterprise standards for naming and storage management. A highly mature company can integrate new data sources for analytics, whether internal or external to the company.

Comuzzi et al. [8] assess in their framework the dimension data. The dimension is split into two sub-dimensions *analytics* and *management*. In the sub-dimension *analytics*, the lowest level *nonexistent* applies to organizations that lack awareness of what kind of big data analytics software can be relevant for their goals. In an organization with medium mature analytics, the range of analytics software available in the company is known. The highest level is achieved when all enterprise users can tap into the analytics seamlessly with support from the IT department. In the sub-dimension *management*, the lowest level

Table 8: Comparison for the dimension data management

Author	Level		
	low	**medium**	**high**
Halper et al. 2013	data files with different formats; no naming standards; storage structures minimally defined	data files with different formats; potential naming standard and storage management	ability to integrate new sources (internal or external) of data for analytics
Comuzzi et al. 2016	data management and related policies siloed and not formally defined	data sources and data types identified and tracked; all data centrally stored and available across organizations	data sources, data types and data policies periodically reviewed to assess usefulness and actual usage
Termer et al. 2018	no mechanisms in place to assure data quality	regular, careful checks of data quality for important quality criteria using special, current quality procedures	periodically testing of data quality for all relevant quality criteria with dynamic fitting of the exiting criteria to handle changing conditions
CMMI Institute 2019	data quality objective rules, and criteria documented	data quality strategy accompanied by corresponding policies processes, and guideli	data quality program milestones and metrics defined and reviewed; continuous improvements implemented
Gentsch 2019	data used in operative systems; different data sources not linked	data used strategically; data sources partially linked	data used strategically; relevant data sources fully linked

refers to companies where data management and related policies are siloed and not formally defined. On a medium level, *defined* refers to organizations that identify and track data sources and data kinds. And where all data are centrally kept and available across the company. Data sources, kinds, and rules are evaluated on a regular basis in companies with the highest data management maturity level to assess their usefulness and actual utilization.

Termer et al. [40] also focus on the dimension data in their framework. Especially outlined here is the sub-dimension *data quality*. Low maturity refers to an organization where no mechanisms are in place to assure data quality. A medium-level applies when regular, careful checks of the data quality for important quality criteria using special and current quality management procedures are conducted. A high level is achieved when the data quality is periodically tested for all relevant quality criteria with the dynamic fitting of the exiting criteria to handle changing conditions.

CMMI Institute [7] distinguishes in their framework two dimensions, *data quality*, and *data operations*. The level performed in the dimension data quality refers to an organization where data quality objectives, rules, and criteria are documented. The medium level *defined* applies when the data quality strategy is followed across the organization and is accompanied by corresponding policies, processes, and guidelines. The high level, *optimized*, is achieved when executives regularly review data quality program milestones and metrics, and continuous improvements are implemented. In the dimension of *data operations*, one sub-dimension is *data lifecycle management*. In this sub-dimension, the *per-*

formed stage is achieved when the data lifecycle for a business process is defined and applied. Stakeholders develop and approve data lifecycle management procedures, which are then governed by data governance bodies and processes at the medium level. When data lifecycle metrics are improved and evaluated by senior management on a regular basis, the greatest degree of *optimized* is reached.

Gentsch [13] also focuses on the dimension data. In a *non-algorithmic enterprise* data are used in operative systems while different data sources are not linked to each other. In a *semi-automated enterprise* data is used strategically. Data sources are partially linked to each other. An *automated enterprise* data is used strategically, e.g. for sales predictions, and relevant data sources are fully linked to each other.

An overview of how the authors assess the dimension of data management is presented in Table 8. This dimension shows huge variances in the assessment criteria. It is the only dimension that is assessed in all frameworks. The characteristics of how the maturity of data management is assessed are very different. Termer et al. [40] and Comuzzi et al. [8] look among others at data quality, while Gentsch assesses the linkage between data sources.

3.2.3.6 Analytics

The maturity dimension analytics evaluates the extent to which data analytics are incorporated into the organization.

Halper et al. [17] the lowest level of maturity refers to an organization where analytics occurs in pockets and silos. The medium stage *early adoption* applies when utilizing descriptive or even predictive analytics in its projects. A *mature* organization makes use of all kinds of data,

Table 9: Comparison for the dimension analytics

Author	**Level**		
	low	**medium**	**high**
Halper et al. 2013	analytics occurring in silos	descriptive or predictive analytics in projects	company uses all kinds of data
Gentsch 2019	simple, isolated and ad-hoc analytics	advanced analytics, data mining machine learning	analytics results automatically used for creating and optimizing business processes

including real-time data, and uses this as part of its decision-making and incorporates it into business processes.

Gentsch [13] assess the dimension analytics. In a *non-algorithmic enterprise* simple, isolated and ad-hoc analytics are applied. In a *semi-automated enterprise* advanced analytics, data mining, and machine learning are used to generate insights. In an *automated enterprise* analytics results are automatically used for creating and optimizing business processes.

An overview of how the authors assess the dimension analytics are shown in Table 9. This dimension shows little variances in the assessment criteria.

3.3 Discussion

Large manufacturing enterprises such as OeTech are looking for an opportunity to leverage their data and become data-driven organizations. In the first step, the status quo needs to be assessed [22]. Based on the current status quo, we can develop a roadmap [34]. For businesses like OeTech, the assessment must be operationally ready, which implies creating a roadmap from the assessment results. The selected frameworks have significant limitations, especially Gentsch [13] and Comuzzi et al. [8]. They are not operational ready because they just present criteria but do not guide how actions can be derived from the result. Furthermore, it is important for the operational readiness that organizations can perform the assessments without interviewing several people in the organization as this creates additional effort within a larger organization due to complex coordination. The framework of the CMMI Institute is highly complex and covers 200 pages. With the limited time available in day-to-day business in large enterprises, this requires significant efforts [7]. Furthermore the assessment uses a language which is rather difficult to understand for non native English speaker. The assessment of Termer et al. [40] is only available in German and focus on sales and marketing which limits the applicability for international companies such as OeTech. Both frameworks of Halper et al. [17] and Termer et al. [40] require the user to reveal the companies name, location, and industry. For large enterprises such as OeTech revealing their data maturity without a non-disclosure agreement could harm the business if these data leak [38].

The different levels of maturity as they are depicted in Figure 4 are not standardized throughout the frameworks. Additionally, the different maturity levels within the frameworks are difficult to differentiate when more than four different levels are considered. This applies for instance to the assessment of Comuzzi et al. [8] which has six different levels.

The discussions show that all selected frameworks limitations regarding their operational readiness. Therefore, we propose developing an operational ready data maturity assessment for large enterprises because existing frameworks are not fit for purpose.

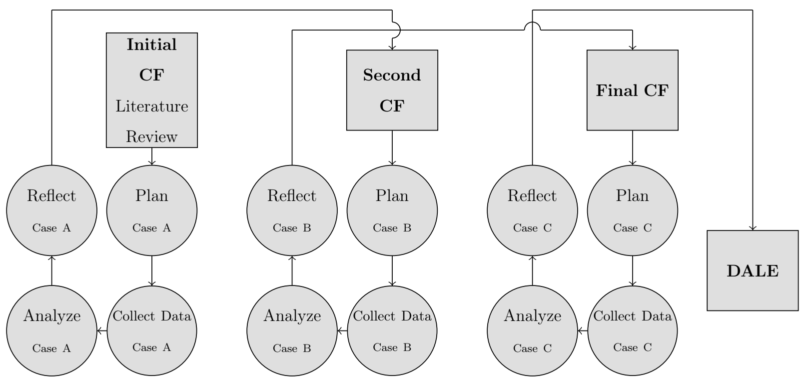

Figure 5: The spiral towards DALE (adapted from [6])

4 Structured Case: Data Maturity Framework

Due to the limitations described in the previous chapter and the lack of applicable frameworks, the idea arose to develop a framework that closes the gap between the existing frameworks and an operational ready data maturity framework applicable to large enterprises. For a company such as OeTech, it is primarily important to understand the status quo regarding the organization's data-drivenness. Therefore the framework needs to be operationally ready, easy to use, be conducted as a self-service, and applicable to a broad range of industries such as the manufacturing industry.

4.1 Applied Methodology

To develop a data maturity framework the structured case methodological framework by Carroll and Swan [6] was applied. In Section 2.3.2 the theory behind the structured cased methodological framework is explained. The goal was to develop an operationally ready, easy-to-use framework. The figure 5 gives an overview of how we practically applied the structured case. From each research cycle, focusing on one case, a new conceptual framework evolves. In the following paragraphs, the steps *Plan* and *Collect Data* are outlined. The steps *Analyze* and *Reflect* are described after the introduction of each framework.

Plan. Using the insights generated from the literature review in Chapter 3 the first step was to design an initial conceptual framework. The initial framework was the starting point for three different research cycles that we planned as three cases.

Case A. In the first research cycle, the focus was on comprehensibility. The lack of comprehensibility as described in the limitations of available frameworks was targeted as a gap to be closed. Therefore, we interviewed interview partners with a general understanding of data and data management in Case A. We decided not to interview data experts to test if - in practice - framework users with less expertise can understand and comprehend the questions.

Case B. The starting point for the second research cycle along case B is based on the second conceptual framework. In this research cycle, the focus was on the applicability of the framework in a wide range. The lack of applicability to various industries, especially the manufacturing industry, was targeted as a gap to be closed. Therefore interview partners with different professions, backgrounds, and from different dimensions were assigned.

Case C. The final research cycle along case C is based on the final conceptual framework. In this research cycle, the focus was on operational readiness. The lack of operational readiness was targeted

Table 10: Cases A, B and C representing the research cycles

Case	Focus	Interviewees	Interview set up
A	Comprehensibility of framework	2 Data Analyst and 1 Senior Manager	remote, structured interview
B	Applicability of framework	1 Data Architect, 1 Business Developer and 2 Senior Manager	remote, structured interview
C	Operational readiness of framework	2 Business Consultants (Data & Analytics) and 1 Data Governance Manager	remote, structured interview

as a gap to be closed. For this case, interview partners working in data management and who are therefore possible framework users were selected.

Collect Data. For each case, at least three informants were interviewed. The informants were assigned according to their expertise level in data management and their industry sector. We transferred each framework into a google form which we shared during a 30min video call with the interviewees. The google form was the main structure for the structured interview. The interview partner was able to read the question by themselves. If there was no concern or question regarding the subdimension we moved on to the next question. We reacted to answers and questions flexibly. For instance, in some interviews, we asked additional questions to deepen the understanding. In the google form for each subdimension, the different increments of maturity were listed as multiple-choice questions. There was one "new" option for each question, which we used to collect the feedback and improvement ideas from the interview partner. Table 10 summarizes the data collection of the three cases.

Analyze The analysis is conducted after all informants for the specific case were interviewed. The first step of the analysis section was to transfer the answers in the google form to a spreadsheet. In the spreadsheet, the findings of the interviewees are compared subdimension by subdimension and then consolidated. Based on the feedback possible adoptions for the framework were analyzed.

Reflect In the final phase, we reflect on the findings of the case concerning the conceptual frameworks. In tables, we present the consolidated feedback and propose possible adaptations to the framework. After doing this independently for every case, we implement the proposed changes in the conceptual framework. In this way, the changes to the framework are tested from one case to the next case.

4.2 Building the Conceptual Frameworks

The starting point of the structured case is the initial conceptual framework, which is based on the literature review of Chapter 3. Six dimensions that we identified in Chapter 3.2.2 built the framework dimensions. Within the subdimension, the authors have assessed different criteria. For the initial conceptual framework, these criteria are clustered, combined, and merged into subdimensions. For each of the 18 subdimensions three levels -low, medium, high- were identified. The decision to use three levels is based on the one hand on the literature review in which we compared three levels of each author. On the other hand, we created the hypothesis that for an easy-to-use framework, three levels are sufficient. Figure 6 shows the six

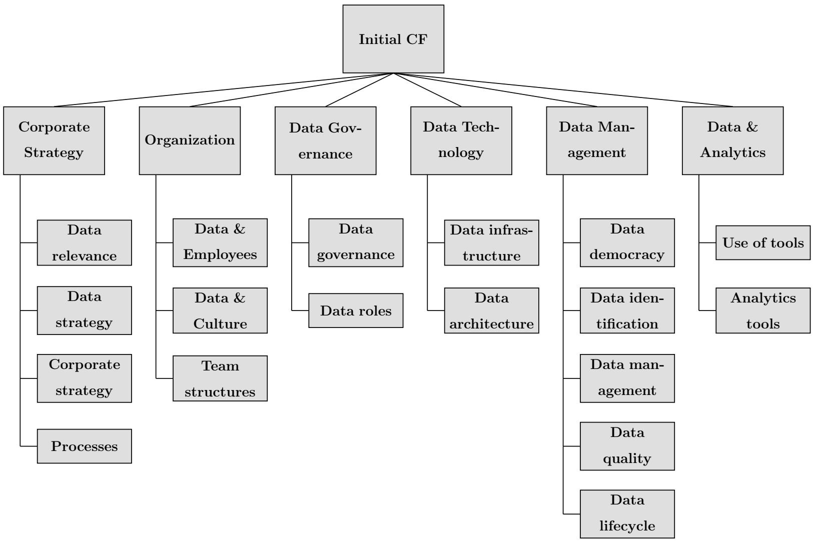

Figure 6: Assessment dimensions within the initial conceptual framework

dimensions and the 18 subdimensions of the data maturity assessment.

4.2.1 Business Glossary

The base for a common understanding of the framework is the business glossary. Terms used in the framework which we clarified throughout the research cycle are explained in the business glossary in alphabetical order.

Data Architecture Data architecture provides the foundation required to more effectively use and share data and optimize the flow of data between a company's internal systems and those of customers, suppliers, and business partners [44].

Data Governance Data governance provides mechanisms and procedures to ensure significant participation across the organization in crucial data asset decisions. Data governance structures are groups of data management stakeholders whose mission is to ensure that the data management program is held responsible to organizational needs and is focused on meeting business and data management objectives. [7].

Data Infrastructure Architecture explains the design of the components and their relationships, whereas infrastructure describes the actual set of components that make up a system. In a nutshell, a system is created on top of an infrastructure with a specific architecture [15].

Data Platform A data platform is a consolidated system that combines scalable flexibility, distributed data storage, and processing capacity to acquire and analyze huge data sets and offer users with accurate and trustworthy data. [21].

Single Source of Truth In the area of data management, it used to be common for enterprises to share their data by replication. The replication of data across different systems causes inconsistency and interoperability challenges. To avoid these challenges, a system should be used as the authoritative source or also referred to as the Single Source of Truth (SSOT) [29].

4.2.2 Structure of the Framework

In the framework, six different dimensions with 18 subdimensions are assessed. For each subdimension three increments - low, medium, high - are stated. The sequence of the dimensions follows a logical structure. We are using the analogy of a house to explain the framework easily. The house is depicted in Figure 7. The first two dimensions of the framework built the foundation of data maturity. These dimensions are Corporate Strategy and Organizational Structure. The fourth dimension of the framework built

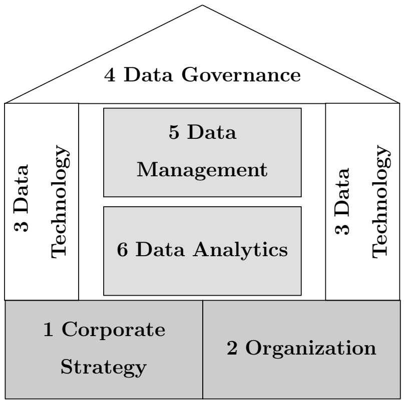

Figure 7: Structure of the data maturity framework

the walls and the house's roof with the dimensions Data Governance and Data Technology. The last two dimensions, Data Management and Data & Analytics reflect the life within the house.

4.2.3 Initial Conceptual Framework

Table 11 shows the entire conceptual framework, as we used it for the first research cycle in Case A. Each dimension, subdimension, and the increments are indexed and represented in column *Index*. For instance, dimension *1 Corporate Strategy* has the subdimension *1-2 Data Strategy* with three increments: 1-2-1, 1-2-2, and 1-2-3. The increments are sorted in ascending order. 1-2-1 is the lowest level and 1-2-3 is the highest level in regards to data maturity.

Table 11: Initial conceptual framework

Begin of Table 11		
Dimension	**Index**	**Criteria**
		1 Corporate Strategy
1-1 Data relevance	1-1-1	Data is not seen as business relevant.
	1-1-2	Data is seen as business relevant.
	1-1-3	Data is seen as a value driver with competitive advantage.
1-2 Data strategy	1-2-1	A data strategy does not exist.
	1-2-2	A rudimentary data strategy exists.
	1-2-3	A defined data strategy exists.
1-3 Corporate strategy	1-3-1	Data is not considered in the corporate strategy.
	1-3-2	Data is used to check if the corporate strategy is fulfilled.
	1-3-3	Data is an economical requirement and strategic imperative.
1-4 Processes	1-4-1	Processes for the utilization of data do not exist.
	1-4-2	Individual data driven processes are introduced.
	1-4-3	There is a huge amount of data-driven innovation in the organization.
		2 Organization
2-1 Data & Employees	2-1-1	Employees have no awareness for data.
	2-1-2	Employees understand how data driven processes can improve the process and utilize related tools. More and more people are involved and start planning to use data
	2-1-3	The organization continuously looks for opportunities to leverage data. The staff is empowered to use data.
2-2 Data & Culture	2-2-1	Relevance of data is not part of organisational values.
	2-2-2	The attitude towards the usage of data is positive throughout the organization.
	2-2-3	Data-driven processes and decisions are at the core of the organizational culture and leadership.

Continuation of Table 11		
Dimension	**No.**	**Criteria**
2-4 Team structure	2-3-1	The teams are isolated and do not exchange data. The same tasks are done by different teams (redundant tasks).
	2-3-2	There is limited exchange of data between the teams. The goal of cross functional use of data is not clear. Specialist departments are power users and act as multipliers. The resources are tied to one department.
	2-3-3	Data and the knowledge exchanged from data is is shared openly. The team structures are dynamic. Knowledge and resources are at the right time at the right place.
3 Data Governance		
3-1 Data governance	3-1-1	No data governance implemented.
	3-1-2	Data governance implemented with executive sponsorship.
	3-1-3	Data governance is anchored in enterprise level governance.
3-2 Data roles	3-2-1	Specific data roles do not exists. (E.g. no Data Scientist, no Data Engineers, no Data Owner, ...)
	3-2-2	A few data roles exist. Generally the staff is analytically oriented
	3-2-3	Data roles are part of many teams. The organization has a CDO.
4 Data Technology		
4-1 Data infrastructure	4-1-1	The organization does not differentiate between infrastructure and data.
	4-1-2	A data platform is in place. The notion of unified architecture and ecosystem is not wide spread. Technologies are not operationalized.
	4-1-3	A data and analytics infrastructure is deployed and operationalized.

Continuation of Table 11		
Dimension	**No.**	**Criteria**
4-2 Data architecture	4-2-1	No data architecture defined.
	4-2-2	The target data architecture is defined. Metrics to control and monitor the compliance to the architectural standard are implemented.
	4-2-3	The organization looks out for now data technologies and potential adoption.
		5 Data Management
5-1 Data democracy	5-1-1	employees have no to very limited access to data
	5-1-2	employees have difficulties to access available data (technical reasons or skill reasons)
	5-1-3	access to available data is very easy (e.g via a data mart)
5-2 Data identification	5-2-1	Data are fragmented throughout the organization. Data is not formally organized. The collection of data is time consuming.
	5-2-2	The organizations has a clear understanding for availability of data in some areas. IT and Business decide together which data should be acquired and stored centrally.
	5-2-3	The centrally collected data and utilization of data is periodically reviewed.
5-3 Data management	5-3-1	No data management platform is available. Data of different sources are not linked to each other.
	5-3-2	Data management platform is available and partially integrated. First data sources are centrally aggregated.
	5-3-3	Data management platform is available and combined internal and external data as single point of truth.
5-4 Data quality	5-4-1	no data quality objectives rules and criteria are documented
	5-4-2	data quality objectives, rules and criteria are followed throughout the organization, policy, processes and guidelines are introduced

Continuation of Table 11		
Dimension	**No.**	**Criteria**
	5-4-3	data quality is regularly reviewed and continuous improvements are implemented
5-5 Data lifecycle	5-5-1	no data lifecycle defined
	5-5-2	data lifecycle management processes are defined and approved
	5-5-3	data lifecycle processes are implemented and periodically defined and reviewed
		6 Analytics
6-1 Use of Analytics	6-1-1	no data analytics
	6-1-2	data analytics introduced in some processes
	6-1-3	analytics are used to drive process decisions
6-2 Analytics Tools	6-2-1	no tools
	6-2-2	tools available and used by some
	6-2-3	use of tools periodically reviewed, employees continuously trained on the use of analytics tools
		End of Table 11

Case A. Analyze. Case A investigates the *comprehensibility of the framework*. Therefore we conducted a structured interview with three different interview partners. We introduced the interview partners to the methodology. Their task was to read the given statement for each subdimension and respond if they comprehend the statement. Since Case A was the first research cycle within the structured case the responses by the interviewees did not only focus on comprehensibility but also on logical and application issues within the framework. All feedback was considered valuable and was taken into account in the reflection. Overall the interviewees provided the feedback that the initial framework is comprehensible but needs to be refined and rephrased for certain subdimensions.

Case A. Reflect. With we the feedback provided by the interviewees, adoptions to the framework were initiated. Table 16 in the Appendix A summarizes both the feedback given by the interviews and the adoptions to the framework. In total 10 out of 18 subdimensions were adopted. We made several refinements to close the gap between increments.

4.2.4 Second Conceptual Framework

The second conceptual framework evolved from the research cycle based on Case A and the adaptions after the reflection. The entire framework and the reflection is shown in Appendix B.

Case B. Analyze. Case B investigates the *applicability of the framework*. Therefore we conducted a structured interview with four different interview partners. we introduced the interview partners to the methodology and the scope of Case B. Their task was to read the given statements for each subdimension and respond if they could apply the framework to their current or previous organization. The feedback and responses of the interviewees did not only focus on the applicability but also logical and comprehensibility issues within the framework. All feedback was considered valuable and was taken into account in the reflection. Overall the interviewees provided the feedback that the framework fits the purpose but needs to be refined and clarified within certain subdimensions.

Case B. Reflect. With the feedback from interviewees, we made adoptions to the framework. Table 18 in the Appendix summarizes both the feedback given by the interviews and the adoptions to the framework. In total six out of 18 subdimensions were adopted. Several refinements were made using specific examples.

4.2.5 Final Conceptual Framework

The third conceptual framework evolved from the research cycle based on Case B and the adaptions after the reflection. The entire framework and the reflection is shown in Appendix C.

Case C. Analyze. Case C investigates the *operational readiness of the framework*. Therefore we conducted a structured interview with three different interview partners. We introduced the interview partners to the methodology and the scope of Case C. Their task was to read the given statements for each subdimension and respond if they could apply the framework to their current or previous organization. The feedback and responses of the interviewees did not only focus on the operational readiness but also logical and comprehensibility issues within the framework. All feedback was considered valuable and was taken into account in the reflection. Overall the interviewees provided the feedback that the framework fits the purpose and is operationally ready.

Case C. Reflect. With the feedback from interviewees, we initiated adoptions to the framework. Table 20 summarizes both the feedback given by the interviews and the adoptions to the framework. In total, we adopted one out of 18 subdimensions. Refinements were made for subdimension 1-4 Processes.

4.3 Summarizing and Presenting DALE

Throughout the three research cycles the conceptual frameworks were adopted based on the feedback of the interview partners.

4.3.1 Analysis and Reflection over all Cases

After the research cycle based on Case C, we made only one adoption to the framework. Therefore we conclude that we have reached the natural end of the structured case as we gained no further insights. We have reached the goal of creating an operational ready data maturity framework. Throughout the interviews conducted we orally clarified several technical terms to set the base for a common understanding. Additionally, the data maturity framework aims to be used in a self-assessment. Therefore we concluded that the data maturity assessment can not stand alone, it requires an introduction and a business glossary.The operational ready data maturity framework is derived from a structured case. The initial framework was based on the data maturity frameworks by Halper et al. [17], Comuzzi et al. [8], Termer et al. [40], CMMI Insititute [7], and Gentsch [13]. The outcome of the structured case in an operationally ready, easy-to-use data maturity framework.

The operational ready data maturity framework targets large enterprises where data maturity is not easy to assess and a roadmap to increase the data maturity is difficult to derive. For instance, OeTech is a manufacturing company that would like to understand its level of data maturity to set the right initiatives to become a data-driven organization.

Table 12: DALE Framework

Begin of Table 12		
Dimension	**Index**	**Criteria**
		1 Corporate Strategy
1-1 Data relevance	1-1-1	Data is not seen as business relevant.
	1-1-2	Data is seen as business relevant.
	1-1-3	Data is seen as a value driver with competitive advantage. This applies for instance when data products are sold.
1-2 Data strategy	1-2-1	A data strategy does not exist.
	1-2-2	A rudimentary data strategy exists.
	1-2-3	A defined data strategy exists.
1-3 Corporate strategy	1-3-1	Data is not considered in the corporate strategy.
	1-3-2	Data is used to check if the corporate strategy is fulfilled.
	1-3-3	Data is an economical requirement and/or a strategic imperative.
1-4 Processes	1-4-1	Data and analytics are not used to steer processes.
	1-4-2	Data and analytics are rarely used to steer processes in some areas.
	1-4-3	Data and analytics is used to steer and improve processes throughout the organization.
		2 Organization
2-1 Data & Employees	2-1-1	Employees lack understanding how to make use of data and analytics.
	2-1-2	Employees in some departments understand how data driven processes can improve the process. More and more employees are involved and start planning to use data.
	2-1-3	The organization continuously looks for opportunities to leverage data. The staff is empowered to use data.

Continuation of Table 12		
Dimension	**No.**	**Criteria**
2-2 Data & Culture	2-2-1	Relevance of data is not part of organisational values.
	2-2-2	The attitude towards the usage of data is positive throughout the organization.
	2-2-3	Data-driven processes and decisions are at the core of the organizational culture and leadership.
2-4 Team structure	2-3-1	The teams are isolated and do not exchange data. The same tasks are redundantly done by different teams.
	2-3-2	There is limited exchange of data between the teams. The goal of cross functional use of data is not clear. Specialist departments are power users and act as multipliers. The resources are tied to one department.
	2-3-3	Data and the knowledge exchanged from data is is shared openly if possible (legal, compliance). The team structures are dynamic. Knowledge and resources are at the right time at the right place.
3 Data Governance		
3-1 Data Governance	3-1-1	No data governance is implemented. No initiatives are planned.
	3-1-2	The organization considers to initiate or has initiated the establishment of a data governance.
	3-1-3	Data governance is anchored in enterprise level governance.
3-2 Data roles	3-2-1	Specific data roles do not exists. (e.g. no Data Scientist, no Data Engineers, no Data Owner, ...)
	3-2-2	A few data roles exist internally. The organization compensates lack of data roles with external resources such as consultants.
	3-2-3	Internal and external data roles are part of many teams. The organization has a CDO.

Continuation of Table 12		
Dimension	**No.**	**Criteria**
		4 Data Technology
4-1 Data infrastructure	4-1-1	The organization is not aware of a data infrastructure. For instance, data are manually and regularly queried from an ERP system to be analyzed in spreadsheets.
	4-1-2	A data platform (a data warehouse for instance) is in place. The notion of unified architecture and ecosystem is not wide spread. Technologies are not operationalized.
	4-1-3	A data and analytics infrastructure is deployed and operationalized.
4-2 Data architecture	4-2-1	No data architecture defined.
	4-2-2	The target data architecture is defined and metrics to control and monitor the compliance to the architectural standard are being established.
	4-2-3	The organization looks out for now data technologies and potential adoption.
		5 Data Management
5-1 Data democracy	5-1-1	Employees have no to very limited access to relevant data
	5-1-2	Employees have difficulties to access available and relevant data (technical reasons or skill reasons)
	5-1-3	Access to available and relevant data is very easy (e.g via a data mart)
5-2 Data identification	5-2-1	Data are fragmented throughout the organization. Data is not formally organized. The collection of data is time consuming.
	5-2-2	The organizations has a clear understanding for availability of data. IT and Business decide together which data should be acquired and stored centrally.
	5-2-3	The centrally collected data and utilization of data is periodically reviewed.

Continuation of Table 12		
Dimension	**No.**	**Criteria**
5-3 Data management	5-3-1	No data management platform available. Data of different sources are not linked to each other.
	5-3-2	Data management platform is available and partially integrated. First data sources are centrally aggregated.
	5-3-3	Data management platform available and combined internal and external data as single point of truth.
5-4 Data quality	5-4-1	No data quality objectives rules and criteria are documented.
	5-4-2	Data quality objectives, rules and criteria are followed in some areas of the organization. Policy, processes and guidelines are introduced.
	5-4-3	Data quality is regularly reviewed and continuous improvements are implemented.
5-5 Data lifecycle	5-5-1	No data lifecycle defined.
	5-5-2	Data lifecycle management processes are defined and approved.
	5-5-3	Data lifecycle processes are implemented and periodically defined and reviewed.
6 Analytics		
6-1 Use of Analytics	6-1-1	Data analytics is used in a few areas of the organization.
	6-1-2	Data analytics introduced in several processes.
	6-1-3	Analytics are used to drive process decisions. The employees are able to interpret analytical results.
6-2 Analytics Tools	6-2-1	No tools.
	6-2-2	Tools available and used by some. The utilization of tools is not reviewed.
	6-2-3	The use of tools is periodically reviewed. The employees are continuously trained on the use of analytics tools.
End of Table 12		

4.3.2 DALE

Table 12 shows the complete operational ready **DA**ta Maturity Framework for **L**arge **E**nterprises (DALE).

4.4 Literature-based Scrutinizing of DALE

This section compares and contrasts the final framework with a range of existing literature. We are looking for instance at agreements and conflicts in the selected dimensions of the framework and their complexity.

4.4.1 Dimensions of DALE

In the first step, we are comparing DALE with the dimension of frameworks in the literature. DALE, as shown in Table 12, consists of six dimensions. An **X** indicates that DALE has a match in the dimension of the framework with another framework. The **(X)** reflects that the dimension in DALE partly matches with a dimension in the literature. DALE and Halper et al. [17] and Comuzzi et al. [8] match in four dimensions and partly in one additional dimension. Halper et al. [17] does not consider the corporate strategy in their framework. Comuzzi et al. do not consider analytics in their framework. Termer et al. [40] have the least matches with DALE. The frameworks match in three dimensions and partly in one dimension. Termer et al. assess culture, staff, and organization in separated dimensions, whereas DALE assesses this consolidated in the subdimension of organization. Termer et al. do not consider Data Governance and Data Analytics in their assessment. We can conclude that the consolidated literature agrees with the dimensions used in DALE. This is natural as DALE derived from the literature and during the research cycle no additional dimensions were added or removed.

Table 13: Matching dimensions in DALE and the literature

Author	Corporate Strategy	Organization	Data Governance	Data Technology	Data Management	Data Analytics
Klein 2021	X	X	X	X	X	X
Halper et al. 2013 [17]		X	X	(X)	X	X
Comuzzi et al. 2016 [8]	X	X	X	X	(X)	
Termer et al. 2018 [40]	X	X		X	(X)	
CMMI Institute 2019 [7]	X		X	X	X	
Gentsch 2019 [13]	X	X			X	X

4.4.2 Complexity of DALE

In the second step, we are contrasting the depth of DALE with the literature. Table 14 depicts a quantitative compilation of dimensions in the framework. The first column shows the quantity of dimension which are assessed. DALE, Termer et al. [40] and CMMI Institute [7] cover six dimensions in the frameworks. Comuzzi et al. [8] cover with four dimensions the least dimensions among the examined frameworks.

The second column shows the quantity of subdimension. The quantity of subdimensions represents the level of detail of a framework. As presented in Figure 6, DALE covers 18 subdimension in the framework. With 38 subdimensions Termer et al. [40] consider the most subdimensions. In his framework, Gentsch [13] does not explicitly have any subdimensions. In comparison to the literature DALE has an average level of detail that neither conflicts nor agrees with any of the examined frameworks. Additionally, we examine the average quantity of subdimensions per dimension which does not deliver any further insights. The fourth column compares the number of levels in which the frameworks specify different levels of maturity. The six examined maturity dimensions vary in the number of stages in between which they distinguish. As highlighted in Section 3.2.2 there is no industry standard regarding the number of stages nor the naming of each stage. In contrast to the frameworks of Halper et al. [17], Comuzzi et al. [8], Termer et al. [40] and CMMI Institute [7] DALE has significantly fewer increments. Nevertheless, we see an agreement with Gentsch [13] who also assesses the maturity in his framework considering three levels. We subsequently conclude that DALE has a reduced level of detail when assessing the level of maturity compared to other frameworks. The fifth column in Table 14 is the product of subdimensions and levels. This column represents the complexity of the framework. Halper et al. [17] for instance consider 25 subdimensions with five levels each. In this particular example, it means that the user or reader of the framework has to go through 125 increments to assess the maturity of their enterprise. DALE has a complexity of 54 whereas Termer et al. [40] has a complexity of 190. Overall, it is seen that Halper et al. [17], Termer et al. [40] and CMMI Institute [7] form a group of complex frameworks with 125 to 190. Comuzzi et al. [8] and DALE form a group with medium complex frameworks. Gentsch's framework has the lowest level of complexity. We conclude that the framework of DALE agrees with the complexity of industry-wide used frameworks.

Table 14: Quantitative comparison of DALE with the Literature

Author	**Qty. of dimensions**	**Qty. of subdimensions**	**Av. qty. of subdimensions**	**Levels**	**Complexity**	**Language**
Klein 2021	6	18	3	3	54	English
Halper et al. 2013 [17]	5	25	5	5	125	English
Comuzzi et al. 2016 [8]	4	9	2,25	5	45	English
Termer et al. 2018 [40]	6	38	6,3	5	190	German
CMMI Institute 2019 [7]	6	25	4,17	6	150	English
Gentsch 2019 [13]	5	0	0	3	15	English

4.4.3 DALE in Practical Use

Considering the practical use of the given frameworks and their operational readiness, we observed that the levels applied in DALE are sufficient to distinguish maturity. The higher the level of complexity the more difficult it is for the user but at the same time, it allows for a more concrete assessment. At this point, we see an opportunity for future work to compare the practical usability of frameworks and the advantages and disadvantages of high or medium complexity.

Table 15: DALE applied to OeTech

Begin of Table 15		
Dimension	**Index**	**Result OeTech**
		1 Corporate Strategy
1-1 Data relevance	1-1-2	Data is seen as business relevant.
1-2 Data strategy	1-2-2	A rudimentary data strategy exists.
1-3 Corporate strategy	1-3-2	Data is used to check if the corporate strategy is fulfilled.
1-4 Processes	1-4-2	Data and analytics are used to steer processes in some areas.
		2 Organization
2-1 Data & Employees	2-1-2	Employees in some departments understand how data driven processes can improve the process. More and more employees are involved and start planning to use data.
2-2 Data & Culture	2-2-2	The attitude towards the usage of data is positive throughout the organization.
2-3 Team structure	2-3-1	The teams are isolated and do not exchange data. The same tasks are redundantly done by different teams.
		3 Data Governance
3-1 Data Governance	3-1-2	The organization considers to initiate or has initiated the establishment of a data governance.
3-2 Data roles	3-2-2	A few data roles exist internally. The organization compensates lack of data roles with external resources such as consultants.
		4 Data Technology

Continuation of Table 15		
Dimension	**No.**	**Result OeTech**
4-1 Data infrastructure	4-1-1	The Organization is not aware of a data infrastructure. For instance, data are manually and regularly queried from an ERP system to be analyzed in spreadsheets.
4-2 Data architecture	4-2-1	No data architecture defined.
		5 Data Management
5-1 Data democracy	5-1-2	Employees have difficulties to access available and relevant data (technical reasons or skill reasons)
5-2 Data identification	5-2-1	Data are fragmented throughout the organization. Data is not formally organized. The collection of data is time consuming.
5-3 Data management	5-3-1	No data management platform available. Data of different sources are not linked to each other.
5-4 Data quality	5-4-1	No data quality objectives rules and criteria are documented.
5-5 Data lifecycle	5-5-1	No data lifecycle defined.
		6 Analytics
6-1 Use of Analytics	6-1-1	Data analytics is used in a few areas of the organization.
6-2 Analytics Tools	6-2-2	Tools available and used by some. The utilization of tools is not reviewed.
End of Table 11		

5 Application and Roadmap

Enterprises such as OeTech need guidelines that guide the company to become a data-driven organization. Therefore, in the first step the status quo is assessed using DALE. From the assessment we are then deriving a short, mid and long-term roadmap.

5.1 Application of DALE Using the Example of OeTech

To assess the Data maturity of OeTech using DALE, we have interviewed two IT Senior Managers. As we consider DALE as an operationally ready, easy-to-use framework we provided the managers with an introduction, the business glossary, and the framework in a google form. The managers then conducted a self-assessment of OeTech. Table 15 exemplary shows the result of the assessment.

Corporate Strategy. At OeTech data is seen as business relevant. The organization has a rudimentary data strategy. Currently, data are neither a strategic nor economic requirement but data is used to check if the corporate strategy is fulfilled. Furthermore, Data and analytics are used to steer processes. This only applies to some areas and not throughout the organization. In all subdimension OeTech scores a medium maturity, which concludes that in the dimension of corporate strategy OeTech has a medium maturity regarding data-drivenness.

Organization. At OeTech employees in some areas already understand how data-driven processed can improve the process. An increasing amount of employees are involved and start planning to use data. On the one hand, the attitude towards data is positive throughout OeTech. On the other hand, however, the teams are isolated and do not exchange data. This causes that different teams redundantly do the same tasks. In regards to the organization, OeTech achieves a medium level in two dimensions. In one dimension OeTech has a low maturity level. We conclude that OeTech has a medium maturity in the Organization towards being a data-driven company.

Data Governance. The current status at OeTech is that the enterprise has already started to initiate data governance. At OeTech a few data roles exist internally and the organization compensates for the lack of data roles with external resources such as consultants. In all subdimensions, OeTech has a medium maturity. Therefore, we conclude that OeTech has overall a medium maturity in their Data Governance.

Data Technology. At OeTech the employees are not aware of data infrastructure. For instance, data from different source systems are manually and repetitively queried to analyze or visualize them in spreadsheets. Furthermore, OeTech has no defined architecture. In all subdimensions regarding Data Technology OeTech scores low. Hence we conclude that OeTech has a low maturity in regards to the applied Data Technology.

Data Management. Employees at OeTech have difficulties accessing available and relevant data, either caused by technical limitations or a lack of skills. OeTech's data are fragmented throughout the enterprise. Data is not formally organized, which caused that the collection of data is time-consuming. Furthermore, there is no data management platform available. Data from different sources are not linked to each other. In four out of five subdimension OeTech scores a low data maturity. We conclude that in consideration of the dimension Data Management

OeTech has a low maturity.

Data Analytics. At OeTech data analytics is used in a few areas of the organization. There are tools available to analyze data. The utilization of the tools is not reviewed. Overall, OeTech scores a low to medium maturity in the area of Analytics.

All in all, the result of the data maturity assessment using DALE is that OeTech has a low to medium data maturity. Especially, in the area of Technology and Data Management, the maturity is low. OeTech does not achieve a high maturity in any of the assessed dimensions. As described in Section 4.2.2 DALE follows the architecture of a house. The foundations - Corporate strategy and Organization - already have a medium maturity whereas the following structures are less mature. For OeTech this means that the management and the organization are ready to be built upon.

5.2 Deriving a Short, Mid, and Long-term Roadmap for OeTech

After we have assessed the status quo of OeTech, we are now deriving three roadmaps that will guide OeTech towards becoming a data-driven organization. The short-term roadmap contains the initiatives for the next two years. The mid-term roadmap contains the actions planned from year two until year four. Finally, we present a long term roadmap that introduces activities for year seven and beyond.

OeTech's 4-year roadmap has the ultimate target that OeTech achieves a medium maturity in all assessed dimensions.

Figure 8 visualizes both, the current situation as well as the mid-term target situation in a spiderweb chart. The dark grey area represents the

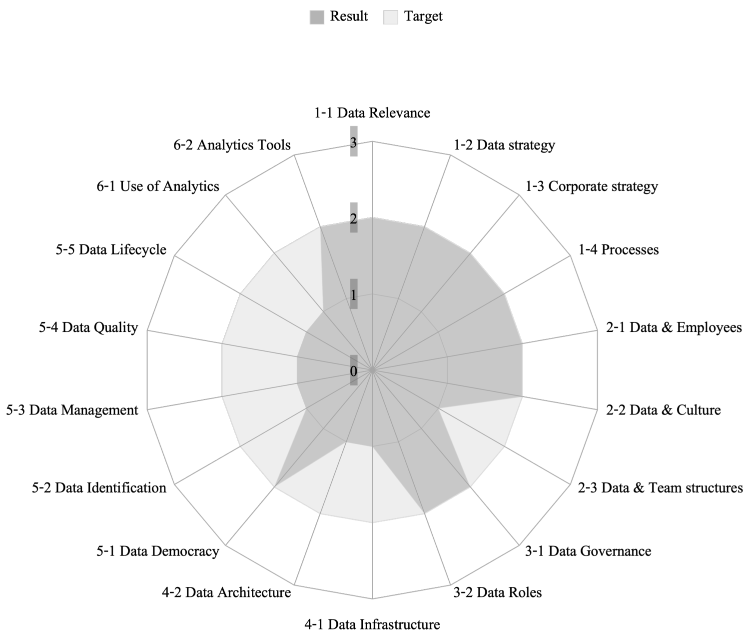

Figure 8: Assessment result and target level maturity

current situation while the lighter gray shade shows the target scenario. The numbers from 0 to 3 imply the following meaning:

(0) *empty* the organization has not been assessed

(1) *low* data maturity

(2) *medium* data maturity

(3) *high* data maturity

The target scenario forms a circle as the intention is to bring all subdimensions to a medium maturity level. The gaps which we can identify relate to eight subdimensions. Four subdimensions are targeted in the 2-year roadmap. The remaining subdimensions are outlined in the 4-year roadmap.

5.2.1 2-year Roadmap for OeTech

To create a short-term roadmap for OeTech to increase the maturity towards a data-driven organization is necessary to create a target scenario that we can realistically achieve. For OeTech we prioritized four subdimensions which need to be developed in the upcoming two years.

Data Infrastructure. For OeTech to reach a medium maturity regarding the data infrastructure the company needs to build up a data platform that contains important data of the organization. The data platform is an enabler for further subdimensions to be improved, such as data management [11].

Data Architecture. OeTech needs to define a data architecture. Furthermore, metrics to control and monitor compliance to the architectural standard need to be established[18] .

Data Identification. To improve the time-consuming data collection OeTech needs to assess which data are available in which areas of the organization. This helps the organization to get a clear understanding of the availability of data [25]. Then IT and business can decide which data should be acquired and stored centrally in the data platform.

Data Management. The data management is closely interlinked with the data infrastructure and the data identification. Once OeTech needs to link data from different sources and store them centrally to increase the data maturity [19].

The 2-year roadmap provides eight areas of improvement for OeTech to move towards a data-driven organization. We are highlighting that the provided roadmap initiatives are interlinked and contribute to each other. The order in which the initiatives are taken is not relevant as they can all be started in parallel.

5.2.2 4-year Roadmap for OeTech

The aim of the 4-roadmap is to bring all dimensions to a medium maturity as depicted in 8.

Team structure. OeTech needs to reduce the isolation of teams. OeTech can do this by forming for instance cross-functional teams. For OeTech the goal needs to be to reduce the amount of redundantly performed work. The more the teams start to exchange and increase communication, the easier it will be to exchange data [24].

Data Quality. OeTech needs to introduce data quality objectives with corresponding rules and criteria. Policies and guidelines must be introduced to increase the maturity concerning data quality [5].

Data Lifecycle. OeTech furthermore needs to implement a data lifecycle management. For certain data, the lifecycle is defined together with the customer. Other data should be clustered for their relevance and critically in the future to set up a data lifecycle management [26].

Use of Analytics. OeTech needs to improve use the use of analytics. On the one hand, OeTech can do this by educating the employees but on the other hand, it requires the management to claim performed analytics by their employees to make decisions [23].

5.2.3 7-year Roadmap for OeTech

We are not defining an explicit 7-year roadmap for OeTech initiatives. In seven years, the status quo needs to be reassessed. Bokrantz et al. [4] predict that in the coming years, digitalization will be impacted by advancement of data analytics, increased emphasis on education and training, and stronger environmental legislation and standards.

In consideration of DALE, OeTech's aim is to increase the data maturity in the fundamental dimensions of corporate strategy and organization.

6 Conclusion

This sections summarizes the findings of the integrated literature review, the result of the structured case methodology and finally the application of DALE.

6.1 Summary

Data-driven companies value, manage and protect their data [27]. They use data to continuously test new ideas, new products, and to solve business problems [9], [30]. Data-driven organization focus on building a culture which accepts and learns from failures [1], [3]. Data-driven companies use their data for the widest range of possible applications [20].

A data-driven enterprise is enabled when data is integrated in the decision making process [3], when the right people are hired, and data literacy exists throughout the organization [37]. Furthermore, the right technology and a defined data architecture are enablers for data-driven organizations [32].

To answer the research question - *Which strategic requirements, organizational structures, and technologies are required to create a roadmap towards becoming a data-driven organization?* - we have investigated five existing data maturity frameworks. These frameworks differ in the assessed dimensions, the number of maturity levels, and operational readiness. We concluded that the available assessments do not fit the purpose and synthesized an operational ready data maturity framework for large enterprises, named DALE. We used DALE to assess a fictive large manufacturing company. DALE assess six different domains of a data driven organization: corporate strategy, organization, data governance, data technology, data management and analytics. From the result a short, mid, and long-term roadmap was defined for OeTech. The short term initiatives target among others the set up of a fitting data architecture and infrastructure to enable for instance the data democracy. The defined mid term initiatives focus leveraging the potential of data by upskilling of employees and bringing analytical tools into use.

The research work provides an easy to use framework, which can be applied to a vast majority of organizations. The framework is backed up with a synthesized view on what are indicators and enablers for a data-driven organizations. Research benefits from this work because the operational readiness of the existing frameworks is assessed and their practical limitations are are outlined. Researchers need to consider the practical application of their work. Organizations benefit from this work by getting an understanding for the data-drivenness. Compared to existing frameworks DALE has a reduced complexity, but looks at six different dimensions which makes it a broader framework compared to the selected frameworks. Future work should consider the necessary complexity for organization and how many resources an organizations needs to invest to conduct a data maturity assessment. Additionally, organizations need share good practices on their initiatives towards becoming a data-driven organization. For instance the "upskilling" of employees is an initiative that is on OeTech's mid-term roadmap. But the question remains how this can be executed. Furthermore, the literature leaves a lot of space for data architecture in large enterprises. This on the one hand needs to be considered by researchers but also by organization which could share best practices in data architecture to peers in the industry. Finally, this thesis has shown that data maturity frameworks exists. They significantly differ in their complexity.

6.2 Future Work

Future studies should aim to help companies like OeTech to become more data driven. Therefore, future work should be devoted to the establishment of organizational standards or tools and technologies for data management. In addition, the application and development of frameworks, such as DALE, should be continuously enhanced by findings from industry and research.

References

[1] Espen Andersen, John Chandler Johnson, Vegard Kolbjørnsrud, and Ragnvald Sannes. The data-driven organization: Intelligence at scale. In *At the Forefront, Looking Ahead*, pages 23–42. Universitetsforlaget, 2018.

[2] Carl Anderson. *Creating a Data-Driven Organization - Practical Advice from the Trenches*. "O'Reilly Media, Inc.", Sebastopol, 2015.

[3] Mikael Berndtsson, Daniel Forsberg, Daniel Stein, and Thomas Svahn. Becoming a data-driven organisation. *Research-in-Progress Papers. 43.*, 2018.

[4] Jon Bokrantz, Anders Skoogh, Cecilia Berlin, and Johan Stahre. Maintenance in digitalised manufacturing: Delphi-based scenarios for 2030. *International Journal of Production Economics*, 191:154–169, 2017.

[5] Li Cai and Yangyong Zhu. The challenges of data quality and data quality assessment in the big data era. *Data science journal*, 14, 2015.

[6] Jenny Carroll and Paul Swatman. Structured-Case: A Methodological Framework for Building Theory in Information Systems Research. In *European Journal of Information Systems volume 9*, pages 116–123, 2000.

[7] CMMI Institute. Data Management Maturity (DMM) Model. *Data Management Maturity (DMM) Model*, 2019.

[8] Marco Comuzzi and Anit Patel. How organisations leverage big data: A maturity model. *Industrial management & Data systems*, 2016.

[9] Thomas H Davenport et al. Competing on analytics. *Harvard business review*, 84(1):98, 2006.

[10] Harold Davis. *Google advertising tools: Cashing in with AdSense, AdWords, and the Google APIs.* " O'Reilly Media, Inc.", 2006.

[11] Mike Ferguson. Architecting a big data platform for analytics. *A Whitepaper prepared for IBM*, 30, 2012.

[12] Amir Gandomi and Murtaza Haider. Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2):137 – 144, 2015.

[13] Peter Gentsch. *AI in marketing, sales and service: How marketers without a data science degree can use AI, big data and bots.* Springer, 2018.

[14] Mert Gökalp, Kerem Kayabay, Mehmet Ali Akyol, P Eren, and Altan KoçyiÄŸit. Big Data for Industry 4.0: A Conceptual Framework. In *2016 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 431–434, 2016.

[15] Jonathan Gray, Carolin Gerlitz, and Liliana Bounegru. Data infrastructure literacy. *Big Data & Society*, 5(2):2053951718786316, 2018.

[16] S Groggert, M Wenking, RH Schmitt, and T Friedli. Status quo and future potential of manufacturing data analyticsâ€"an empirical study. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 779–783. IEEE, 2017.

[17] Fern Halper and David Stodder. Tdwi analytics maturity model guide. *TDWI research*, pages 1–20, 2014.

[18] William H Inmon, Daniel Linstedt, and Mary Levins. *Data Architecture: A Primer for the Data Scientist: A Primer for the Data Scientist*. Academic Press, 2019.

[19] Rod Koch. Draining the data swamp: An organization's data lake is only as good as the preparation and maintenance planning that go into creating it. *Strategic Finance*, 99(12):62–64, 2018.

[20] Steve LaValle, Eric Lesser, Rebecca Shockley, Michael S Hopkins, and Nina Kruschwitz. Big data, analytics and the path from insights to value. *MIT sloan management review*, 52(2):21–32, 2011.

[21] Martin Lidl, Tom Robinson, and James Moorhouse. The next generation data platform. *Deloitte publication: Inside Magazine*, 2018.

[22] Frank Lindner. *Projektwissensmanagement: Status quo, Gestaltungsfaktoren und Erfolgsdeterminanten des Wissensmanagements in der Projektabwicklung*, volume 8. LIT Verlag Münster, 2010.

[23] Lionel Lionel et al. Developing an undergraduate data analytics program for non-traditional students. *ISEDJ*, 18(3):18, 2020.

[24] Márcio Lopes Pimenta, Andrea Lago da Silva, and Wendy L Tate. Developing and managing cross-functional teams: A multi-case study of brazilian manufacturing companies. *Journal of technology management & innovation*, 9(2):1–16, 2014.

[25] H Gilbert Miller and Peter Mork. From data to decisions: a value chain for big data. *It Professional*, 15(1):57–59, 2013.

[26] Martin Hubert Ofner, Kevin Straub, Boris Otto, and Hubert Oesterle. Management of the master data lifecycle: a framework for analysis. *Journal of Enterprise Information Management*, 2013.

[27] Kelle O'Neal. Becoming a data-driven organization: A five-step process, 2017.

[28] Boris Otto. Quality and Value of the Data Resource in Large Enterprises. *Information Systems Management*, 32(3):234–251, 2015.

[29] Candy Pang and Duane Szafron. Single Source of Truth (SSOT) for Service Oriented Architecture (SOA). In Xavier Franch, Aditya K Ghose, Grace A Lewis, and Sami Bhiri, editors, *Service-Oriented Computing*, pages 575–589, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg.

[30] DJ Patil and Hilary Mason. *Data Driven.* " O'Reilly Media, Inc.", 2015.

[31] Travis Pearson and Rasmus Wegener. Big data: the organizational challenge. *Bain Co*, 2013.

[32] Tobias Pentek, Christine Legner, and Martin Fadler. Pmi's journey towards a data-driven enterprise. *PMI White Paper*, 2018.

[33] John Durham Peters. Communication as dissemination. *G. Shepherd, J. St. John, & T. Striphas (Eds.). Communication asâ€¦ Perspectives on Theory*, pages 211–222, 2006.

[34] Robert Phaal and Gerrit Muller. An architectural framework for roadmapping: Towards visual strategy. *Technological forecasting and social change*, 76(1):39–49, 2009.

[35] David Reinsel, John Gantz, and John Rydning. The digitization of the world from edge to core. *IDC White Paper*, 2018.

[36] Cynthia Russell. An overview of the integrative research review. *Progress in transplantation (Aliso Viejo, Calif.)*, 15:8–13, 2005.

[37] Feliks Sejahtera, Wei Wang, Marta Indulska, and Shazia Sadiq. Enablers and inhibitors of effective use of big data: Insights from a case study. In *PACIS 2018 PROCEEDINGS*, 06 2018.

[38] Asaf Shabtai, Yuval Elovici, and Lior Rokach. *A survey of data leakage detection and prevention solutions*. Springer Science & Business Media, 2012.

[39] Statista. Google: ad revenue 2001-2019 | statista. https://www.statista.com/statistics/266249/advertising-revenue-of-google/. (Accessed on 08/09/2020).

[40] Frank Termer and Tobias Weiss. Reifegradmodell zum digital analytics & optimization maturity index (daomi). https://www.bitkom.org/sites/default/files/file/import/20181018-Reifegradmodell-zum-Digital-Analytics-Optmization-Maturity-In.pdf. (Accessed on 08/02/2020).

[41] Coleen E Toronto and Ruth Remington. *A step-by-step guide to conducting an integrative review*. Springer, 2020.

[42] Google Trends. data driven organization - explore - google trends. https://trends.google.com/trends/explore?date=all&q=data%20driven%20organization. (Accessed on 08/02/2020).

[43] Patrick Ulrich, Wolfgang Becker, Alexandra Fibitz, Eva Reitelshöfer, and Felix Schuhknecht. Data analytics systems and sme type–a design science approach. *Procedia Computer Science*, 126:1162–1170, 2018.

[44] John van den Hoven. Data architecture: Standards for the effective enterprise. *Information Systems Management*, 21(3):61, 2004.

Appendices

A Initial Conceptual Framework

Table 16: Reflection on Case A along initial conceptual framework.

Begin of Table 16		
Dimension	**Explanation**	**Adaption to Framework**
	1 Corporate Strategy	
1-1 Data relevance	difficult criteria to assess if data is the only "product" sold by an organization	1-1-3 add additional possibility if data is the only product
1-3 Corporate strategy	within the public sector data could be a strategic imperative without being an economical requirement	1-3-3 add additional information to clarify
	2 Organization	
2-1 Data & Employees	2-1-2 considers the utilization of tools which is part of 6-1	remove "and utilize related tools..."; replace "people" by "employees"
	3 Data Governance	
3-1 Data Governance	the increments are too far apart: beginning to implement data governance is not considered	new 3-1-2 considers the implementation, 3-1-3 is replaced by previous 3-1-2
	4 Data Technology	
4-2 Data architecture	the increments between 4-2-1 and 4-2-2 are too far apart	rephrase 4-2-1 to reduce the graduation
	5 Data Management	

Continuation of Table 16		
Dimension	**Explanation**	**Adaption to Framework**
5-1 Data Democracy	data democracy should be addressed earlier in the framework, the increments do not consider confidentiality of data	explain how the sequence should be thought off, add 'if allowed'
5-2 Data identification	the increments between 5-2-2 and 5-2-3 are too far apart	rephrase 5-2-2 using 'in some areas'
5-3 Data Management	the meaning of "single source of truth" is not comprehensive	before the framework a list of definitions should be provided
5-4 Data quality	the increments between 5-4-1 and 5-4-2 are too far apart	5-4-2: replace "throughout the organization" by "in some areas"
	6 Analytics	
6-2 Analytics Tools	theoretical availability vs. practical use is not considered	6-2-2: add "the utilization of tools is not reviewed"
End of Table 16		

B Second Conceptual Framework

Table 17: Second conceptual framework

Begin of Table 11		
Dimension	**Index**	**Criteria**
		1 Corporate Strategy
1-1 Data relevance	1-1-1	Data is not seen as business relevant.
	1-1-2	Data is seen as business relevant.
	1-1-3	Data is seen as a value driver with competitive advantage. This applies for instance when data products are sold.
1-2 Data strategy	1-2-1	A data strategy does not exist.
	1-2-2	A rudimentary data strategy exists.
	1-2-3	A defined data strategy exists.
1-3 Corporate strategy	1-3-1	Data is not considered in the corporate strategy.
	1-3-2	Data is used to check if the corporate strategy is fulfilled.
	1-3-3	Data is an economical requirement and/or a strategic imperative.
1-4 Processes	1-4-1	Processes for the utilization of data do not exist
	1-4-2	Individual data driven processes are introduced.
	1-4-3	There is a huge amount of data-driven innovation in the organization.
		2 Organization
2-1 Data & Employees	2-1-1	Employees have no awareness for data.
	2-1-2	Employees understand how data driven processes can improve the process. More and more employees are involved and start planning to use data.
	2-1-3	The organization continuously looks for opportunities to leverage data. The staff is empowered to use data.

Continuation of Table 11		
Dimension	**No.**	**Criteria**
2-2 Data & Culture	2-2-1	Relevance of data is not part of organisational values.
	2-2-2	The attitude towards the usage of data is positive throughout the organization.
	2-2-3	Data-driven processes and decisions are at the core of the organizational culture and leadership.
2-4 Team structure	2-3-1	The teams are isolated and do not exchange data. The same tasks are redundantly done by different teams.
	2-3-2	There is limited exchange of data between the teams. The goal of cross functional use of data is not clear. Specialist departments are power users and act as multipliers. The resources are tied to one department.
	2-3-3	Data and the knowledge exchanged from data is is shared openly. The team structures are dynamic. Knowledge and resources are at the right time at the right place.
3 Data Governance		
3-1 Data governance	3-1-1	No data governance is implemented. No initiatives are planned.
	3-1-2	The organization considers to initiate or has initiated the establishment of a data governance.
	3-1-3	Data governance is anchored in enterprise level governance.
3-2 Data roles	3-2-1	Specific data roles do not exists. (e.g. no Data Scientist, no Data Engineers, no Data Owner, ...)
	3-2-2	A few data roles exist. Generally the staff is analytically oriented
	3-2-3	Data roles are part of many teams. The organization has a CDO.
4 Data Technology		
4-1 Data infrastructure	4-1-1	The organization does not differentiate between infrastructure and data.
	4-1-2	A data platform is in place. The notion of unified architecture and ecosystem is not wide spread. Technologies are not operationalized.

Continuation of Table 11		
Dimension	**No.**	**Criteria**
	4-1-3	A data and analytics infrastructure is deployed and operationalized.
4-2 Data architecture	4-2-1	No data architecture defined.
	4-2-2	The target data architecture is defined and metrics to control and monitor the compliance to the architectural standard are being established.
	4-2-3	The organization looks out for now data technologies and potential adoption.
		5 Data Management
5-1 Data democracy	5-1-1	Employees have no to very limited access to relevant data
	5-1-2	Employees have difficulties to access available and relevant data (technical reasons or skill reasons)
	5-1-3	Access to available and relevant data is very easy (e.g via a data mart)
5-2 Data identification	5-2-1	Data are fragmented throughout the organization. Data is not formally organized. The collection of data is time consuming.
	5-2-2	The organizations has a clear understanding for availability of data. IT and Business decide together which data should be acquired and stored centrally.
	5-2-3	The centrally collected data and utilization of data is periodically reviewed.
5-3 Data management	5-3-1	No data management platform available. Data of different sources are not linked to each other.
	5-3-2	Data management platform is available and partially integrated. First data sources are centrally aggregated.
	5-3-3	Data management platform available and combined internal and external data as single point of truth.

Continuation of Table 11		
Dimension	**No.**	**Criteria**
5-4 Data quality	5-4-1	No data quality objectives rules and criteria are documented.
	5-4-2	Data quality objectives, rules and criteria are followed in some areas of the organization. Policy, processes and guidelines are introduced.
	5-4-3	Data quality is regularly reviewed and continuous improvements are implemented.
5-5 Data lifecycle	5-5-1	No data lifecycle defined.
	5-5-2	Data lifecycle management processes are defined and approved.
	5-5-3	Data lifecycle processes are implemented and periodically defined and reviewed.
6 Analytics		
6-1 Use of Analytics	6-1-1	No data analytics.
	6-1-2	Data analytics introduced in some processes.
	6-1-3	Analytics are used to drive process decisions.
6-2 Analytics Tools	6-2-1	No tools.
	6-2-2	Tools available and used by some. The utilization of tools is not reviewed.
	6-2-3	The use of tools is periodically reviewed. The employees continuously trained on the use of analytics tools.
End of Table 11		

Table 18: Reflection on Case B along the second conceptual framework.

Dimension	Explanation	Adaption to Framework
	2 Organization	
2-1 Data & Employees	difficult to answer because it does not apply to all employees e.g. blue collar, does not consider that not all employees fully understand how to leverage analytics	rephrase 2-1-1 "employees lack understanding..." , 2-1-2 add "in some departments"
2-3 Team structures	difficult to answer when data can not be shared due to legal or compliance relationships	2-3-3 add "if possible (legal, compliance)"
3-2 Data Roles	consider to add working with external companies to share roles (e.g. consulting)	3-2-2 add "The organization uses external resources to compensate lack of data roles"; rephrase 3-2-3 "internal and external data roles"
	4 Data Technology	
4-1 Data infrastructure	4-1-1 abstract terminology, consider to use and example	rephrase 4-1-1 "Organization is not aware of a data infrastructure e.g. data are manually and regularly queried from an ERP system to analyze them using Excel"
	5 Data Management	
5-4 Data quality	difficult to answer as the graduation between 5-4-2 and 5-4-3 is minor	rephrase 5-4-2 "processes to improve data quality are not yet implemented"
	6 Analytics	
6-1 Use of Analytics	The assessment does not consider if employees are able to interpret the results, very unlikely that no analytics is used in the organization	Rephrase 6-2-1, rephrase 6-2-2

C Final Conceptual Framework

Table 19: Final conceptual framework

Begin of Table 17		
Dimension	**Index**	**Criteria**
		1 Corporate Strategy
1-1 Data relevance	1-1-1	Data is not seen as business relevant.
	1-1-2	Data is seen as business relevant.
	1-1-3	Data is seen as a value driver with competitive advantage. This applies for instance when data products are sold.
1-2 Data strategy	1-2-1	A data strategy does not exist.
	1-2-2	A rudimentary data strategy exists.
	1-2-3	A defined data strategy exists.
1-3 Corporate strategy	1-3-1	Data is not considered in the corporate strategy.
	1-3-2	Data is used to check if the corporate strategy is fulfilled.
	1-3-3	Data is an economical requirement and/or a strategic imperative.
1-4 Processes	1-4-1	Processes for the utilization of data do not exist
	1-4-2	Individual data driven processes are introduced.
	1-4-3	There is a huge amount of data-driven innovation in the organization.
		2 Organization
2-1 Data & Employees	2-1-1	Employees lack understanding how to make use of data and analytics.
	2-1-2	Employees in some departments understand how data driven processes can improve the process. More and more employees are involved and start planning to use data.
	2-1-3	The organization continuously looks for opportunities to leverage data. The staff is empowered to use data.

Continuation of Table 17		
Dimension	**No.**	**Criteria**
2-2 Data & Culture	2-2-1	Relevance of data is not part of organisational values.
	2-2-2	The attitude towards the usage of data is positive throughout the organization.
	2-2-3	Data-driven processes and decisions are at the core of the organizational culture and leadership.
2-4 Team structure	2-3-1	The teams are isolated and do not exchange data. The same tasks are redundantly done by different teams.
	2-3-2	There is limited exchange of data between the teams. The goal of cross functional use of data is not clear. Specialist departments are power users and act as multipliers. The resources are tied to one department.
	2-3-3	Data and the knowledge exchanged from data is is shared openly if possible (legal, compliance). The team structures are dynamic. Knowledge and resources are at the right time at the right place.
3 Data Governance		
3-1 Data governance	3-1-1	No data governance is implemented. No initiatives are planned.
	3-1-2	The organization considers to initiate or has initiated the establishment of a data governance.
	3-1-3	Data governance is anchored in enterprise level governance.
3-2 Data roles	3-2-1	Specific data roles do not exists. (e.g. no Data Scientist, no Data Engineers, no Data Owner, ...)
	3-2-2	A few data roles exist internally. The organization compensates lack of data roles with external resources such as consultants.
	3-2-3	Internal and external data roles are part of many teams. The organization has a CDO.
4 Data Technology		
4-1 Data infrastructure	4-1-1	The Organization is not aware of a data infrastructure. For instance, data are manually and regularly queried from an ERP system to be analyzed in spreadsheets.

Continuation of Table 17		
Dimension	**No.**	**Criteria**
	4-1-2	A data platform (a data warehouse for instance) is in place. The notion of unified architecture and ecosystem is not wide spread. Technologies are not operationalized.
	4-1-3	A data and analytics infrastructure is deployed and operationalized.
4-2 Data architecture	4-2-1	No data architecture defined.
	4-2-2	The target data architecture is defined and metrics to control and monitor the compliance to the architectural standard are being established.
	4-2-3	The organization looks out for now data technologies and potential adoption.
		5 Data Management
5-1 Data democracy	5-1-1	Employees have no to very limited access to relevant data
	5-1-2	Employees have difficulties to access available and relevant data (technical reasons or skill reasons)
	5-1-3	Access to available and relevant data is very easy (e.g via a data mart)
5-2 Data identification	5-2-1	Data are fragmented throughout the organization. Data is not formally organized. The collection of data is time consuming.
	5-2-2	The organizations has a clear understanding for availability of data. IT and Business decide together which data should be acquired and stored centrally.
	5-2-3	The centrally collected data and utilization of data is periodically reviewed.

Continuation of Table 17		
Dimension	**No.**	**Criteria**
5-3 Data management	5-3-1	No data management platform available. Data of different sources are not linked to each other.
	5-3-2	Data management platform is available and partially integrated. First data sources are centrally aggregated.
	5-3-3	Data management platform available and combined internal and external data as single point of truth.
5-4 Data quality	5-4-1	No data quality objectives rules and criteria are documented.
	5-4-2	Data quality objectives, rules and criteria are followed in some areas of the organization. Policy, processes and guidelines are introduced.
	5-4-3	Data quality is regularly reviewed and continuous improvements are implemented.
5-5 Data lifecycle	5-5-1	No data lifecycle defined.
	5-5-2	Data lifecycle management processes are defined and approved.
	5-5-3	Data lifecycle processes are implemented and periodically defined and reviewed.
		6 Analytics
6-1 Use of Analytics	6-1-1	Data analytics is used in a few areas of the organization.
	6-1-2	Data analytics introduced in several processes.
	6-1-3	Analytics are used to drive process decisions. The employees are able to interpret analytical results.
6-2 Analytics Tools	6-2-1	No tools.
	6-2-2	Tools available and used by some. The utilization of tools is not reviewed.
	6-2-3	The use of tools is periodically reviewed. The employees are continuously trained on the use of analytics tools.
End of Table 17		

Table 20: Reflection on Case C along the final conceptual framework.

Dimension	Explanation	Adaption to Framework
	1 Corporate Strategy	
1-4 Processes	1-4-1 targets processes to utilize data whereas 1-4-2 and 1-4-3 target process steering based on data	rephrase 1-4-1 to clarify the scope