B2B data consultancy performed by SMEs for growth purposes – A systematic literature review

Author: Jan Hein Swillens
University of Twente
P.O. Box 217, 7500AE Enschede
The Netherlands

ABSTRACT

This paper guides SMEs in performing B2B data consultancy for growth purposes, based on a systematic literature review. It examines three key pillars: Strategy, Value and Fundament. Each pillar has essential sub-pillars, fourteen of those sub-pillars have been used in the research. The goal is to identify best practices, trends and challenges in data usage, highlighting areas that are well-researched and areas that need further exploration. A holistic approach of continuous learning is advised for effective data consultancy, in order to successfully help clients grow in becoming data-driven.

Graduation Committee members:
1st Supervisor: Dr. Vincent Göttel
2nd Supervisor: Dr. Martin Stienstra

Keywords

Systematic literature review, data consultancy, data expertise pillars, data strategy, data value, data fundament

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
B2B	Business-to-business
BI	Business Intelligence
MDM	Master Data Management
ML	Machine Learning
SLR	Systematic Literature Review
SME	Small and medium-sized enterprises

1. INTRODUCTION

Humanity is transitioning from an industrial to an information civilisation, and adaptation is required to avoid falling behind. When trying to solve new problems, traditional techniques will lead to new obstacles and confusion. The future society, including organisations that desire competitive advantages, has to integrate the virtual world of data with the real world composed of materials to keep up and solve those problems (Miao et al., 2022). Data can be defined as objects of the real world that can be retrieved, stored and elaborated through a software process and communicated via a network (Riedel & Rass, 2019). The amount of global data used shows exponential growth, leading to the constant need for organisations to think about deriving value from that data. Therefore, the data is also considered the "oil" of the digital age (Kambli, 2019; Zhang et al., 2022).

However, not all organisations have the expertise to recognise, let alone develop, the possibilities of data usage. Researchers, for example, use the terms 'big data', 'data analytics' and 'business analytics' interchangeably (Zhang et al., 2021). However, what do those terms entail, and where can organisations start to implement or develop them? Consulting firms specialised in data usage can offer this information and help their clients develop.

Current research describes how data can be used and what it consists of. However, there is an absence of an overview of what factors data consulting can include. This paper provides insight into such an overview by extensively combining the different research.

In order to do so, a collaboration with Little Rocket has been realised. Little Rocket is a business-to-business (B2B) data consulting SME located in Enschede. Small and medium-sized enterprises (SMEs) can offer their services to their customers by using their own experiences and reflecting on how competitors perform similar services. SMEs form the majority of firms in most countries and the large majority of jobs. Therefore, governments worldwide often increase their efforts to promote and support SME expansion (Bayraktar & Algan, 2019).

Little Rocket has its expertise split up into five pillars and twenty-five sub-pillars, which can be found in Table A.1 (of the Appendix), including a brief description of what the pillars entail. The sub-pillars form a subdivision of the pillars to distinguish further what parts a pillar consists of. These (sub-)pillars have mainly been formed by Little Rocket's practical experience and are used to communicate with (possible) clients.

Little Rocket uses the pillars to check how (possible) clients score in all five categories. This assessment is mainly done in an open conversation where the client describes its current situation in (lack of) data usage. Little Rocket can then explain the growth possibilities that their help can bring. Often, clients do not have the financial option to develop all pillars (Willetts et al., 2020). However, the desired growth can already be obtained by starting, developing or maximising one or two of them. This is dependent on the client, its strategy, resources and goals.

Ideally, the sub-pillars would be used in a maturity framework, which would be an assessment tool for evaluating how far a client's level of progress is for specific parts (the sub-pillars) of data usage (Contributor, 2019). All sub-pillars would have the same maturity levels, making it possible to have an overview for a client quickly. The next step is identifying a suitable improvement strategy (Storbjerg et al., 2016).

Due to time constraints, this research will only be conducted using the first three pillars, namely strategy, value, and fundament, as well as their fourteen respective sub-pillars. These three pillars can be clustered together in doing research quite well, as can be seen in the data science hierarchy of needs in Figure A.3 (of the Appendix). This model, inspired by Maslow's famous hierarchy of needs pyramid, suggests that a lower layer (starting from the bottom) must be fulfilled to start with the next layer. The Figure includes sub-pillars of the first three pillars, indicating a solid base for this report. However, suggestions will be made for future research, including logging useful articles that include useful information about the pillars 'knowledge' and 'team'.

Furthermore, Figure A.3 (of the Appendix) shows how artificial intelligence (AI) is a very advanced technique, as it is in the highest layer. Feng et al. (2024) describe AI as a field of computer science that creates systems and algorithms to perform tasks that typically require human intelligence. These tasks could be problem-solving, learning, understanding language, and decision-making, contributing to various applications across different sectors. Machine Learning (ML) is a crucial technology that supports AI. It is a computational method that allows machines to 'think' or act without being specifically programmed to perform specific actions (Król & Zdonek, 2020).

For now, this research will use the chosen sub-pillars to investigate how literature defines the sub-pillars, what they include, and what steps they require to develop further. It will also provide an analysis of possible connections between sub-pillars. The following research question is proposed for this research: *Based on which pillars can SMEs consult their clients to use data for growth purposes?*

By performing **literature-based research** to answer this research question, Little Rocket, and data consultancies in general, are supposed to be able to improve their services in several aspects. By backing up their expertise with scientific research, they can convince possible clients to use their services even better. Besides that, new information could be found on how (sub-)pillars relate, possibly improving efficiency on how separate pillars are improved for a client. Lastly, making it possible to use (literature-based) maturity levels can allow clients to understand their current and desired position even better.

There is also the possibility that certain sub-pillars are not researched much, meaning they might not be that interesting or that there still lay opportunities. Besides that, practice might show that the data science hierarchy of needs (Figure A.3 of the Appendix) layers do not always need to complete a layer below in order to start a project on the higher layer; as Little Rocket indicates, it is possible to develop some sub-pillars and not necessarily all of them. This is an interesting point to consider in the literature research.

The theoretical implication of this report lies in having a research-based overview which confirms or challenges the used literature, whilst the practical implication lies in the possibility of applying the theory in the field. Clients can be informed by information that is backed up by literature (written by experts in specific research fields) that can be read online. A limitation can be that not all twenty-five sub-pillars can be researched in depth. Besides that, the research used in this report is based on domain-specific use cases. These use cases might differ slightly from the diverse sectors in which a data consultancy's client operates. However, that information can lead to valuable results as well.

The remainder of the paper will be structured as follows. After the introduction, the methodology chapter explains the research design, type of literature review used, and how data is collected and analysed. In the third chapter, the findings of the literature will be stated. In the following chapter, the findings will be further analysed and discussed, as well as how they contribute to theory and practice. Besides that, possible limitations and recommendations for future research will be brought up. Lastly, the conclusion includes a summary of the main points of this paper.

2. METHODOLOGY

2.1 Systematic Literature Review

Considering the need for researching and analysing (in the direction of) the pillars and the lack of summarised data for these pillars, a **systematic literature review** (SLR) is required. A SLR is "a review of an existing body of literature which follows a reproducible and transparent methodology in searching and assessing its quality, maintaining a high level of objectivity" (Kraus et al., 2020, p. 1026).

The difference between traditional and systematic literature reviews (also called systematic reviews) is mostly in how data is collected and the possibilities of replication (Okoli, 2015). Traditional reviews are less strict in collecting and performing quality control of sources, making it highly possible for the author to choose subjectively (Tranfield et al., 2003). Kraus et al. (2020) argue that an advantage can be that the author is experienced and can use their intuition in choosing sources. However, nowadays, electronic databases can assist in performing SLR quicker and more transparently. Therefore, the result of the data used is objective, which is preferred in this research.

Literature shows that the amount of stages and steps a systematic review should contain can vary. For example, Okoli (2015) and Pittaway et al. (2016) consider four stages, whilst Tranfield et al. (2003) consider three stages. The number of steps in those stages also differs. However, according to Kraus et al. (2020), the main description of all different versions refers to the same main steps: *planning the review, conducting the review* and *reporting the findings*. A possible extra step can be added regarding disseminating the findings (see Table 2).

Table 2. Process of SLR (Kraus et al., 2020)

Stage 1: Planning the review
1.1 Identify the need
1.2 Develop protocol
Stage 2: Identifying and evaluating studies
Stage 3: Extracting and synthesizing data
3.1 Conducting data extraction
3.2 Conducting data synthesis
Stage 4: Disseminate the review findings

The first and second stages are described up until Chapter *2. Methodology*. After that, the extracting of data can be found in Chapter 3. *Findings*. Then, the data is synthesised in Chapter *4. Discussion*, where it will further be reflected on, and connections between the data can be made. It will conclude with the research dissemination by stating recommendations (for future research).

When conducting the SLR on the first fourteen sub-pillars of Little Rocket, it will also be essential to reflect on whether connections occur between sub-pillars. Pittaway et al. (2016) argue that a good and careful SLR can form a new theoretical construct, which is one of the goals of this paper. Stating connections can further improve the quality of this theoretical construct. To do so, it is also essential that the limitations of papers are admitted and transparently listed (Hermann & Hatak, 2014).

2.2 Data Collection

When collecting data, eligibility criteria have to be decided upon. These inclusion and exclusion criteria can assess whether a source is relevant and valuable for this research.

In this research, document types such as journal articles, book chapters, conference papers, and editorials, all in the English language, have been included. Sources had to include keywords related to the sub-pillars to be relevant for this research. All of the sources have been published between (1995 – 2024).

Furthermore, databases Web of Science and Scopus were chosen as the preferred information sources due to their wide variety of sources and range of filters. Sometimes, Google Scholar was added as an information source to increase the search range. The increasing amount of databases also increased the chances of finding double articles.

Specific search strings were made for all fourteen sub-pillars used. These include the sub-pillars themselves and keywords related to the sub-pillar to keep the research broad and prevent the funnelling of sources by sub-pillars that Little Rocket uses. Using truncation symbols and proximity symbols made it possible to note the string as short and efficient as possible while keeping the search area comprehensive (*Database Search Tips: Truncation*). The search strings are in the top-right corner of Figures B.7, B.8 and B.9 (of the Appendix). After using the search string, extra filters and limitations could be applied based on the amount and quality of the results. The filters used were sorted by 'times cited', publication year, article-only, and 'relevance'. Those exact Figures also show what extra filters and limitations were used per sub-pillar.

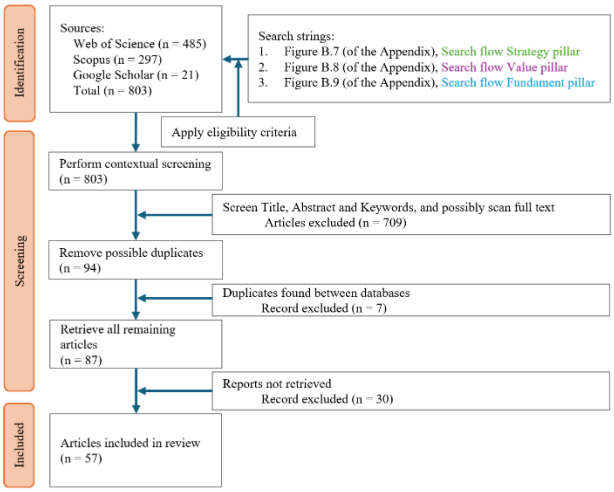

Figure 1. SLR selection process

Figure 1 represents the SLR selection process. It consists of information of the specific search flows of the three used pillars. These search flows are stated in Figures B.7, B.8 and B.9 (of the Appendix).

By combining these Figures, Figure 1 built the selection process for all sources of the SLR. The design of these Figures is inspired by the PRISMA Flow Diagram of Sarkis-Onofre et al. (2021) and the source selection process of Boiral et al. (2017).

Through the selection process, 57 sources were found to be relevant to this research. The combination of 485 sources of Web of Science, 297 sources of Scopus and 21 sources of Google Scholar led to an initial cluster of 803 sources after applying the eligibility criteria to the search string results. In the next step of identifying sources, could the screening take place. Sources considered not valuable by screening the title, abstract, keywords and sometimes the whole text, had been excluded.

When sources stated the same, such as Figure A.3 (of the Appendix) and the first Figure in the source of Jain et al. (2020), the source that included most (sub-)pillars would be chosen. In this case, Figure A.3 (of the Appendix) of Informatica.com (2024) was chosen as this version also includes valuable information for the team pillar.

The exclusion of 709 sources made the cluster of 803 potential sources shrink to 94 sources. After that, duplicate sources were removed. After that, all left sources had to be retrieved, which was not always possible. Out of the 87 sources, 30 were not available online for free. This left over the 57 sources.

Figure 2 shows the timely overview of these 57 sources. It includes an outlier of a source published in 1995. Often, sources with more recent publication dates were preferred. However, this example of Kotter (1995) shows that some theories can be used for decades, as it was cited very often by recent studies.

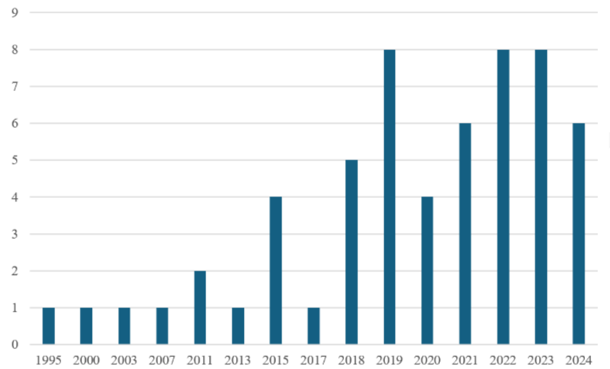

Figure 2. Timely overview of SLR sources

All 57 sources were read, whilst significant findings were noted in the content analysis (Table C.7 of the Appendix). Besides that, they were noted in Table B.5 (of the Appendix), with a red 'X' on the coordinate of that source with the specific sub-pillar for which that source was selected. However, sources often also included information and connections of other sub-pillars, which were noted in the respective colour of that other sub-pillar. For example, the first source, 'A Coherent Strategy for Data Security through Data Governance (Trope et al., 2007),' was initially selected after using the search string and selection process of Figure B.7 (of the Appendix) for *1.1 Governance*. After analysing the source, it appeared that the sub-pillar *3.3 Security, Privacy, Compliance* was also important, and therefore marked in Table B.5 (of the Appendix) with a blue 'X'. Note that the pillars 'knowledge' and 'team' have been added in grey so that future research can be performed using marked sources in their columns. In the lowest two rows of the Table, the sums of red X's, and other X's in general, can be found. So, this indicates how many sources were explicitly found for a sub-pillar and the sum of sources that a sub-pillar was considered important in.

Based on Table B.5 (of the Appendix), Table B.6 (of the Appendix) can be made, indicating how often sub-pillars have been considered important in sources initially found using another sub-pillar's search string. The left/first column (with sub-pillars) indicates the sources with red X's in Table B.5 (of the Appendix). The top row (with vertically written sub-pillars) indicates the sum of another sub-pillar being considered important.

2.3 Data Analysis

Performing the SLR and analysing the content of the sources leads to several types of conclusions. Firstly, the amount of available sources after using certain search strings and eligibility criteria indicates how often the sub-pillar is researched. Secondly, after performing the follow-up filtering of removing sources by scanning the title, abstract, and content, it becomes clear how many sources can actually be of good use for this research. Unfortunately, a limitation lies in having to delete the doubles and removing unretrievable articles.

The content of the sources selected by the search flow process can be analysed, and important findings are noted in the content analysis (Table C.7 of the Appendix). Besides that, the links between sub-pillars in those sources also offer valuable information; connections can be found with this information. It will be clear what sub-pillars are mentioned often and which are not, and what sub-pillars are often related to each other. Follow-up questions can arise on whether less searched sub-pillars are 1) less important in general and therefore researched less, 2) might be important but understudied, here may lay opportunities, or 3) more useful sources were not found do to errors in the search flow, such as the phrasing of the search strings. Lastly, the content of several sources might confirm, challenge or extend one another; this could also lead to useful future research recommendations.

The analyses that are found in *4. Discussion* are backed up by the summarised content of all sources in the following Chapter 3. *Findings*. With this information, SME data consultants can assess if clients already fulfil important parts of sub-pillars or if they are missing some fundamental parts that can improve their success in data usage.

3. FINDINGS

This chapter states the findings of the conducted SLR per individual pillar with its respective sub-pillars. As Table B.5 (of the Appendix) suggests, sources have also been used across sub-pillars to point out connections and provide more depth.

3.1 Strategy Sub-Pillars

3.1.1 Governance

Research divides data governance into several streams, as it is considered a broad concept. Zhang et al. (2022) consider data governance as the legitimate and disciplined management of data into a strategic asset for a firm, including the intervention of procedures, rules and values. In short, the exercising of authority, responsibility and control of the management of the data (Abraham et al., 2019; Al-Ruithe et al., 2018). The goal is to increase data quality and trust by overseeing security strategies (Trope et al., 2007; van Helvoirt & Weigand, 2015). Li et al. (2019) separated big data governance over three domains and imagined the ability for it to function as the ability for humans to run: drive domain (consider whether it can run), support domain (determines how fast it can run) and capability domain (determines the length it can run). Basim Alwan and

Ku-Mahamud (2020) state that big data contains large, complex volumes. The complexity and size can often be measured by volume, velocity, variety and veracity; the 4 V's.

In the research of Zorrilla and Yebenes (2022) on data governance systems for Industry 4.0 (I4.0), data governance is stated to gain valuable insights and thus perform better decision-making. Besides that, it acts as a crucial step before being able to apply more advanced technologies such as AI (Monah et al., 2022). Zorrilla and Yebenes (2022) state the general requirements a governance system must meet, namely the stakeholder's needs, generating value using data, consisting of several components that work together as a whole, being dynamic, clearly distinguishing between data management and governance, being adjusted to the organisation's needs and taking into account the whole organisation.

More specific requirements a data governance system must meet can be categorised into four groups: Principles, Governance, Management and Monitoring.

The principles determine what must be considered when implementing a data governance system. They are guidelines that direct the firm's action, behaviour, and philosophy regarding data usage, management, and governance.

Governance ensures that the data governance structure and objectives are aligned with the firm's business plan, including monitoring to verify that this is done right. It also includes the collection of required company information to use for decision-making. Besides that, the organisational model in which the data governance system will operate has to be settled. International (2017) provide a high-level overview of these models. Furthermore, policies, standards, and functions of evaluating, directing and supervising must be defined. Management includes many sectors: metadata requirement, classification, data quality, data security and privacy, and data life cycle. Lastly, key performance indicators (KPIs) must be defined for monitoring to check if data usage performs well, if the strategy is lived up to, and if this is done according to the policies set.

Many companies have modelled their governance system in a maturity model. These models often define between the four and six steps of maturity. Mahanti (2021) discusses and compares the models of six companies: Kalido, DataFlux, Microsoft, Informatica, Oracle and IBM. It is stated that organisations in highly regulated industries tend to have more mature data governance than organisations in low-regulated industries. They have also proposed a model that combines the information of the six mentioned companies. This model can be found in Table A.3 (of the Appendix). International (2017) their framework in Figure A.4 (of the Appendix) has similarities. It seems that data governance is an overarching term for many other sub-pillars.

3.1.2 *Vision, Strategy Development*

Data strategy has to be aligned with the firm's overall organisation's strategy (Bhargava et al., 2020). It provides a path to achieve the organisation's goal and long-term vision (Kambli, 2019). The extensive paper "What's Your Data Strategy?" written by DalleMule and Davenport (2017), is a framework that considers trade-offs between "defensive" and "offensive" data usage as well as balancing control and flexibility. With this article, a data strategy can be formed.

Defensive strategies minimise downside risk by ensuring compliance with regulations, detecting and limiting fraud by data usage and building systems that prevent theft. They concentrate on compliance, IT, legal and financial concerns.

On the other hand, offence strategies are primarily customer-focused (as well as profitability) and more real-time, such as in sales and marketing. The offensive positioning focuses on market power by taking a more proactive position (Munhoz de Medeiros et al., 2020). It includes performing data analysis and modelling to generate customer insights or, for example, integrating customer and market data to make dashboards that support managerial decision-making.

Every firm must find a balance between the two, which is not necessarily evenly divided. Some sectors, for example, finance or healthcare, push strategies towards defence due to the regulations. Sectors with solid customer competition move towards offence; this is visualised in Figure A.5 (of the Appendix). Note that a company's position on this spectrum is rarely static.

The more data is standardised, the better it can be used for defensive purposes. On the contrary, the more flexible the data, the more readily it is for transitions and thus an offensive purpose.

Munhoz de Medeiros et al. (2020) discovered that the offensive strategy directly and positively impacts gaining a competitive advantage while mediating the relationship between the defensive strategy and that competitive advantage.

3.1.3 *Data quality*

Azeroual et al. (2018, p. 1277) define data quality as a "multi-dimensional measure of the suitability of data to fulfil the purpose bound in its acquisition/generation". It is possible that the suitability changes over time as the needs change. Reliability measurements must be performed using defined standards to determine the data's quality. As times change, these defined standards must also be able to be changed.

Often, business decisions strongly rely on some proportions of data to be available inside an organisation. Thus, low levels of data quality can have far-reaching consequences, such as missing business opportunities or making poor decisions (Riedel & Rass, 2019).

The nature of data can be categorised in several ways. For example, manufacturing has three types: Raw Data Items, Component Data Items and Information Products. Raw data items are small data units used to create information and component data items. Component data items are stored temporarily until a final product is made. Information products are the consequences of performing manufacturing activities. Another way of classifying data is by structuring it into structured, semi-structured, or unstructured data. Riedel and Rass (2019) compared data quality frameworks for all business environments, filtering which works best for every situation based on the above data categories.

Combining literature in their article, Azeroual et al. (2018) state that data quality requirements can be divided into four categories. These are accuracy, relevancy, representation and accessibility. Furthermore, there are four dimensions of the data quality. These are completeness (data has sufficient breadth, depth and scope for the task), correctness (data is reliable and correct), timeliness (the data has appropriate age for the task) and consistency (data are presented in the same format and compatible with previously used data). Ridzuan and Zainon (2024) have researched sixteen papers in this field and state that the four dimensions are indeed necessary to measure data quality. Thus, the literature states the same findings but in slightly different ways or wording.

3.1.4 *Metadata Management*

Metadata can offer significant benefits, such as searching for datasets in web repositories, schema matching, data integration, and exploring data lakes (Khalid & Zimányi, 2024). Simply put, metadata is "data about data" (Quimbert et al., 2020). A less simplistic definition is as follows: "Metadata is structured information that locates, describes, explains or otherwise makes it easier to use, retrieve or manage an information resource." (Allen & Cervo, 2015a, p. 1).

There are many metadata categories, with different experts having different categories. In the paper of Allen and Cervo (2015a), Ralph Kimball, a respected author and expert in the field, was used. He distinguishes business metadata, technical metadata and process/operational metadata. These categories should not be handled in the same way and should be prioritised accordingly, for example, by considering the source of data and the data elements that are part of that source.

Metadata management captures the context required to understand the elements of a dataset and their usage and is, therefore, a critical element of data governance. Besides proactively preventing issues by properly documenting enterprise data, it also identifies areas that need improvements (such as data quality, stewardship or governance).

Metadata management is about "capturing the definition, context, description, and lineage of all information spread across the enterprise" (Allen & Cervo, 2015a, p. 163). This information is then stored in an easily accessible repository so that it can be retrieved and distributed quickly when needed.

Metadata management aims to minimise adverse effects and support as many key activities as possible. It is also a very critical task to perform to have a successful BI program due to the following reasons (Loshin, 2013):

- Metadata covers information (e.g. conceptual, logical and physical) as loose datasets, which can be transformed into data models useful for analysis.
- Metadata captures the structure of the data, which can be used for BI and warehousing.
- Metadata management processes provide a way to split different meanings associated with source data. This provides methods for analysts to ensure coherence once that data is available for analytics and reporting.
- Differences can be captured in how data is manipulated over time. This is critical, especially for data warehouses with extensive historical data spans.
- Metadata traces the evolution of information to validate and verify its results retrieved from an analytical process.

3.1.5 Change Management

Change management is critical to keep up with a continuously changing and dynamic business environment. It is, therefore, considered an essential capability for an organisation. Continuously assessing and renewing the organisation's direction, capabilities and structure as a response to the changing demands of stakeholders are vital points of organisational change management (Mitra et al., 2019).

Traditional change management principles are about prioritising structured planning, having clear communication, and engaging stakeholders when implementing a change in the organisation. Kotter (1995) developed a strategic-level model that describes eight steps to follow those principles and transform an organisation (Ademola, 2024; Mitra et al., 2019; Pachamanova et al., 2022; Raineri, 2011). These steps are as follows: 1) creating a critical need for action, 2) building a strong leadership team, 3) creating a clear strategy, 4) sharing the strategy widely, 5) enabling others to implement the strategy, 6) planning for, and creating short-term wins, 7) consolidating improvements while still producing more change and 8) institutionalising new approaches.

Next to Kotter's model is Lewin's Force Field analysis, also known as the three-step model, a popular and effective model to use in change management. The three steps include unfreezing, changing and refreezing processes and systems in an organisation, after which the change is adopted. In the first step, the existing, stable equilibrium in the organisation is altered whilst maintaining current behaviour and attitude. In the second step, cognitive reconstructing occurs, where the actors acquire information and evidence which support their idea that the change is needed and possible. Finally, in the third step, the organisation reaches a new equilibrium; all changes have been made permanent (Mitra et al., 2019).

Ademola (2024) states that, with the evolving organisational landscape, the practices of change management shift towards 1) the adoption of agile change management methodologies, 2) the embracing of a culture of continuous learning and adaptability and 3) the integration of AI in change management processes. AI appliance leads to opportunities for organisations to leverage advanced technologies for change management purposes. These technologies facilitate more advanced data analysis, decision-making and predictive modelling, making change management processes more informed and efficient. More information will follow in Chapter 3*.2.2 Data Analytics*.

3.2 Value Sub-Pillars

During the SLR selection process, it became clear that the two sub-pillars 2.2 *Descriptive and Diagnostic Analysis* and 2.3 *Data Science, Forecasting, Statistics, NLP, ML, AI* (as stated in Table A.1 of the Appendix) are very often mentioned together in literature (notice how four sources state that, in Table B.5 of the Appendix). Mainly because the first of the two sub-pillars focuses on (analysing) historical data, but the second-mentioned sub-pillar goes a step further and makes predictions and forecasts. Table A.4 (of the Appendix) visualises how literature recognises the steps in developing data analytics. Therefore, in the content analysis (Figure C.7 of the Appendix) and the following sub-chapters, the two sub-pillars have been merged into one sub-pillar: *2.2 Data Analytics*. Further explanation and recommendations are stated in Chapter *4. Discussion*.

3.2.1 Analytics Translation, Business Consultancy

Analytics translation and business consultancy include using use cases, the extent to which these use cases bring helpful insight, and how they result in data consulting strategic decisions. Dobing and Parsons (2000, p. 1) define a use case as a "description of a sequence of actions constituting a complete task or transaction in an application."

Schwarz et al. (2023) made a framework that can be used to systematically and structured evaluate if a use case has potential and is feasible. It includes nine criteria that are clustered into three different groups: potential, technical feasibility, and economic feasibility. The potential is from an organisational perspective, technical feasibility specifies obstacles to using data and analytics, and the economic feasibility checks the costs and timeliness. Figure A.6 (of the Appendix) states how the nine criteria are divided into those three groups.

The added value is an estimated value. Use case value can be distinguished into four dimensions: value network, value proposition, value capture or value creation. The strategic fit, also part of the 'potential' header, evaluates if the use case fits the firm's philosophy, image, objectives, management interests, etcetera.

Data availability refers to the degree to which it is possible to access data instantly. The accessibility is based on the sensitivity of the data (public/open, internal, confidential or restricted data). Data quality has already been discussed in Chapter *3.1.3 Data Quality*. Data security measurements are in Chapter *3.3.3 Security, Privacy, Compliance*. Furthermore, it is helpful to perform gap analyses to evaluate the tools, technology, and expertise needed to conclude all criteria for technical feasibility. Lastly, the cost and time estimations must be made, which is difficult in the early stages. Estimations are still based on the technical feasibility criteria, making the economic feasibility dependent on that.

Furthermore, research shows that by improving the data architecture coherence across a firm, increased ML capabilities can be used, and more applications and use cases can be developed (Cao & Iansiti, 2022).

3.2.2 Data Analytics

Analytics play a crucial role in decision-making for all kinds of businesses. They help to identify growth and improvement opportunities, optimising operational efficiency and allocation of resources, understand customer behaviour and preferences, mitigate risks and identify (potential) threats and monitor and evaluate the performance of strategies (Cam et al., 2021), as cited in Wolniak and Grebski (2023b)).

As stated in the introduction of Chapter *3.2 Value Sub-Pillars*, literature often describes the path to data analytics maturity with the same steps. As seen in Table A.4 (of the Appendix), all four sources include the steps of descriptive, predictive and prescriptive. The diagnostic step is also included, except by Farrokhizadeh and Öztayşi (2022). As outliers, Wolniak and Grebski (2023b) have included a 'real-time' step, whereas Król and Zdonek (2020) implemented the 'cognitive' step after the prescriptive step. In practice, all steps complement each other and co-exist (Król & Zdonek, 2020). Each step builds upon the previous step; new steps address new needs and challenges (Wolniak & Grebski, 2023b).

This report considers all six steps and compares and elaborates definitions of the used articles. The first step, the descriptive analysis, is a foundational step in data analysis. It offers a comprehensive overview of existing conditions by turning raw data into valuable summaries. Statistical methods, such as mean, mode, median, etcetera, are used to find patterns in the data sets (Farrokhizadeh & Öztayşi, 2022; Ibeh et al., 2024). Data visualisation tools often give a clear overview of trends and key performance indicators (Wolniak & Grebski, 2023b). This type of analysis answers the question: "What happened?" (Król & Zdonek, 2020).

Real-time analytics, also known as instant analytics or streaming analytics, analyse data as it is received without delay. This way, organisations can monitor or respond to events and transactions in real-time, which can be helpful in dynamic environments such as financial markets and online customer transactions (Wolniak & Grebski, 2023b).

Although Farrokhizadeh and Öztayşi (2022) include it in descriptive analytics. Ibeh et al. (2024), Król and Zdonek (2020) and Wolniak and Grebski (2023b) state that diagnostic analytics involve around the question: "Why did it happen?". It detects regularities and relationships between variables by analysing historical data (Król & Zdonek, 2020). Techniques involved are statistical analysis and data mining, along with others, to uncover patterns, correlations and causal relationships (Wolniak & Grebski, 2023a). Liu et al. (2018, p. 280) describe data mining as "the process of discovering patterns in big data sets by possibly extracting information and knowledge from a large number of noisy, incomplete, fuzzy, random amounts of data". It is a combination of statistics, database technology and AI. Diagnostic analytics can help find root causes of, for example, sales fluctuations of firms or patient readmissions in hospitals (Ibeh et al., 2024).

The next step is predictive analysis, which uses statistical algorithms and historical data to predict future events or outcomes by modelling and forecasting (Król & Zdonek, 2020; Wolniak & Grebski, 2023b). The question is: "What will happen in the future?" (Król & Zdonek, 2020). Forecasting represents predicting and anticipating future trends (Angelica & Mariluzia, 2022). The prediction uses more sophisticated techniques than previous steps, such as regression analysis and the appliance of ML algorithms (Ibeh et al., 2024). Firms can use this information to, for example, anticipate market trends, customer behaviour, and product demand (Wolniak & Grebski, 2023b).

Prescriptive analytics use simulations and ML to suggest actions to achieve desired results, such as risk reductions, sustainable growth, or securing competitive advantage in a volatile market (Ibeh et al., 2024). It supports complex decision-making processes and aims to automate actions (Wolniak & Grebski, 2023b). The question to be answered is: "What actions should be taken?" (Król & Zdonek, 2020).

Lastly, cognitive analytics is mentioned by Król and Zdonek (2020). They describe it as a step further than the previously mentioned prescriptive analysis. Here, AI technologies and high-performance data analysis lead to the automation of decision-making processes and increased efficiency by using intelligent machines. AI systems are easily scalable and can multiply their reach several times without much human intervention (Kambli, 2019). Khaleghian and Shan (2023) point out the importance of a solid foundation of data usage before it is possible to adapt AI and ML algorithms.

Challenges in data analytics, as Wolniak and Grebski (2023a) for example describe for performing diagnostic analysis, can be time-intensity, heavy dependence on data quality and availability, the complexity of the analyses, data privacy and security and preventing bias and misinterpretation in data usage.

3.2.3 Business Intelligence

Business Intelligence (BI) includes collecting, storing, and analysing data. It can help the data analytics described in Chapter *3.2.2 Data Analytics*, dependent on the structuring and accessibility, leading to improved strategic performance management (Kambli, 2019; Sharma & Djiaw, 2011).

One of the goals of BI is to combine data from several datasets of different business areas for reporting and analysis. To do so, designers and developers need to understand the source system's flow of information, and proper metadata is also essential (Loshin, 2013). BI-driven dashboards are then made, which are interactive interfaces that visualise and analyse the performance metrics of a firm. They include indicators, graphs, and tables in combination with interactive features to provide decision-makers with a consistent yet flexible representation of the firm. As decision-makers rely on these dashboards, they must be correct so that possible business ideas are not turned down unjustly (Burnay et al., 2024). Furthermore, Cao and Iansiti (2022) note that the interaction between these analytic dashboards and predictive models (as stated in Chapter *3.2.2 Data Analytics)* involves product-specific routines, communication channels and workflows. These are often shaped over time, as the data team has to adapt to new technologies and configurations with the existing systems.

Lastly, sources were sought regarding the level of dashboard implementation by organisations. Nevertheless, these were not found.

3.3 Fundament Sub-Pillars

3.3.1 Data Platform

Ma (2023) states that a data platform is a new transformational architecture. Modern data platforms can be composed of several functional elements, but they remain very centrally built in an organisation (Loukiala et al., 2021). It is a platform that is organic and integrated, used for the sharing of capabilities, which empowers an organisation with nine capabilities: 1) data service capability, 2) capability to develop data applications, 3) capability to process data, 4) capability to develop data, 5) self-learning and automated improvement capability, 6) capability to accumulate assets, 7) capability to auto-track data quality, 8) capability to integrate data and 9) IT system & DT system risk isolation capability.

There are several ways a data platform can be applied. Firstly, it can assist a business department in performing data analysis. Secondly, it helps business, technical and external departments flexibly create applications. Thirdly, the technical compartment can construct the application's capabilities to accumulate data assets and their values at a constant rate.

Ma (2023) made some recommendations for constructing a data platform. Start by defining a clear objective: What results can be achieved? Then, a standardised data source and data pooling applications are required. It needs to analyse and integrate different data sources, delete duplicates, and share the data to improve the organisation's response efficiency. Furthermore, checking if the data platform is adequately constructed is essential. If any errors are involved, derivative issues can be expected by the generated applications. Liu et al. (2018) add that, in the case of big data platform construction, the focus has to lie on key technologies such as big data architecture, big data modelling and storage, large data analysis and processing, and big data applications.

3.3.2 Data Architecture, Data Models

DalleMule and Davenport (2017, p. 4) state that "a company's data architecture describes how data is collected, stored, transformed, distributed and consumed". Miao et al. (2022) note that a simple and universal structure should be used for data rights confirmation, management, sharing, and security. It is also a critical component that drives co-invention from data and possibly machine learning (Cao & Iansiti, 2022).

In the research done by Cao and Iansiti (2022), the term data architecture is divided into three aspects: coherence, security and cloud computing. Coherence includes the capabilities by which data processing is integrated, standardised and automated. Costs and frictions of data-related co-invention are reduced by coherence in data architectures. It makes it easier for the data team to make predictive models, leading to more ML capabilities and use cases across product development and multiple business cases. Security is discussed in Chapter *4.3.3 Security, Privacy, Compliance*. Furthermore, cloud computing includes migrating cloud infrastructure to external providers, allowing organisations to reach more data flexibility and robustness.

As part of the data architecture landscape, data models can present the data requirements of an application. These models present all parts of applications, their characteristics and their relationships (March, 2003). A well-designed model is of great value in securing the quality and integrity of data, as well as ease of maintenance and scalability. Therefore, data architecture is also especially important in developing ML capabilities and delivering higher-quality services and products (Cao & Iansiti, 2022).

3.3.3 Security, Privacy, Compliance

Data security entails implementing measures that protect digital information from unauthorised parties' access, disclosure, alteration, or destruction. It includes forming strategies that encrypt, control access, and have firewalls and authentication methods to ensure confidentiality, integrity and data availability (Feng et al., 2024).

Bautista-Villalpando and Abran (2021) propose a security framework that can be used to protect cloud computing services. Often, cyberattacks happen on these services, which are very difficult for organisations to prevent. This is due to the reliance on insecure networks while using new technologies. The life cycle of the framework is first to identify the data security requirements in cloud computing services, then manage these risks, and then evaluate the performance of this data security. This cycle continues due to the possibility of changing security requirements. Kambli (2019) and Schwarz et al. (2023) add that there are also mandatory regulatory requirements set by the EU, such as the General Data Protection Regulation (GDPR).

Research by Zhang et al. (2021) shows that, on average, businesses that use big data analytics can significantly reduce their systematic risk (e.g. exposure to macro-level economic risks) when investing in data privacy and security. When using AI, this is an even more crucial investment to safeguard sensitive information and keep users' trust (Feng et al., 2024).

3.3.4 Data Catalogue

Data catalogues collect, create and maintain metadata. By centralising this metadata, they offer a stable inventory of data assets whilst making data discovery, interpretation and governance simpler (Boufassil et al., 2023; Ehrlinger et al., 2021; Quimbert et al., 2020; Yee et al., 2023).

According to Ehrlinger et al. (2021), a metadata schema is first required to implement data cataloguing. The following eight steps are advised to make a catalogue; the first five steps represent the forming of a metadata schema: 1) define data context variables, 2) define data attributes, representing data quality, accessibility, sensitivity and accessibility, 3) tag the data; decide which metadata is attached to which particular level (e.g. entity-level), 4) define rules that regulate the accessibility of data or audit, 5) the previous steps combined form an enterprise data model, this is the final data catalogue schema, 6) populate the catalogue with data, 7) expose the data catalogue to users and 8) keep improving the data catalogue by receiving feedback, revisions and reviews.

3.3.5 Master Data Management

MDM is an initiative to improve the core data quality (i.e. master data) of an organisation and to increase data usage across an organisation's business processes (Haug et al., 2023; Vilminko-Heikkinen & Pekkola, 2019). Usually, this data is used across multiple business processes and includes, for example, customer, supplier and product master data (Krause & Becker, 2021).

By assessing the MDM maturity, organisations can identify what aspects of MDM they are developed in and what parts need more attention. This assessment is often performed by using the Master Data Management Maturity Model (MD3M) by Spruit and Pietzka (2015) (Iqbal et al., 2019; Pratama et al., 2018; Rahman et al., 2019). The MD3M considers five key topics, thirteen focus areas and sixty-five capabilities in MDM. The five key topics are data model, data quality, usage and ownership, data protection and maintenance.

Although, based on the literature review done in the report, the MD3M seems to be the most used model, Zúñiga et al. (2018)

note that different sectors might need slightly different maturity frameworks. Allen and Cervo (2015b) agree that too many factors differ between different organisations or domains in a single organisation. This means that there is no one-size-fits-all plan for MDM with multiple domains. However, at all times, when an organisation is expanding to a multiple-domain MDM, an effective top-down management style is always required.

4. DISCUSSION

As found in the literature, data consultancy contains a lot of different topics, which often include several definitions and layers of maturity. In this Chapter will the findings be analysed, providing the answer on the research question, as described in Chapter *2.3 Data Analysis*.

It becomes clear that some sub-pillars are researched (far) more than others, and some sub-pillars are more dependent on one another. Combining information from the SLR found sources in Tables B.5 and B.6 (of the Appendix) can make some general conclusions.

Firstly, it is noticeable how many sources were identified during the identification phase of the search flows (Figures B.7, B.8 and B.9 of the Appendix). In general, the first database used was Web of Science, and in the case of too few useful sources, Scopus was also used. After that, some (specific) sources might still be missing, so Google Scholar helped with providing extra possibilities. The search flow of the individual pillars shows that for strategy-related sub-pillars, almost all sources were found on Web of Science. This was different for the value-related sub-pillars, as Scopus was the database used the most. The fundament-related searches used Web of Science and Scopus for half of the found sources. After that, it also became clear that using only the eligibility criteria and search strings, the strategy and fundament pillars have around three times as many sources as the value pillar. Meanwhile, those two pillars also applied more eligibility criteria than the value pillar, which only applied the criteria of the English language and sometimes the document type.

Furthermore, Table B.7 (of the Appendix) includes the sum of sources that were found specifically for a sub-pillar (the red X's) in the second-last row. Here, it becomes clear that the sub-pillars *1.1 Governance* and *3.5 MDM* could find plenty of valuable sources, as eight and nine sources have been used in this research. There were even more possible sources to scan on their title, keywords and abstract. However, due to time constraints, this was not possible. Besides that, enough information seemed to have already been found. For other sub-pillars, this was not as easy. Especially the sub-pillars *2.1 Analytics Translation, Business Consultancy* and *2.4 Business Intelligence* lacked sources, and only two were found for both sub-pillars. All other sub-pillars were between three and five sources used for the research.

The absence of availability for possible useful sources, found in the screening phase of the search flows, also did not help in this case. The search flow for the strategy pillar only contains six sources that could not be retrieved. This led to twenty-five sources still being available and included in the research. However, half of the sources were not available for the search flow of the value pillar. These ten unavailable sources were mainly spread over *2.1 Analytics Translation, Business Consultancy* and *2.4 Business Intelligence*, leading to the earlier mentioned difficulties on source shortage. Fourteen out of thirty-six sources could not be retrieved for the fundament pillar. A significant number, but the twenty-two remaining sources still served as a solid information base.

The source searching for the value pillar was done after the source searching for the strategy and fundament pillar. This way, it was possible to research the value pillar using the connections with sources of other pillars. The lowest row of Table B.5 (of the Appendix) visualises this, as it sums up the number of times a sub-pillar was considered important in a source. That way, it was possible to retrieve sufficient information for *2.1 Analytics Translation, Business Consultancy* and *2.4 Business Intelligence*.

It became clear that the two sub-pillars *2.2 Descriptive and Diagnostic Analysis* and *2.3 Data Science, Forecasting, Statistic, NLP, ML, AI* were very often mentioned together; literature combined the two sub-pillars. Table B.5 (of the Appendix) visualises this with the red X's next to each other for several sources. A connection that is not noted for any other sub-pillars. Besides that, the linkage is also visualised in the very centre of Table B.6 (of the Appendix). Here, it can be seen that for both sub-pillars, four sources relate to the other one of the sub-pillars. This led to combining the sub-pillars from Chapter 3.*2 Value Sub-Pillars* onwards.

Besides that, more information can be retrieved from Table B.6 (of the Appendix) from the right two columns. These show that governance-related sources often include other sub-pillars; six sub-pillars have been considered important eighteen times in the governance sources. Besides that, MDM-related sources have included five sub-pillars ten times, which is also a significant amount. However, for four of those references, *1.1 Governance* was counted, narrowing down the sources' width. Interestingly, this is not the case the other way around, as governance-found sources never seem to include MDM as an important topic.

Besides these general findings, based on the SLR and sub-pillars found in sources, theoretical and practical implications can also be made. These are based on the most critical information of all used sources of the SLR, which are placed in Chapter *3. Findings*.

4.1 Theoretical Implications

Table A.3 and Figure A.4 (of the Appendix) state that Data Governance consists of quite some topics. The Table and Figure overlap in those topics, where sometimes a different choice of words is used to describe the same. The fact that these Table and Figure include so much overlap is partly confirmed by the information in Chapter *3. Findings* make it seem like they are reliable sources. As a data consulting SME, using them as a base overview when consulting clients on data usage might be helpful, as (the development of) multiple sub-pillars is considered at once.

Furthermore, sources often referred to advanced technologies, including AI and ML, as valuable and necessary tools to fulfil specific data analytics. However, the literature states that the proper fundament of several sub-pillars is required to use such technologies. This information partly confirms the structure of the Data Science Hierarchy of Needs (Figure A.3 of the Appendix). AI (and Deep Learning) are at the top of the pyramid, meaning all steps below are needed to use AI. When a client wants to use simple ML algorithms (the second-highest layer in the Figure), which is the case starting from diagnostic analysis, the bottom four layers must be fulfilled. However, the third-highest layer of the Figure already states the term ‘Analytics’. This paper challenges the hierarchy in what is considered ‘Analytics’ in this stage; according to the sources used for the SLR (also used for Table A.4 of the Appendix), only descriptive (and perhaps real-time) analytics would be involved in that third-highest layer.

Furthermore, note how all steps in Chapter *3.2.2 Data Analytics* are written down extensively. For a client's request in data usage, this sub-pillar conducts what is required of many other sub-pillars.

Lastly, Figure A.6 (of the Appendix) showcases the criteria for evaluating whether a client should apply a use case. This paper includes research that concludes that the more advanced the use case, the more dependent the use case is on data quality, availability, privacy, security and advanced technologies. The expertise can be questioned in future research, as this report does not include literature on pillars *4. Knowledge* and *5. Team*.

4.2 Practical Implications

The findings can also be used for practical applications.

Organisations in highly regulated industries tend to be more mature regarding data governance. Other organisations, more often in lower regulated industries, have their governance processes undeveloped. There are no formally defined roles, data ownership, or metrics for measurement and tracking. When combining this information with the data-strategy spectrum (Figure A.5 of the Appendix), it can be stated that based on research, hospitals are more likely to have a more mature data governance than banks or (more probably) retailers. This is because defensive strategies are often found in highly regulated sectors and vice-versa. These are all examples on that spectrum, which are not static. However, it does improve the chances that an SME data consultant either gives more precise advice or is more efficient time-wise. All by first assessing the sector a client is active in. For example, when assessing al sub-pillars at a client in the health sector, it might be clear quickly that the governance is well organised already, but growth opportunities lie mainly at other sub-pillars. Alternatively, when this is not the case (and the data governance is underdeveloped), it might show why the organisation performs worse than similar parties in its sector. When consulting, remember that research shows that having an offensive strategy can positively impact gaining a competitive advantage. Therefore, for example, a retailer with underdeveloped governance might have gained a competitive advantage. The strategy could remain the same, but the advantage can be maintained/expanded by increasing the data governance structure. Several variables play a role in correctly consulting the client or at least sketching the future situation in the best way possible.

Metadata and data quality also play a role in this situation. When assessing the client's sector, the literature furthermore points out that more standardised data is required for defensive purposes. Thus, mainly for companies that are often located in highly regulated sectors. Meanwhile, flexible data is useful for other organisations to adapt more quickly.

Besides the client's work field, there can also be a distinction on whether the client uses big data. As described in Chapter *1. Introduction* and Chapter *3.1.1 Governance*, terms such as 'big data' and 'data analytics' are used interchangeably in the literature for describing data-driven decision-making. The size and complexity of the data can distinguish 'big data' from 'data' by 4 V's. This could change the way an SME data consultant categorises its clients. However, it seems like future research is needed on where to precisely draw the line on whether the 4 V's are met, or that they should partly be met, and how this might affect the consulting (and for what sub-pillars this is of more importance than others).

Another practical application is that literature often combines several sub-pillars regarding data platform, architecture, data lakes and warehouses (DalleMule & Davenport, 2017; Liu et al., 2018; Loukiala et al., 2021). When SMEs consult their clients, especially clients that are not data-driven, it might be helpful to use the overarching term "Data Ecosystem Infrastructure". This term might be more understandable when a client first hears about data storage technology and allocation in general.

4.3 Limitations and Future Research

Some limitations must be discussed, including potential biases and errors in this report. Some limitations are recommendations that can assist in (defining) future research.

Firstly, the base of this research lies in the pillars and sub-pillars mentioned by Little Rocket. These (sub-)pillars have been formed based on years of experience by the data professionals at Little Rocket. They know the market, use cases, challenges and opportunities in data consultancy. However, it remains the only source used to guide the SLR in this report. As mentioned in Chapter *1. Introduction*, and can be seen in Figures B.7, B.8 and B.9 (of the Appendix), there has been tried not to 'funnel' on only these sub-pillars when searching for sources. Also, other words were used in the search strings. Nevertheless, for future research, it can provide extra depth to gain input of other data consultancy SMEs as well, rather than just one.

Additionally, sources were often unavailable when searching for useful sources and performing the SLR selection process. Based on the available title and abstract, sources showed high potential for being useful for the rest of this research. Nevertheless, it was impossible to open the source due to a lack of license in 30 of the 87 cases. Having lost almost one-third of potentially useful sources, more subscriptions/licences would be available in an ideal future situation to support this research.

Furthermore, due to time constraints, it was impossible to research all five pillars and twenty-five sub-pillars of Little Rocket. During the research, plenty of sources referred to 'knowledge' and 'team' related topics that could not be discussed further. Some examples are Ehrlinger et al. (2021), who state that having responsible persons for data is a crucial aspect of data catalogue implementation and Ademola (2024), who notices that upskilling and reskilling of the workforce is needed in order to effectively embrace advanced technology, such as AI, to improve change management. In order to support future research, Table B.5 (of the Appendix) has sources that contain useful information about the two pillars. Another example is Figure A.3 (of the Appendix) which includes such information, useful for future research. With extra effort, not needed for this research, can future research be boosted.

Next to that, when consulting a client, from the perspective of a data consultancy SME, it can be useful to have some maturity steps for each sub-pillar. That way, clients can quickly recognise their current situation and the steps required to reach a desired situation. Whatever is aligned with their resources and set strategy. Sometimes, theory has provided these maturity levels, such as in Chapter 3*.2.2 Data Analytics*, were it was possible to summarise literature into Table A.4 (of the Appendix). Based on the literature used in this research, finding such levels for all sub-pillars was impossible. Therefore, researchers should continue leveraging qualitative research (including interviews and in-depth case studies) with quantitative research (such as data analytics) to discover more maturity paths in data usage.

Lastly, some sources have included their specific limitations and recommendations for future research. Combining this information with the literature can provide some useful extra insights. In the field of *1.1 Governance*, Zhang et al. (2022) state that further study needs to be done about the evolution of a firm's digital technology investment strategies and governance mechanisms. This could include the allocation of decision-making authority and dynamically adjusting organisational structures. Al-Ruithe et al. (2018) add that, due to the lack of research on cloud data governance, an urge is created to develop a framework that can be used to develop a cloud data governance strategy. This framework has to highlight main pillars, attributes and processes to design a more specific data governance program.

In terms of the *1.3 Data Quality* sub-pillar, Riedel and Rass (2019) mention a lack in research on data quality dimensions on data quality, and thus regulatory compliance. Günther et al. (2019) do note that this aggregation of results into quality dimensions leads to loss of information. Future work should look into methods that improve the traceability of data quality dimensions, such as visualisations of results, to prevent this. Ridzuan and Zainon (2024) agree that this leads to enhanced decision-making, improved business outcomes and competitive advantage for organisations.

For *2.2 Data Analytics*, it is found that there will be an increased amount of available analytics maturity models, personalised for specific domains and sectors, including more entities in the implementation of data analytics in organisations (Król & Zdonek, 2020).

For future work on *3.2 Data Architecture, Data Models* should focus lie on identifying effective approaches to re-architecture parts of a technological system that are difficult to remove but incompatible with new technologies (Cao & Iansiti, 2022).

Boufassil et al. (2023) and Ehrlinger et al. (2021) argue that more research and cooperation will be required for *3.4 Data Catalogue* to continuously improve its design and solve new difficulties in the constantly changing data management environment.

Lastly, the required research is on *3.5 Master Data Management*. Several sources state that the research done was in a single sector, a single case study or just a small amount of case companies, possibly leading to restricted generalisability (Haug et al., 2023; Spruit & Pietzka, 2015; Vilminko-Heikkinen & Pekkola, 2019). However, they state that even research with a single case can be a starting point for further research. In their specific research, Spruit and Pietzka (2015) note that it would be an improvement if the interviewed experts were asked for more feedback on their Master Data Management Maturity Model. In particular, on whether the model would work differently on different companies with different sizes. This could lead to uncovering possible inconsistencies regarding the model's capabilities. Another interesting aspect could be that different capabilities would have different weights. A new matrix could arise, emphasising parts different from the current one.

5. CONCLUSION

This report focussed on which pillars SMEs can consider when consulting their clients to grow their organisation using data. A systematic literature review was conducted, using Little Rocket's first three pillars and fourteen sub-pillars to select sources. These sources were analysed, linked to each other and summarised. All findings were also analysed, leading to a complete overview on what sub-pillars should consist of, how they relate to each other, and which ones can be considered more important. Recommendations, as well as limitations, on (extending) this research have been provided. This research strived to provide a literature perspective on improving SME data consultancy.

6. ACKNOWLEDGEMENTS

I would like to thank my supervisors, Dr. Göttel and Dr. Stienstra, for their guidance throughout my thesis. Next to that, I would like to thank Little Rocket, especially my supervisor Jelle Scholten, for their help from a business perspective. Lastly, I would like to thank my roommates, friends and foremost my family, for their support throughout my bachelor.

7. REFERENCES

* Marked sources have been used in the SLR

1. Abraham, R., Schneider, J., & vom Brocke, J. (2019). Data governance: A conceptual framework, structured review, and research agenda. *International Journal of Information Management*, *49*, 424-438. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2019.07.008
2. *Ademola, O. (2024). Change Management Trends in the AI Modern World: Adapting to the Future of Work. *Advances in Multidisciplinary & Scientific Research Journal Publications*, *10*, 41-50. https://doi.org/10.22624/AIMS/V10N1P3
3. *Al-Ruithe, M., Benkhelifa, E., & Hameed, K. (2018). Data Governance Taxonomy: Cloud versus Non-Cloud. *Sustainability*, *10*(1), Article 95. https://doi.org/10.3390/su10010095
4. *Allen, M., & Cervo, D. (2015a). Chapter 10 - Metadata Management. In M. Allen & D. Cervo (Eds.), *Multi-Domain Master Data Management* (pp. 161-178). Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-0-12-800835-5.00010-5
5. *Allen, M., & Cervo, D. (2015b). Chapter 11 - Performance Measurement. In M. Allen & D. Cervo (Eds.), *Multi-Domain Master Data Management* (pp. 181-194). Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-0-12-800835-5.00011-7
6. *Angelica, B.-C., & Mariluzia, P. (2022, 2022//). The Impact of Data Science, Big Data, Forecasting, and Predictive Analytics on the Efficiency of Business System. Digitalization and Big Data for Resilience and Economic Intelligence, Cham.
7. *Azeroual, O., Saake, G., & Wastl, J. (2018). Data measurement in research information systems: metrics for the evaluation of data quality. *Scientometrics*, *115*(3), 1271-1290. https://doi.org/10.1007/s11192-018-2735-5
8. Basim Alwan, H., & Ku-Mahamud, K. R. (2020). Big data: Definition, characteristics, life cycle, applications, and challenges. IOP Conference Series: Materials Science and Engineering,
9. *Bautista-Villalpando, L.-E., & Abran, A. (2021). A Data Security Framework for Cloud Computing Services. *Computer Systems Science and Engineering*, *37*(2). https://doi.org/10.32604/csse.2021.015437
10. Bayraktar, M., & Algan, N. (2019). *The Importance Of SMEs On World Economies*. https://doi.org/10.36880/C11.02265
11. *Bhargava, H., Rubel, O., Altman, E., Arora, R., Boehnke, J., Daniels, K., Derdenger, T., Kirschner, B., LaFramboise, D., Loupos, P., Parker, G., & Pattabhiramaiah, A. (2020). Platform data strategy. *Marketing Letters*, *31*, 1-12. https://doi.org/10.1007/s11002-020-09539-3
12. Boiral, O., Guillaumie, L., Heras-Saizarbitoria, I., & Tayo, C. (2017). Adoption and Outcomes of ISO 14001: A Systematic Review. *International Journal of Management Reviews*, *In press*, 14-2017. https://doi.org/10.1111/ijmr.12139
13. *Boufassil, A., Bouhafer, F., Cherradi, M., & Haddadi, A. E. (2023, 8-10 Nov. 2023). Data Catalog: Approaches, Trends, and Future Directions. 2023 17th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS),
14. *Burnay, C., Lega, M., & Bouraga, S. (2024). Business intelligence and cognitive loads: Proposition of a dashboard adoption model. *Data & Knowledge Engineering*, *152*, 102310. https://doi.org/https://doi.org/10.1016/j.datak.2024.102310
15. *Cao, R., & Iansiti, M. (2022). Digital transformation, data architecture, and legacy systems. *Journal of Digital Economy*, *1*(1), 1-19. https://doi.org/https://doi.org/10.1016/j.jdec.2022.07.001
16. Contributor, T. (2019). *Maturity grid (maturity model)*. https://www.techtarget.com/whatis/definition/maturity-model
17. *DalleMule, L., & Davenport, T. H. (2017). What's Your Data Strategy? *Harvard Business Review*, 112-121.
18. *Database Search Tips: Truncation*. MIT Libraries. Retrieved 18-4 from https://libguides.mit.edu/c.php?g=175963&p=1158679
19. *Dobing, B., & Parsons, J. (2000). Understanding the Role of Use Cases in UML: A Review and Research Agenda. *J. Database Manag.*, *11*, 28-36. https://doi.org/10.4018/978-1-931777-12-4.ch008
20. *Ehrlinger, L., Schrott, J., Melichar, M., Kirchmayr, N., & Wöß, W. (2021). *Data Catalogs: A Systematic Literature Review and Guidelines to Implementation* Database and Expert Systems Applications - DEXA 2021 Workshops,
21. *Farrokhizadeh, E., & Öztayşi, B. (2022, 2022//). Review of Descriptive Analytics Under Fuzziness. Intelligent and Fuzzy Systems, Cham.
22. *Feng, D., Li, H., Lu, R., Liu, Z., Ni, J., & Zhu, H. (2024). Data security and privacy computing in artificial intelligence. *Journal of Information and Intelligence*, *2*(2), 99-101. https://doi.org/https://doi.org/10.1016/j.jiixd.2024.02.007
23. *Günther, L. C., Colangelo, E., Wiendahl, H.-H., & Bauer, C. (2019). Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises. *Procedia Manufacturing*, *29*, 583-591. https://doi.org/https://doi.org/10.1016/j.promfg.2019.02.114
24. *Haug, A., Staskiewicz, A. M., & Hvam, L. (2023). Strategies for Master Data Management: A Case Study of an International Hearing Healthcare Company. *Information Systems Frontiers*, *25*(5), 1903-1923. https://doi.org/10.1007/s10796-022-10323-z
25. Hermann, F., & Hatak, I. (2014). Doing a Research Literature Review. In (pp. 94-117). https://doi.org/10.4337/9781782540625.00012
26. *Ibeh, C. V., Asuzu, O. F., Olorunsogo, T., Elufioye, O. A., Nduubuisi, N. L., & Daraojimba, A. I. (2024). Business analytics and decision science: A review of techniques in strategic business decision making. *World Journal of Advanced Research and Reviews*, *22*(3). https://doi.org/10.30574/wjarr.2024.21.2.0247

27. Informatica.com. (2024). *Bridge the Gap to Real-World AI with the Help of Data Integration*
28. International, D. (2017). *DAMA-DMBOK Data Management Body of Knowledge*. Technics Publications.
29. *Iqbal, R., Yuda, P., Aditya, W., Hidayanto, A. N., Handayani, P. W., & Harahap, N. C. (2019, 21-22 Sept. 2019). Master Data Management Maturity Assessment: Case Study of XYZ Company. 2019 2nd International Conference on Applied Information Technology and Innovation (ICAITI),
30. Jain, A., Aneja, S., Fuller, C., Dicker, A., Chung, C., Kim, E., Kirby, J., Quon, H., Lam, C., Louv, W., Ahern, C., Xiao, Y., McNutt, T., Housri, N., Ennis, R., Kang, J., Tang, Y., Higley, H., Berny-Lang, M., & Camphausen, K. (2020). Provider Engagement in Radiation Oncology Data Science: Workshop Report. *JCO Clinical Cancer Informatics*, *4*, 700-710. https://doi.org/10.1200/CCI.20.00051
31. *Kambli, R. (2019). Value Creation Through Effective Data Strategy. *ISACA*, *5*.
32. *Khaleghian, H., & Shan, Y. (2023). Developing a Data Quality Evaluation Framework for Sewer Inspection Data. *Water*, *15*(11).
33. *Khalid, H., & Zimányi, E. (2024). Repairing raw metadata for metadata management. *Information Systems*, *122*, 102344. https://doi.org/https://doi.org/10.1016/j.is.2024.102344
34. *Kotter, J. P. (1995). Leading Change - Why Transformation Efforts Fails *Harvard Business Review*.
35. Kraus, S., Breier, M., & Dasí-Rodríguez, S. (2020). The art of crafting a systematic literature review in entrepreneurship research. *International Entrepreneurship and Management Journal*. https://doi.org/https://doi.org/10.1007/s11365-020-00635-4
36. *Krause, R., & Becker, C. (2021, 2021//). Scenarios for the Use of Master Data Management in the Context of the Internet of Things (IoT). Innovation Through Information Systems, Cham.
37. *Król, K., & Zdonek, D. (2020). Analytics Maturity Models: An Overview. *Information*, *11*(3), 142. https://www.mdpi.com/2078-2489/11/3/142
38. *Li, Q., Lan, L., Zeng, N., You, L., Yin, J., Zhou, X., & Meng, Q. (2019). A Framework for Big Data Governance to Advance RHINs: A Case Study of China. *IEEE Access*, *7*, 50330-50338. https://doi.org/10.1109/ACCESS.2019.2910838
39. *Liu, P., Wang, X., Wen, F., Liu, Y., Sun, Z., Zhang, C., Yan, M., & Fan, L. (2018). Development and Application of Big Data Platform for "Bohai granary". *Wireless Personal Communications*, *103*(1), 275-293. https://doi.org/10.1007/s11277-018-5441-y
40. *Loshin, D. (2013). Chapter 9 - Metadata. In D. Loshin (Ed.), *Business Intelligence (Second Edition)* (pp. 119-130). Morgan Kaufmann. https://doi.org/https://doi.org/10.1016/B978-0-12-385889-4.00009-0
41. *Loukiala, A., Joutsenlahti, J.-P., Raatikainen, M., Mikkonen, T., & Lehtonen, T. (2021, 2021//). Migrating from a Centralized Data Warehouse to a Decentralized Data Platform Architecture. Product-Focused Software Process Improvement, Cham.
42. *Ma, X. (2023). The Development Phases of a Data Platform. In X. Ma (Ed.), *Methodology for Digital Transformation: Implementation Path and Data Platform* (pp. 241-259). Springer Nature Singapore. https://doi.org/10.1007/978-981-19-9111-0_16
43. *Mahanti, R. (2021). Data Governance Maturity Models. In R. Mahanti (Ed.), *Data Governance Success: Growing and Sustaining Data Governance* (pp. 99-126). Springer Singapore. https://doi.org/10.1007/978-981-16-5086-4_4
44. *March, S. (2003). Data Modeling - The Entity-Relationship Data Model. *Encyclopedia of Information Systems*. https://doi.org/10.1016/B0-12-227240-4/00034-4
45. *Miao, F., Yang, W., Xie, Y., & Fan, W. (2022, 2022//). Research on the Realization Path and Application of a Data Governance System Based on Data Architecture. Data Science, Singapore.
46. *Mitra, A., Gaur, S. S., & Giacosa, E. (2019). Combining organizational change management and organizational ambidexterity using data transformation. *Management Decision*, *57*(8), 2069-2091. https://doi.org/10.1108/MD-07-2018-0841
47. *Monah, S. R., Wagner, M. W., Biswas, A., Khalvati, F., Erdman, L. E., Amirabadi, A., Vidarsson, L., McCradden, M. D., & Ertl-Wagner, B. B. (2022). Data governance functions to support responsible data stewardship in pediatric radiology research studies using artificial intelligence. *Pediatric Radiology*, *52*(11), 2111-2119. https://doi.org/10.1007/s00247-022-05427-2
48. *Munhoz de Medeiros, M., Maçada, A. C., & Freitas Junior, J. C. (2020). The effect of data strategy on competitive advantage. *The Bottom Line*, *ahead-of-print*. https://doi.org/10.1108/BL-12-2019-0131
49. Okoli, C. (2015). A Guide to Conducting a Standalone Systematic Literature Review. *Communications of the Association for Information Systems*, *37*. https://doi.org/10.17705/1CAIS.03743
50. *Pachamanova, D., Tilson, V., & Dwyer-Matzky, K. (2022). Machine Learning, Ethics, and Change Management:A Data-Driven Approach to Improving Hospital ObservationUnit Operations. *INFORMS TRANSACTIONS ON EDUCATION*, *22*.
51. Pittaway, L., Holt, R., & Broad, J. (2016). Synthesising Knowledge in Entrepreneurship Research: The Role of Systematic Literature Reviews. *Handbook of Research on Small Business and Entrepreneurship*.
52. *Pratama, F. G., Astana, S., Yudhoatmojo, S. B., & Hidayanto, A. N. (2018, 1-2 Nov. 2018). Master Data Management Maturity Assessment: A Case Study of Organization in Ministry of Education and Culture. 2018 International Conference on Computer, Control, Informatics and its Applications (IC3INA),
53. *Quimbert, E., Jeffery, K., Martens, C., Martin, P., & Zhao, Z. (2020). Data Cataloguing. In Z. Zhao & M. Hellström (Eds.), *Towards Interoperable Research Infrastructures for Environmental and Earth Sciences: A Reference Model Guided Approach for Common Challenges* (pp. 140-161). Springer International Publishing. https://doi.org/10.1007/978-3-030-52829-4_8
54. *Rahman, A. A., Dharma, P. G., Fatchur, R. M., Freedrikson, A. N., Ari, B. P., & Ruldeviyani, Y.

(2019, 12-13 Oct. 2019). Master Data Management Maturity Assessment: A Case Study of A Pasar Rebo Public Hospital. 2019 International Conference on Advanced Computer Science and information Systems (ICACSIS),
55. *Raineri, A. (2011). Change management practices: Impact on perceived change results. *Journal of Business Research*, *64*, 266-272. https://doi.org/10.1016/j.jbusres.2009.11.011
56. *Ridzuan, F., & Zainon, W. M. N. W. (2024). A Review on Data Quality Dimensions for Big Data. Procedia Computer Science,
57. *Riedel, C., & Rass, S. (2019). An Overview of Data Quality Frameworks. *IEEE Access*, *PP*, 1-1. https://doi.org/10.1109/ACCESS.2019.2899751
58. Sarkis-Onofre, R., Catalá-López, F., Aromataris, E., & Lockwood, C. (2021). How to properly use the PRISMA Statement. *Systematic Reviews*, *10*(1), 117. https://doi.org/10.1186/s13643-021-01671-z
59. *Schwarz, D., List, M., & Mueller, R. (2023). *A Framework for the Systematic Evaluation of Data and Analytics Use Cases at an Early Stage*. https://doi.org/10.24251/HICSS.2023.658
60. *Sharma, R., & Djiaw, V. (2011). Realising the strategic impact of business intelligence tools. *VINE*, *41*, 113-131. https://doi.org/10.1108/03055721111134772
61. *Spruit, M., & Pietzka, K. (2015). MD3M: The master data management maturity model. *Computers in Human Behavior*, *51*, 1068-1076. https://doi.org/https://doi.org/10.1016/j.chb.2014.09.030
62. Storbjerg, S. H., Brunoe, T. D., & Nielsen, K. (2016). Towards an engineering change management maturity grid. *Journal of Engineering Design*, *27*(4-6), 361-389. https://doi.org/10.1080/09544828.2016.1150967
63. Tranfield, D., Denyer, D., & Smart, P. (2003). Towards a Methodology for Developing Evidence-Informed Management Knowledge by Means of Systematic Review. *British Journal of Management*, *14*(3), 207-222. https://doi.org/https://doi.org/10.1111/1467-8551.00375
64. *Trope, R. L., Power, E. M., Polley, V. I., & Morley, B. C. (2007). A coherent strategy for data security through data governance. *Ieee Security & Privacy*, *5*(3), 32-39. https://doi.org/10.1109/msp.2007.51
65. *van Helvoirt, S., & Weigand, H. (2015, 2015//). Operationalizing Data Governance via Multi-level Metadata Management. Open and Big Data Management and Innovation, Cham.
66. *Vilminko-Heikkinen, R., & Pekkola, S. (2019). Changes in roles, responsibilities and ownership in organizing master data management. *International Journal of Information Management*, *47*, 76-87. https://doi.org/https://doi.org/10.1016/j.ijinfomgt.2018.12.017
67. Willetts, M., Atkins, A., & Stanier, C. (2020). *Barriers to SMEs Adoption of Big Data Analytics for Competitive Advantage*. https://doi.org/10.1109/ICDS50568.2020.9268687
68. *Wolniak, & Grebski. (2023a). The concept of diagnostic analytics. *Scientific Papers of Silesian University of Technology Organization and Management Series*, *2023*. https://doi.org/10.29119/1641-3466.2023.175.41
69. *Wolniak, & Grebski. (2023b). The five stages of business analytics. *Scientific Papers of Silesian University of Technology. Organization and Management Series*, *2023*. https://doi.org/10.29119/1641-3466.2023.178.41
70. *Yee, M., Surkis, A., Lamb, I., & Contaxis, N. (2023). The NYU Data Catalog: a modular, flexible infrastructure for data discovery. *Journal of the American Medical Informatics Association : JAMIA*, *30*. https://doi.org/10.1093/jamia/ocad125
71. *Zhang, Q., Sun, X., & Zhang, M. (2022). Data Matters: A Strategic Action Framework for Data Governance. *Information & Management*, *59*(4), 103642. https://doi.org/https://doi.org/10.1016/j.im.2022.103642
72. *Zhang, Y., Zhang, C., & Xu, Y. (2021). Effect of data privacy and security investment on the value of big data firms. *Decision Support Systems*, *146*, 113543. https://doi.org/https://doi.org/10.1016/j.dss.2021.113543
73. *Zorrilla, M., & Yebenes, J. (2022). A reference framework for the implementation of data governance systems for industry 4.0. *Computer Standards & Interfaces*, *81*, 103595. https://doi.org/https://doi.org/10.1016/j.csi.2021.103595
74. *Zúñiga, D. V., Cruz, R. K., Ibañez, C. R., Dominguez, F., & Moguerza, J. M. (2018). *Master Data Management Maturity Model for the Microfinance Sector in Peru* Proceedings of the 2nd International Conference on Information System and Data Mining, Lakeland, FL, USA. https://doi.org/10.1145/3206098.3206127

APPENDIX A

Table A.1. Little Rocket expertise pillars

Pillar	Sub-pillars
1. Strategy	1.1 Governance
	1.2 Vision, Strategy Development
	1.3 Data quality
	1.4 Metadata Management
	1.5 Change Management
The strategy pillar focuses on developing comprehensive plans to manage and utilise data usage within an organisation effectively. It involves aligning data usage initiatives with business objectives, setting long-term goals and setting policies and procedures for data integrity. Data becomes a strategic asset for making informed decisions.	
2. Value	2.1 Analytics Translation, Business Consultancy
	2.2 Descriptive and Diagnostic Analysis
	2.3 Data Science, Forecasting, Statistics, NLP, ML, AI
	2.4 Business Intelligence
The value pillar involves transforming raw data into actionable information, improving the decision-making process and optimising operations. It does so by recognising use cases and developing dashboards that use historical data, analyse patterns, and (eventually) perform predictions using advanced technology.	
3. Fundament	3.1 Data Platform
	3.2 Data Architecture, Data Models
	3.3 Security, Privacy, Compliance
	3.4 Data Catalogue
	3.5 Master Data Management
The fundament pillar describes the technical base and infrastructure for effective data usage. Data operations rely on a stable, reliable fundament to store, process and manage data. This pillar also includes the security and compliance aspects of working with data.	
4. Knowledge	4.1 Basics and Mindset (Literacy)
	4.2 Leadership Program
	4.3 Ongoing Education
The knowledge pillar addresses the skill development needed to maintain a data-driven workforce. It includes the awareness of the possibilities of using data in an organisation, especially for the management team. Leadership has to be adjusted by this, and ongoing education for all employees assists in keeping up with the evolving data technologies.	
5. Team	5.1 Organisation
	5.2 CDO / Data Lead
	5.3 Data Engineers
	5.4 Data Scientists
	5.5 Data / Business Analysts
	5.6 Analytics Translators
	5.7 Data Stewards
The team pillar defines and supports the organisational roles and responsibilities an organisation must have for effective data management. It ensures the organisation has the required talent and clearly defined roles to foster a data-centric culture supporting its decision-making.	

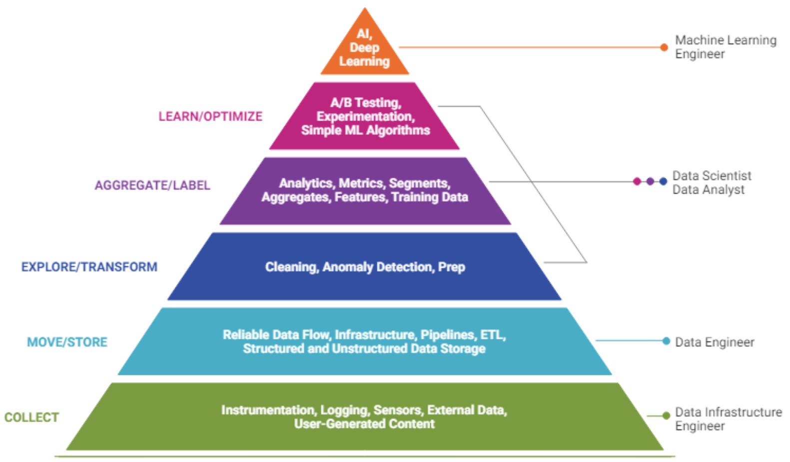

Figure A.3. Data science hierarchy of needs (Informatica.com, 2024, p. 7)

Table A.3. Data governance maturity model proposal (Mahanti, 2021, p. 123)

	Dimensions	Maturity level				
A	Data governance—overall	1	2	3	4	5
B	Data policies	1	2	3	4	5
C	Data standards	1	2	3	4	5
D	Data ownership	1	2	3	4	5
E	Data stewardship	1	2	3	4	5
F	Data compliance	1	2	3	4	5
G	Data architecture—data access and security	1	2	3	4	5
H	Data quality	1	2	3	4	5
I	Metadata	1	2	3	4	5

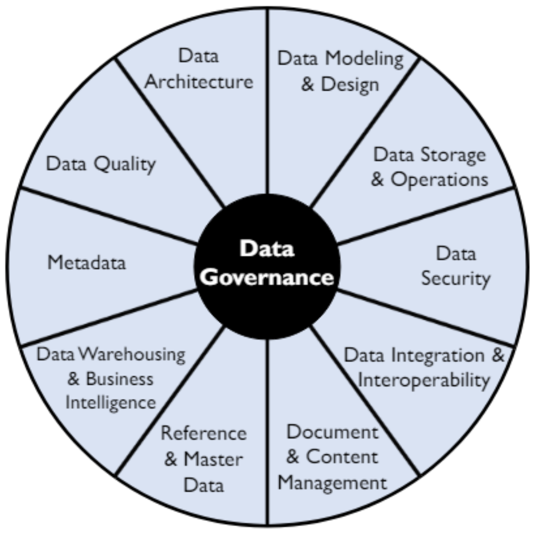

Figure A.4. DAMA-DMBOK2 data management framework (International, 2017, p. 67)

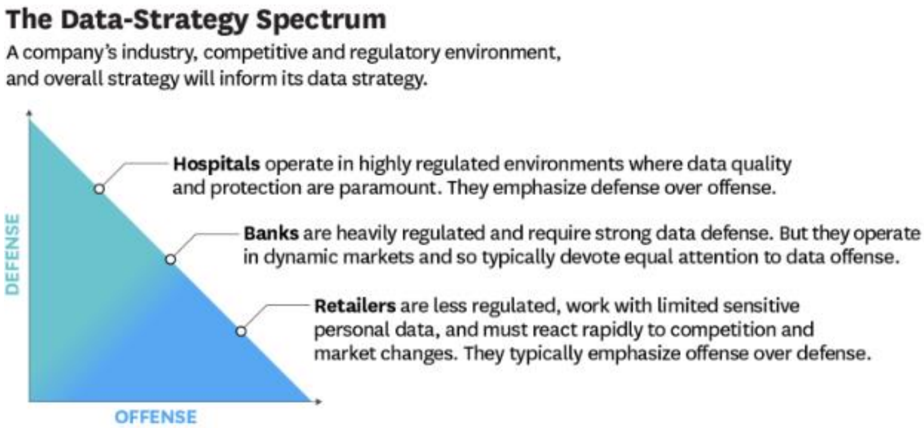

Figure A.5. Data-strategy spectrum (DalleMule & Davenport, 2017, p. 10)

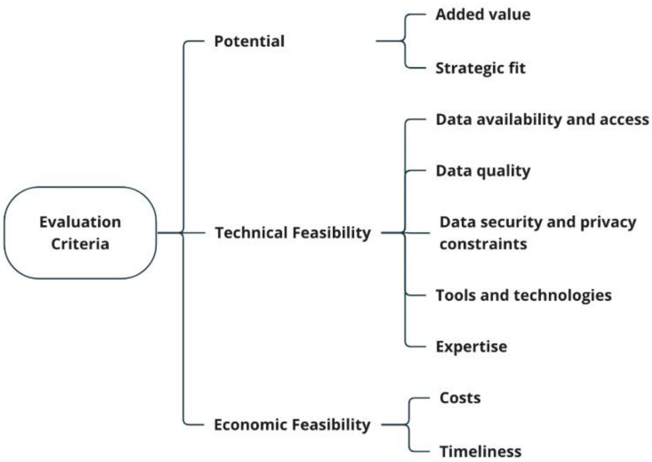

Figure A.6. Evaluation criteria for data and analytics use cases (Schwarz et al., 2023, p. 5)

Table A.4. Data analytics steps

	Data analytics steps					
Article	Descriptive	Real-time	Diagnostic	Predictive	Prescriptive	Cognitive
(Ibeh et al., 2024)	X		X	X	X	
(Król & Zdonek, 2020)	X		X	X	X	X
(Wolniak & Grebski, 2023)	X	X	X	X	X	
(Farrokhizadeh & Öztayşi, 2022)	X			X	X	

APPENDIX B

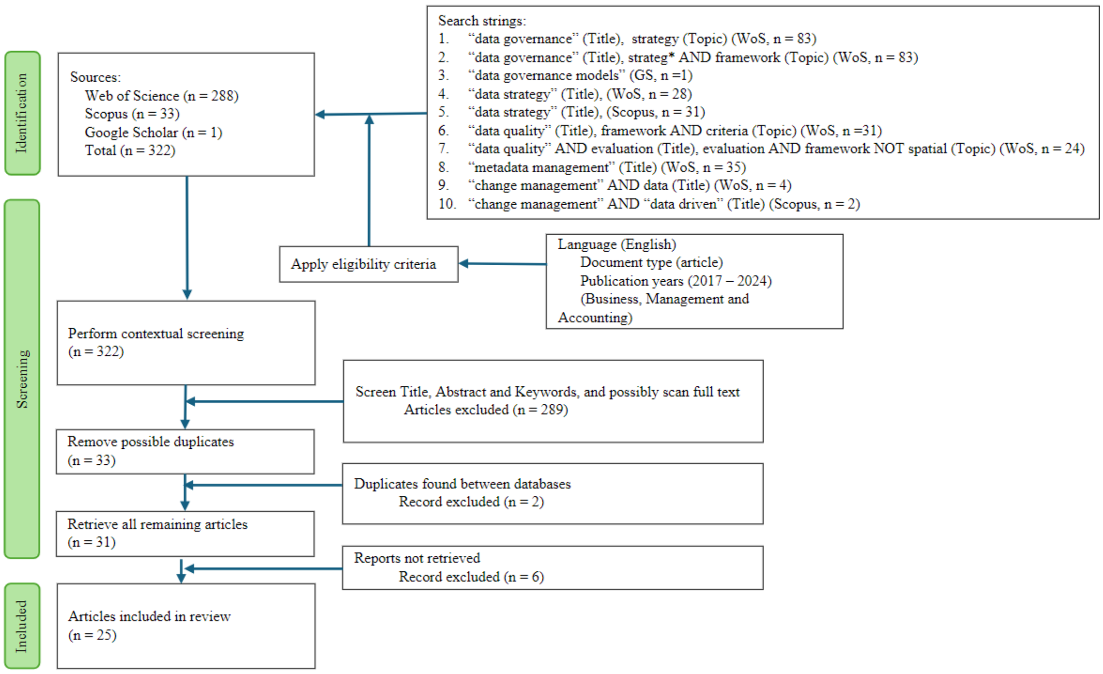

Figure B.7. Search flow Strategy pillar

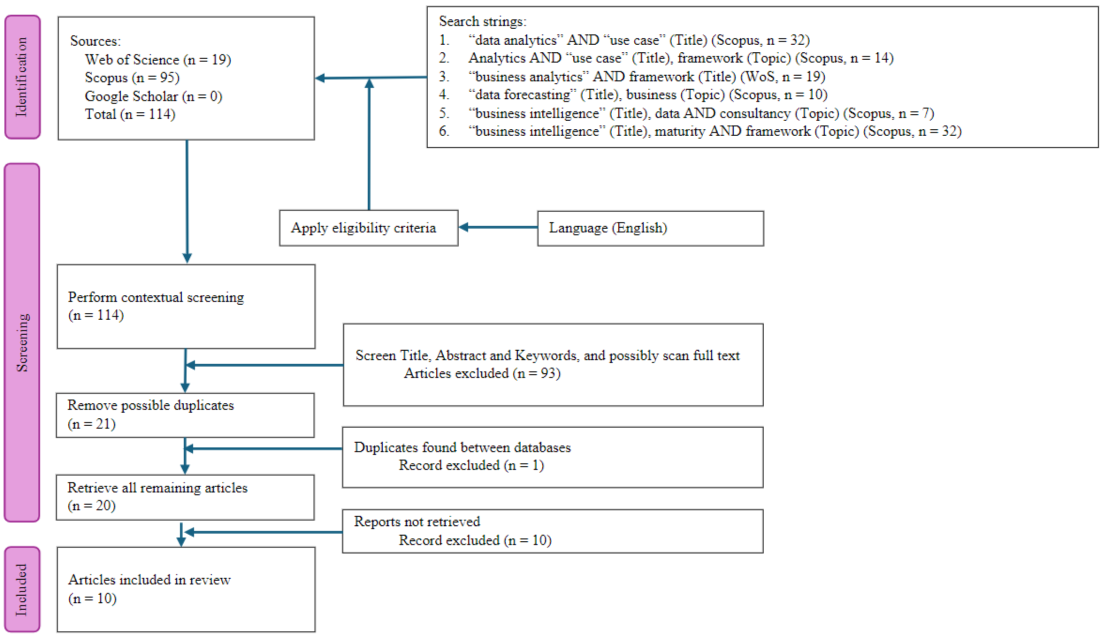

Figure B.8. Search flow Value pillar

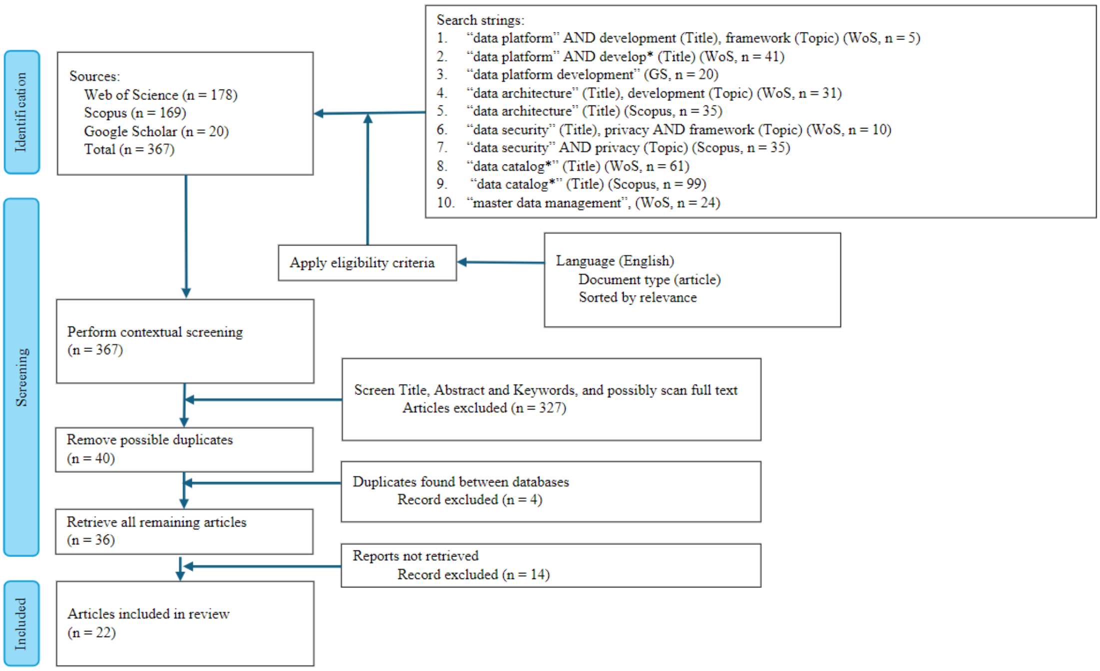

Figure B.9. Search flow Fundament pillar

Table B.5. Source x sub-pillar overview

	Strategy					Value				Fundament						
Source	1.1 Governance	1.2 Vision, Strategy Development	1.3 Data Quality	1.4 Metadata Management	1.5Change Management	2.1 Analytics Translation, Business Consultancy	2.2 Descriptive and Diagnostic Analysis	2.3 Data Science, Forecasting, Statistics, NLP, ML, AI	2.4 Business Intelligence	3.1 Data Platform	3.2 Data Architecture, Data Models	3.3 Security, Privacy, Compliance	3.4 Data Catalogue	3.5 Master Data Managment	**Knowledge**	**Team**
A Coherent Strategy for Data Security through Data Governance (Trope et al., 2007)	X											X				
Data Governance Taxonomy: Cloud versus Non-Cloud (Al-Ruithe et al., 2018)	X											X				
Operationalizing Data Governance via Multi-level Metadata management (van Helvoirt & Weigand, 2015)	X			X												
A Framework for Big Data Governance to Advance RHINs: A Case Study of China (Li et al., 2019)	X	X									X					
Data Matters: A Strategic Action Framework for Data Governance (Zhang et al., 2022)	X	X										X			X	
Data governance functions to support responsible data stewardship in pediatric radiology research studies using artificial intelligence (Monah et al., 2022)	X							X				X				X
A reference framework for the implementation of data governance systems for industry 4.0 (Zorrilla & Yebenes, 2022)	X	X		X				X			X	X				X
Data Governance Maturity Models (Mahanti, 2021)	X		X	X							X	X				X
Platform data strategy (Bhargava et al., 2020)		X														
What's Your Data Strategy? (DalleMule & Davenport, 2017)		X	X								X			X	X	X
The effect of data strategy on competitive advantage (Munhoz de Medeiros et al., 2020)		X														
Value Creation Through Effective Data Strategy (Kambli, 2019)	X	X						X	X			X				
An overview of data quality framework (Riedel & Rass, 2019)			X													
Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises (Günther et al., 2019)			X													
Developing a Data Quality Evaluation Framework for Sewer Inspection Data (Khaleghian & Shan, 2023)			X					X								
Data measurement in research information systems: metrics for the evaluation of data quality (Azeroual et al., 2018)			X													
A Review on Data Quality Dimensions for Big Data (Ridzuan & Zainon, 2024)			X													
Repairing raw metadata for metadata management (Khalid & Zimányi, 2024)				X												
Chapter 9 – Metadata (Loshin, 2013)				X					X							
Chapter 10 – Metadata Management (Allen & Cervo, 2015)	X			X							X					X
Combining organizational change management and organizational ambidexterity using data transformation (Mitra et al., 2019)		X			X										X	
Change Management Trend in the AI Modern World: Adapting to the Future of Work (Ademola & Professor, 2024)					X			X							X	X
Machine Learning, Ethics, and Change Management: A Data-Driven Approach to Improving Hospital Observation Unit Operations (Pachamanova et al., 2022)					X	X	X	X								X
Change management practices: Impact on perceived change results (Raineri, 2011)					X		X									
Leading Change - Why Transformation Efforts Fails (Kotter, 1995)					X											
A Framework for the Systematic Evaluation of Data and Analytics Use Cases at an Early Stage (Schwarz et al., 2023)		X	X			X			X			X			X	
Understanding the Role of Use Cases in UML: A Review and Research Agenda (Dobing & Parsons, 2000)						X										
Business analytics and decision science: A review of techniques in strategic business decision making (Ibeh et al., 2024)							X	X								
Analytics Maturity Models: An Overview (Król & Zdonek, 2020)							X	X								
The five stages of business analytics (Radosław Wolniak & Wies Grebski, 2023)							X	X								
Review of Descriptive Analytics Under Fuzziness (Farrokhizadeh & Öztayşi, 2022)			X				X	X								
THE CONCEPT OF DIAGNOSTIC ANALYTICS (Radosław Wolniak & Wes Grebski, 2023)			X				X					X				
The Impact of Data Science, Big Data, Forecasting, and Predictive Analytics on the Efficiency of Business System (Angelica & Mariluzia, 2022)								X								
Realising the strategic impact of business intelligence tools (Sharma & Djiaw, 2011)									X							
Business intelligence and cognitive loads: Proposition of a dashboard adoption model (Burnay et al, 2024))						X			X						X	
The Development Phases of a Data Platform (Ma, 2023)										X	X					
Development and Application of Big Data Platform for "Bohai granary" (Liu et al., 2018)								X		X	X	X				
Migrating from a Centralized Data Warehouse to a Decentralized Data Platform Architecture (Loukiala et al., 2021)										X						X
Research on the Realization Path and Application of a Data Governance System Based on Data Architecture (Miao et al., 2022)	X										X					
Digital transformation, data architecture, and legacy systems (Cao & Iansiti, 2022)							X	X			X				X	X
Data modelling – The Entity-Relationship Data Model (March, 2003)											X					

	1.1 Governance	1.2 Vision, Strategy Development	1.3 Data Quality	1.4 Metadata Management	1.5Change Management	2.1 Analytics Translation, Business Consultancy	2.2 Descriptive and Diagnostic Analysis	2.3 Data Science, Forecasting, Statistics, NLP, ML, AI	2.4 Business Intelligence	3.1 Data Platform	3.2 Data Architecture, Data Models	3.3 Security, Privacy, Compliance	3.4 Data Catalogue	3.5 Master Data Managment	**Knowledge**	**Team**
A Data Security Framework for Cloud Computing Services (Bautista-Villalpando & Abran, 2021)												X				
Data security and privacy computing in artificial intelligence (Feng et al., 2024)								X				X				
Effect of data privacy and security investment on the value of big data firms (Zhang et al., 2021)		X										X				
Data Catalogs: A Systematic Literature Review and Guidelines to Implementation (Ehrlinger et al., 2021)				X									X			X
Data Catalog: Approaches, Trends, and Future Directions (Boufassil et al., 2023)				X				X					X			
Data Cataloguing (Quimbert et al., 2020)				X									X			
The NYU Data Catalog: a modular, flexible infrastructure for data discovery (Yee et al., 2023)				X									X			X
Chapter 11 – Performance Measurement (Allen & Cervo, 2015)	X		X	X										X		X
Strategies for Master Data Management: A Case Study of an International Hearing Healthcare Company (Haug et al., 2023)	X													X		X
Scenarios for the Use of Master Data Management in the Context of the Internet of Things (IoT) (Krause & Becker, 2021)							X							X		
Changes in roles, responsibilities and ownership in organizing master data management (Vilminko-Heikkinen & Pekkola, 2019)	X													X		X
Master Data Management Maturity Assessment: Case Study of XYZ Company (Iqbal et al., 2019)														X		
Master Data Management Maturity Assessment: A Case Study of A Pasar Rebo Public Hospital (Rahman et al., 2019)														X		
Master Data Management Maturity Assessment: A Case Study of Organization in Ministry of Education and Culture (Pratama et al., 2018)														X		
MD3M: The master data management maturity model (Spruit & Pietzka, 2015)			X								X	X		X		X
Master Data Management Maturity Model for the Microfinance Sector in Peru (Zúñiga et al., 2018)	X		X				X					X		X		X
Number of articles found for a specific sub-pillar (X)	8	4	5	3	5	2	5	5	2	3	3	3	4	9	0	0
Number of times sub-pillar was considered important in an article	15	10	13	11	4	4	10	15	5	3	11	15	4	10	7	16

Table B.6. Sub-pillar x sub-pillar overview

	1.1 Governance	1.2 Vision, Strategy Development	1.3 Data quality	1.4 Metadata Management	1.5 Change Management	2.1 Analytics Translation, Business Consultancy	2.2 Descriptive and Diagnostic Analysis	2.3 Data Science, Forecasting, Statistics, NLP, ML, AI	2.4 Business Intelligence	3.1 Data Platform	3.2 Data Architecture, Data Models	3.3 Security, Privacy, Compliance	3.4 Data Catalogue	3.5 Master Data Management	Total sum of sub-pillars mentioned	Number of sub-pillars mentioned
Governance 1.1		3	1	3	0	0	0	2	0	0	3	6	0	0	18	6
Vision, Strategy Development 1.2	1		1	0	0	0	0	1	1	0	1	1	0	1	7	7
Data quality 1.3	0	0		0	0	0	0	1	0	0	0	0	0	0	1	1
Metadata Management 1.4	1	0	0		0	0	0	0	1	0	1	0	0	0	3	3
Change Management 1.5	0	1	0	0		1	2	2	0	0	0	0	0	0	6	4
Analytics Translation, Business Consultancy 2.1	0	1	1	0	0		0	0	1	0	0	1	0	0	4	4
Descriptive and Diagnostic Analysis 2.2	0	0	2	0	0	0		4	0	0	0	1	0	0	7	3
Data Science, Forecasting, Statistics, NLP, ML, AI 2.3	0	0	1	0	0	0	4		0	0	0	0	0	0	5	2
Business Intelligence 2.4	0	0	0	0	0	1	0	0		0	0	0	0	0	1	1
Data Platform 3.1	0	0	0	0	0	0	0	1	0		2	1	0	0	4	3
Data Architecture, Data Models 3.2	1	0	0	0	0	0	1	1	0	0		0	0	0	3	3
Security, Privacy, Compliance 3.3	0	1	0	0	0	0	0	1	0	0	0		0	0	2	2
Data Catalogue 3.4	0	0	0	4	0	0	0	1	0	0	0	0		0	5	2
Master Data Management 3.5	4	0	3	1	0	0	1	0	0	0	0	1	0		10	5
Total sum of sub-pillars mentioned	7	6	9	8	0	2	8	14	3	0	7	11	0	1		
Number of sub-pillars mentioned	4	4	6	3	0	2	4	9	3	0	4	6	0	1		

APPENDIX C

Table C.7. Content analysis

Article	Sub-pillar search	Most important findings
	1.1 Governance	
A Coherent Strategy for Data Security through Data Governance (Trope et al., 2007)		Merger occurs; InterCepht (aerospace and defence contractor) asked for an expert's view on emerging risks. Four trends that posed security risks: - Data stored on weakly protected portable devices - Deperimeterization undermines the reliability of perimeter-based defences - Decreasing the reliability of user identification and password protection - Introducing new technology Essential activities for the early stages of the merger: - Note that many laws contain security requirements but only express these implicit - Form a coherent security strategy - Prevent being deferred or unfocussed on due diligence Governance: has to oversee the security strategies Change management: the role of boards of directors now extends to actively managing the company's data.
Data Governance Taxonomy: Cloud versus Non-Cloud (Al-Ruithe et al., 2018)		The article focuses on cloud/non-cloud stored data, which could be implemented in maturity steps. Requirements for data governance: - A strategy framework that can be implemented → helps to create a mission, achieve clarity, maintain scope and focus as well as define measurable successes - Model of role responsibilities to identify people accountable for producing, defining and using data in the organisation The source provides several definitions of data governance. Most clear one: "*the exercise of authority, control and shared decision-making (planning, monitoring and enforcement) over the management of data assets.*" ➔ General consensus: "… among authors is that data governance refers to the entirety of responsibilities and decision rights concerning the management of data assets in organisations." Policies, standards, and operating processes are required to ensure data accuracy, availability, and security, which are part of the governance strategy.
Operationalizing Data Governance via Multi-level Metadata management		The data governance domain consists of 3 focus areas: - People - Process

(van Helvoirt & Weigand, 2015)		- Technology The goal of data governance → increasing data quality and trust. IBM's holistic approach to Big Data governance: 1. Define business problem 2. Obtain executive sponsorship 3. Align teams 4. Understand data risk and value 5. Implement analytical/operational projects 6. Measure results
A Framework for Big Data Governance to Advance RHINs: A Case Study of China (Li et al., 2019)		Healthcare big data governance practices in China have been researched, leading to a framework that includes three domains (Drive, capability and support) and twelve elements → goal: contributing to realising the business value of healthcare big data. Imagina big data governance as the ability for a human to run; - Drive domain = Can it run? - Support domain = How fast can it run - Capability domain = How far can it run? Drive domain - Strategy planning governance - Open transaction governance - Industry support governance - Governance of laws and regulation Capability domain - Healthcare big data organisation → IT architecture - Healthcare big data collection - Healthcare big data storage - Healthcare big data process and analysis - Healthcare big data usage Support domain - Healthcare big data resource planning - Healthcare big data standard system - Healthcare big data privacy security protection Note: the healthcare application of this case study might also be helpful for other sectors. The framework overlaps with other articles.
Data Matters: A Strategic Action Framework for Data Governance (Zhang et al., 2022)		Case study on a Chinese gold mining company, researching how firms deploy data governance, what strategic actions they take to do so and how they adapt to dramatic data increases. Organisations need to formulate a data governance strategy to cope with the significant increase in data or turn data into an asset. Defining the concept of data governance is difficult due to different viewpoints. In general, there are two streams: - Data governance is the exercise of decision-making authority and control over data to transform it into an asset → focus on the applicability of data and value data can have - Data governance ensures data availability by addressing data quality, privacy and security

		Four key data governance activities supported by two strategic actions - Concepts - Data governance falls under corporate governance and, therefore, includes the activities that create and execute this function - Includes the object of who holds rights/control over the data, who can use it and how - Mechanistically - structures that connect data to business, procedures, norms to deploy data governance - relationships supporting participating parties in developing governance collaborations - Consequences - having data governance around high-quality data makes firms able to adapt to business eruptions whilst developing new solutions to grow the business For traditional companies to organise data governance: - Start with data collaboration (collaboration within the firm, based on data, as well as collaboration with external actors, based on data) → besides opportunities to leverage external resources and strengths, it also leads to needed infrastructure for the firm to perform their work - Perceptions of data potential in the firm and the capabilities to perform these activities conclude how big the step is between data collaboration and data governance - Data-related awareness is essential: - if managers do not understand the importance of data governance, employees will also ignore the value of data - With growing data, firms must continuously scan which data is useful and which can be turned into company assets.
Data governance functions to support responsible data stewardship in pediatric radiology research studies using artificial intelligence (Monah et al., 2022)		Health sector, including AI development. Advanced data governance and security are required to safeguard private (patient) information. It is crucial to understand basic data governance principles to build capacity enabling AI technologies. ➔ Many AI initiatives get stalled due to a lack of data governance and thus concerns in data security →"especially where data is stored across different systems, departments and institutions and in various formats and standards." For patient information, a distributed or centralised network approach can be used. There is no better version. It is case-dependent. The rest of the article is too specific on radiologic research/advice.
A reference framework for the implementation of data governance systems for industry 4.0 (Zorrilla & Yebenes, 2022)		Implementation for data governance in Industry 4.0 is crucial → real-time generated data is used for insights and thus decision-making. It, therefore, must be managed and governed as a strategic asset. A data governance system must establish: - "Identifying what activities and specific aspects related to data should be governed." - Which roles are involved in decision-making (data stewards, owners, committees, etc.) - How these roles are related to the decision-making; who has authority, responsibility, etc General requirements for a data governance system: - Stakeholder needs are satisfied; value is generated from the data - Several components work together well - It's dynamic; in case of changes, it must be considered how much the system gets impacted - Structures and activities between data governance and Data Management must be distinguished - Adjusts to the needs of the organisation - It takes the whole organisation into account Specific (~~I4.0~~, not used (too specific)) requirements for a data governance system: Four groups to consider:

		1. Principles: requirements collected that have to be met by principles that govern the data governance system "Guiding conduct, behaviour and philosophy of the firm, regarding data usage, management and governance. Principles orient towards a data-centric (everything starts with data; it is the core of the business) architecture. All principles must be aligned with the general data governance's goals and objectives." 2. Governance: Strategic alignment and organisational requirements, Data Governance and stewardship requirements, data governance program goals, objectives, and strategies must be aligned with the firm's business plan. There must also be a monitoring system to check on this. Lastly, the firm's needs must be collected and available for decision-making. Clear roles and bodies must be defined as those that govern or are governed in data-related situations. Furthermore, a right organisational model must be defined to apply the data governance system. Also, policies and standards and the functions of evaluating, directing, and supervising have to be defined. 3. Management: includes classification and metadata requirements, data quality requirements, Security, privacy and data risk requirements, Data Life Cycle (DLC) requirements 4. Monitoring: requirements to monitor, evaluate and assess Set policies and processes as well as KPIs to check on the performance of the data usage; is the strategy lived up by and performed according to the internal policies?
Data Governance Maturity Models (Mahanti, 2021)		The article explains using/making maturity frameworks in general and analyses the governance frameworks of several companies (Kalido, DataFlux, Microsoft, Informatica, Oracle, IBM). The article precisely analyses those frameworks and proposes a combined version, including five maturity levels and nine dimensions. This is according to the Capability Maturity Model (CMM). Note: table 4.5 provides an added 'Metrics' column to provide written expected change (/improvements).
	1.2 Vision, Strategy Development	
Platform data strategy (Bhargava et al., 2020)		Article on platforms: these create value by allowing consumers and external producers to interact through their infrastructure Decisions to make related to data usage: - Type of data to collect (Individual vs Aggregate data) - How should the data be sorted (On-premise vs In the cloud) - How should the data be shared (use data' hoarding' (for in-house innovation) Vs data sharing (for external innovation by third parties)) - How should the data be accessed? (Through API's or not?) - How should the data be monetised (ad-supported Vs consumer payment models) Also, for NONPLATFORM businesses: - "… data strategy must align with the firm's overall competitive business strategy." Traditional (non-platform) firms have contractual relationships set up differently; traditional firms often interact with external partners, which often involves individual bilateral relationships and contracts. In contrast, platforms have thousands/millions of participants; contracts are automated and simple. Relationships are about data exchange. Strategic priorities for platforms: - Balance efficiency (operational) with strategic decisions - Competition and Strategy
What's Your Data Strategy?		The article describes a framework to build a robust data strategy applicable across industries and levels of maturity.

(DalleMule & Davenport, 2017)		Two key issues: Defence vs offence, Flexibility vs control. Defence: - Minimising the downside of risk Ensuring compliance with regulations (for example, rules about data privacy integrity of financial reports) Use analytics to limit/detect fraud Building systems that prevent theft Defence ensures the integrity of data flow (internally in the firm) by identifying, standardising and governing data sources of authority (e.g. customer, supplier information or sales data). Concentrates on legal, financial, compliance and IT concerns. Offence: - Supporting business objectives (increase of revenue, profitability, customer satisfaction) Includes activities that generate customer insights (for example, data analysis/modelling) and integrate disparate customer/market data to support management's decision-making through (for example) interactive dashboards. Mostly relevant for customer-focused functions of the business and real-time sales and marketing. BALANCE IS KEY for every company; how can that be achieved? - The trick is to divide funding and people (it does not have to be equally divided!) Strong industry regulations (finance, healthcare) push firms to defensive strategies, and strong competition for customers shifts if towards offensive strategies. Defence requires standardised data; Offensive requires flexible data Information about Data Architecture and Information Architecture useful for the sub-chapters of those specific sub-pillars in sub-chapters *4.1 Strategy Sub-Pillars* and *4.3 Fundament Sub-Pillars*
The effect of data strategy on competitive advantage (Munhoz de Medeiros et al., 2020)		The article analyses how data strategy can affect competitive advantage. Several supported hypotheses have been stated and discussed in a structured manner. - Source "What's your data structure?", used in this SLR already, is an often cited source. In conclusion, three theoretical contributions have been stated: - "Conceptualisation, operationalisation and validation of construct related to the data strategy" - By performing surveys and data analysis on these results, empirical evidence shows that developing defensive and offensive positionings in data strategy has a positive, direct relation and impact on competitive advantage. - Offensive strategy directly and positively impacts competitive advantage whilst completely mediating the relation between defensive strategy and competitive advantage.
Value Creation Through Effective Data Strategy (Kambli, 2019)		Figure 3 proposes 9 points a data strategy should consider. - Centralisation - Data Sharing - Real-time - Data structure - Automation - Security

		- AI - Regulatory requirements - Backup recovery These points also provide useful definitions and information for other sub-pillars than 'strategy.'
	1.3 Data quality	
An overview of data quality framework (Riedel & Rass, 2019)		The article compares several types of data quality frameworks that differ in definition, assessment and ways of improving data quality (focus on methodologies useful for a wide range of business environments). POSMAD approach supports decision-making: - Planning - Obtaining - Storing and sharing - Maintaining - Applying - Disposing of data Twelve frameworks of different sources are used. Three types of data (Raw data item, Component data item, Information product). Three types of data structure (Structured data, Semi-structured data, unstructured data). Five dimensions of data quality that can indicate the level of quality of data: - Completeness - Accuracy - Timeliness - Consistency - Accessibility
Data quality assessment for improved decision-making: a methodology for small and medium-sized enterprises (Günther et al., 2019)		The paper proposes a methodology that simplifies the way data quality is evaluated and improves the understandability of its results → mainly for usability for SMEs. Paper useful for extra information on the data quality dimensions.
Developing a Data Quality Evaluation Framework for Sewer Inspection Data (Khaleghian & Shan, 2023)		Data quality evaluation framework to point out reliability problems. Quality dimensions (Accuracy, Consistency, Completeness, Uniqueness, Validity) used. A very useful hierarchy of needs pyramid is shown → to be used/snowballed for sub-pillar *2.3 Data Science, Forecasting, Statistics, NLP, ML, AI*.
Data measurement in research information systems: metrics for the evaluation of data quality (Azeroual et al., 2018)		Research information systems (RIS) are an integral part of a university's IT landscape, yet many of such institutions have not implemented RIS's - Many information and data sources are interwoven, this can lead to the data sources harming the quality of the data in different RISs Research is about investigating how data quality can be investigated (in the context of RIS) and explaining various dimensions of data quality based on literature. Contains a clear definition of Data Quality

		- Explains how times change, as well as data requirements - Data reliability measurements have to be able to be changed as well over time Table 1 states the categories in data quality: - Accuracy - Relevancy - Representation - Accessibility
A Review of Data Quality Dimensions for Big Data (Ridzuan & Zainon, 2024)		Snowballed article on "Data measurement in RIS: metrics for the evaluation of data quality (Azeroual et al., 2018)" to check on the other (possible) data quality dimensions by recent (2024) research.
	1.4 Metadata management	
Repairing raw metadata for metadata management (Khalid & Zimányi, 2024)		The paper introduces MDPrep, a system that can explore the usability and applicability of data preparation techniques. Goal: improving metadata quality 3-step approach: - Detect and identify problematic metadata elements and structural issues - Employ a keyword-based approach to enhance metadata elements and a syntax-based approach to rectify metadata issues that occur structurally - To compare the outcomes to ensure improved readability and reusability of prepared metadata files Data preparation pre-processed operations are performed early in data processing pipelines. Advantages of data preparation: - Detecting errors - Uniform formatting of values - Normalising numeric values Data preparation tasks: - Data cleaning - Data transformation - Data standardisation - Data enrichment
Chapter 9 – Metadata (Loshin, 2013)		Helpful information on the origin and utility of metadata and how metadata management is of importance to BI (other sub-pillar). States all (8) different types of metadata and works these out.
Chapter 10 – Metadata Management (Allen & Cervo, 2015a)		A book chapter that addresses metadata management in an MDM model - It discusses an approach to define, identify and manage enterprise-level metadata (including data models, dictionaries, reference data and more) - States how well-organised metadata contributes to the efficiency and success of data governance/analysis - Metadata management in Multi-Domain MDM context - Metadata example in a typical company

		- Internal and external sources - Data models → useful for other sub-pillar - Interfaces and Transformations - Analytics and Reporting - Correspondence and Legal Documents - Business Definitions and Processes - Business and Data-Quality Rules - Business Applications' User Interfaces - Metadata Management Goals → most useful
	1.5 Change management	
Combining organisational change management and organisational ambidexterity using data transformation (Mitra et al., 2019)		Combination of organisational change and organisation ambidexterity by using data transformation. Organisations have to engage with existing data because continuously: Therefore, external factors often trigger change and are discontinuous; being reactive is required! →"The outcome must be driven toward preparing for the change through data engagement, implementation and reinforcement." Requirements to succeed: - Have a strategy - Set up the correct operating model - Have a clear change management work-stream scope - Continue monitoring the progress by using defined milestones and (acceptance) criteria Several (traditional) models for performing change management: - Lewin's Force Field Analysis, a.k.a. 3-step model - Beckhard's change plan - Kotter's 8-step model - Continuous Change Process Model In literature, change is conceptualised in 2 fundamental ways: 1. As a rational strategic process, where the organisation chooses a new course of action and adapts to the change 2. Change is an evolutionary process where organisations (typically) resist the change happening around them In real life, however, organisations often adapt via strategic processes or fail to see the need for change.
Change Management Trend in the AI Modern World: Adapting to the Future of Work (Ademola, 2024)		The article focuses on the trends in change management in the context of AI, noting difficulties and prospects for organisations that embrace/integrate AI in future work. - Challenges and opportunities Overview Change management: traditional view and evolution in the context of AI - Traditional models include Kotter. Key factors on how to integrate AI in change management: - Upskilling/reskilling workforce

		- Fostering a culture of adaptability and of continuous learning - Maintain clear communication lines with employees; what are their concerns/expectations during AI implementation Note: organisations are obligated to address the ethical implications of AI; ensuring transparency and accountability in the deployment of AI is required. Three current trends in change management: - Adoption of agile change management methodologies - Embrace a culture of adaptability and of continuous learning - Integration of AI in change management processes Adapting to the future of work →helpful text for pillar Team and Knowledge → write a recommendation.
Machine Learning, Ethics, and Change Management: A Data-driven Approach to Improving Hospital Observation Unit Operations (Pachamanova et al., 2022)		Data analytics lifecycle: - Discovery - Data preparation - Model planning - Model building - Communication of results - Operationalisation "There is a dearth of cases that can be used to teach how to effect organisational change to implement data-driven recommendations in analytics, data science, and operations management courses" → potential future research? Kotter's 8-step model is used. Table 2 includes an overview of (typical) team (→ pillar) roles, including a description.
Change management practices: Impact on perceived change results (Raineri, 2011)		Includes some pretty different models and theories on change management and how they suggest performing the change. Most findings have already been discussed in other sources. Confirmed: Kotter's 8-step model is important again → clear pattern between sources for this sub-pillar.
Leading Change - Why Transformation Efforts Fails (Kotter, 1995)		Very famous 8-step model by Kotter to implement in Change Management.
	2.1 Analytics Translation, Business Consultancy	
A Framework for the Systematic Evaluation of Data and Analytics Use Cases at an Early Stage (Schwarz et al., 2023)		The article provides an evaluation framework that can be used for systematic screening (at an early stage) using nine criteria. It states that previous research on this exact field was missing, motivating them to do it. Often, use cases at companies fail or are not even developed at all. A lot of the time, a good evaluation process that has a systematic and structured approach.

		Nine criteria and some explanation: - Added value The value of a use case is distinguished by four dimensions: value creation, value capture, value proposition or value network. An estimated value is used for this criterion. - Strategic fit Evaluate if the use case fits with a firm's philosophy, image, objectives and management interests. - Data availability and access Data and sources have to be analysed first. Data availability refers to the degree to which it is instantly possible to access data. - Data quality Evaluated by the dimensions of completeness, understandability, timeliness, validity or integrity - Data security and privacy constraint - Tools and technologies Perform gap analysis - Expertise Perform gap analysis - Costs Dependent on technical feasibility criteria - Timeliness Dependent on technical feasibility criteria ➔ Use Figure 1 to visualise these evaluation criteria in the report.
Understanding the Role of Use Cases in UML: A Review and Research Agenda (Dobing & Parsons, 2000)		Source is briefly used to define a use case.
	2.3 Data Analytics	
Business analytics and decision science: A review of techniques in strategic business decision making (Ibeh et al., 2024)		Part of the table with Data Analytics Steps. The article combines business analytics and decision science. Includes AI, ML, NLP. Uses the steps descriptive, diagnostic, predictive, and prescriptive for defining data analytics. States decision science techniques: decision analysis, risk analysis, cost-benefit analysis, optimisation modelling, simulation modelling.
Analytics Maturity Models: An Overview (Król & Zdonek, 2020)		Part of the table with Data Analytics Steps. Uses the steps descriptive, diagnostic, predictive, prescriptive, and cognitive, for defining data analytics. Analyses analytics models of 11 different sources. The three most important factors in a firm that uses analytics are human resources, infrastructure, and proper organisation.
The five stages of business analytics (Wolniak & Grebski, 2023b)		Part of the table with Data Analytics Steps

		The goal of Analytics: Help identify growth and improvement opportunities, optimise operational efficiency and resource allocation, understand customer behaviour and preferences, mitigate risks, identify (potential) threats, and monitor and evaluate the performance of strategies. Uses the steps descriptive, real-time, diagnostic, predictive, and prescriptive, for defining data analytics - It also states the evolution and comparison of these steps.
Review of Descriptive Analytics Under Fuzziness (Farrokhizadeh & Öztayşi, 2022)		Part of the table with Data Analytics Steps. Uses the steps descriptive, predictive, and prescriptive for defining data analytics. Focuses on the first step of descriptive analysis, especially on the fuzziness (vague nature) of data.
THE CONCEPT OF DIAGNOSTIC ANALYTICS (Wolniak & Grebski, 2023a)		The paper focuses on diagnostic analysis, the gains it can bring but also the challenges: - Time-consuming - Real-time insights may be limited - Data quality and availability are uncertain sometimes. This impacts the reliability of the analysis. - Complexity of the analysis - Concerns about data privacy and security - Preventing bias and misinterpretation - Having difficulties in identifying causal relationships
The Impact of Data Science, Big Data, Forecasting, and Predictive Analytics on the Efficiency of Business System (Angelica & Mariluzia, 2022)		The paper presents the latest methods and technologies in Big Data analytics to increase productivity and efficiency in organisations. - Focuses on forecasting in predictive analysis. Used to estimate values of indicators by managing key performance indicators (by using specific algorithms). Research conducted in the field of financial-banking.
	2.4 Business Intelligence	
Realising the strategic impact of business intelligence tools (Sharma & Djiaw, 2011)		The balanced scoreboard approach led to the finding that strategic performance management required BI to be sound. It helps firms identify strengths and gaps concerning their environment. Some BI outcomes: - Better productivity - Increased market share - New product development improvements - Reduced costs and increased profits - Better customer handling
Business intelligence and cognitive loads: Proposition of a dashboard adoption model		Data-driven decision support systems (data-driven DSS), which are BI, improve the ability to objectify decisions with facts. BI dashboard designs are considered important as they are the visible tip of the BI data-driven DSS iceberg and the entire system's adoption level. In this paper is, dashboard content investigated, and to what extent it affects the adoption or rejection of usage of BI data-driven DSS.

(Burnay et al., 2024)		Dashboards are interactive interfaces that visualise and analyse a firm's performance metrics. They include indicators, graphs and tables in combination with interactive features to provide decision-makers with a consistent yet flexible representation of the firm. - If the dashboards are not correct yet used for decision making, good ideas may be declined due to 'misinformation.'
	3.1 Data platform	
The Development Phases of a Data Platform (Ma, 2023)		Data platform strategy is NOT a pile of 'pure technical concepts'; "It is the accumulation of core capabilities, data, and user information of the companies with the shared services approach, avoiding duplication of construction in every business department and reducing the costs associated with new businesses such that most of the business requirements are fulfilled by the business team." Data platforms can respond quickly to customer requirements whilst avoiding burdensome procedures of describing business requirements in the business department → saves communication costs and improves the business precision level of facilitating the business and technical departments in their collaboration of developing DT applications. Multi-dimensional ways to interpret data platforms: - The data platform is a new transformational architecture - The data platform is a sharing platform for capabilities - The data platform is an organic, integrated platform - The data platform is a new type of concept in the construction of technologies Nine basic capabilities of a data platform: 1. Data service capability 2. Data application development capability 3. Data processing capability 4. Data development capability 5. Self-learning and automated improvement capability 6. Assets accumulation capability 7. Data quality auto-tracking capability 8. Data integration capability 9. IT system & DT system risk isolation capability Three types of application: - Help the business department to perform data analysis flexibly - Assist business, technical and external departments to create applications flexibly - Allows the technical department to construct the capabilities of the application constantly and to accumulate data assets and their values Corporate misconceptions on building a data platform can result in construction risk → avoid this at all costs, as errors in construction can lead to derivative problems of applications produced by the data platform.' Common failures that occur when constructing a data platform are mentioned.

Development and Application of Big Data Platform for "Bohai granary" (Liu et al., 2018)		Introduction on Big Data; already provided at the sub-pillar governance (also an application for the agricultural industry) To provide higher performance (than traditional data analysis techniques) and better reliability environment (for large data analysis and processing), a large data platform architecture is needed → support horizontal growth. Definition of data mining → useful is sub-pillar 'Data Science, Forecasting, Statistics, NLP, ML, AI Big data platform construction → focus has to lie on key technologies such as big data architecture, big data modelling and storage large data analysis and processing, and big data applications). - All these technologies are explained in further subchapters. Further explanation about the data platform made especially for the "Bohai granary project"; too specific to use
Migrating from a Centralized Data Warehouse to a Decentralized Data Platform Architecture (Loukiala et al., 2021)		The paper states that centralised solution technologies (including data warehouses and data lakes) do not always scale well. There might be (competitive) opportunities/advantages in decentralised. Empirical evidence shows the potential of a decentralised data platform architecture whilst indicating core pieces that need central management. Key concepts of data warehousing, data lakes, data platform architecture, service-oriented architecture and microservices are further discussed.
	3.2 Data Architecture, Data models	
Research on the Realization Path and Application of a Data Governance System Based on Data Architecture (Miao et al., 2022)		Nowadays, the broad concept of data governance goes beyond the traditional scope → involves at least "establishment of data asset status, management system and mechanism, sharing and openness, security and privacy protection." Data registration system is the core composition of data architecture; public key encryption and authentication system are key components of the system. A data governance system based on a data architecture supports more complex, collaborative, cross-domain business application scenarios. Data architecture definition: a basic/simple and universal architecture that performs data rights confirmation, data management, data sharing and data security Note: before sharing data, is must first be managed Data rights confirmation and data security have similar interests/goals and are considered together in the rest of the paper Useful for the introduction: Chapter 4.1 - States how society is transitioning from a material civilisation to a data civilisation → Figure 1 - Table 2 gives clear difference between information and data Chapter 4.3 - relationship between data and applications
Digital transformation, data architecture, and legacy systems		The paper identifies architecture as key component of "intangible capital that drives technical co-invention from data and machine learning." Surveys measure the architecture of large, traditional corporations in the manufacturing, retail trade, finance and healthcare sectors.

(Cao & Iansiti, 2022)		Findings: - Costs and frictions of data related co-invention are reduced by coherent data architectures → easier for data team to make predictive models Moving from the bottom to the top quartile of participating companies, higher data architecture coherence leads to more intensive ML capabilities or ~14% more use cases across product development and business processes as customer relations. Legal hardware components negatively affect data architecture coherence when running complex software → result: legacy servers lower ML capabilities of complex corporations two times more than for simple corporations. - Among complex corporations, 6% more of ML capabilities are lost when using third-party server maintenance. Note: data coherence includes the architecture's capabilities around processing data streams, which consist of several layers and pipelines, combining data sets through several sources and locations, as well as developing and deploying ML models at scale. The survey divides data architecture into three aspects: coherence, security, and cloud computing.
Data modelling – The Entity-Relationship Data Model (March, 2003)		A useful source for defining data modelling - What does it include, what is its function? Combine and compare the definition with different, more recent sources. Have there been any developments in this definition over the years?
	3.3 Security, Privacy, Compliance	
A Data Security Framework for Cloud Computing Services (Bautista-Villalpando & Abran, 2021)		Specified article for cloud computing services. The introduction includes numbers on the number of cyberattacks; these attacks are very difficult to counter by organisations because they often fundamentally rely on insecure networks and new technologies. Often, these attacks happen on cloud computing services; the paper includes a framework for protecting cloud computing. The life cycle of the security framework is first to identify the data security requirements in cloud computing services, then manage these risks, and then evaluate the performance of this data security. This cycle continues due to the possibility of changing security requirements.
Data security and privacy computing in artificial intelligence (Feng et al., 2024)		Short guest editorial with very up-to-date and useful information about defining data security, specifically in the usage of AI - Useful definitions for both security and AI
Effect of data privacy and security investment on the value of big data firms (Zhang et al., 2021)		The paper analyses the usage of Big Data Analytics (BDA) and related technologies to raise concerns about data privacy and security (DPS) - It compares firms that have or have not made BDA investments.

		The main findings include that, on average, DPS investments significantly reduce the firm's systematic risk. The larger the BDA usage in a firm, the smaller the risk reduction. - Systematic risk includes, for example, the firm's exposure to macro-level economic risks.
	3.4 Data Catalogue	
Data Catalogs: A Systematic Literature Review and Guidelines to Implementation (Ehrlinger et al., 2021)		The paper researches the necessary and optional conceptual components for data catalogues and provides guidelines for implementing these. Some findings in general: - It is stated that there is little research done on data catalogues - Data catalogues "collect, create and maintain metadata" → so cannot go without it! - The target group who use data catalogues are primarily business users (and not IT/data specialists) - Data catalogues have to be adequately managed; the success depends on the people maintaining it A metadata schema is first required to implement a data catalogue. There are eight steps, of which the first five steps refer to the metadata schema: 1. The firm defines data context variables, containing data-system relationships, business context, technical context 2. Define data attributes, representing data quality, accessibility, sensitivity, reliability 3. Tag the data; decide which metadata is attached to certain data at a specific level (e.g. column level, entity level, data set level) 4. Define rules that regulate the accessibility of data or audit 5. Combine the previous steps into one enterprise data model, and this is the final data catalogue schema 6. Populate the catalogue with data 7. Expose the populated catalogue to users 8. Improve the catalogue through feedback, revisions and reviews
Data Catalog: Approaches, Trends, and Future Directions (Boufassil et al., 2023)		Briefly used for definitions, further access to the article is denied.
Data Cataloguing (Quimbert et al., 2020)		Book chapter Describes the definition, connection to metadata (and definition of metadata). Describes differences between metadata standards and schemas. Examples of using data catalogue tools; too specific to use.
The NYU Data Catalog: a modular, flexible infrastructure for data discovery (Yee et al., 2023)		A specific case of New York University using a data catalogue that has been improving since 2015 - Tailored metadata schema Again, the importance of the team managing the data sets is mentioned, stated in the discussion.
	3.5 Master Data Management	

Chapter 11 – Performance Measurement (Allen & Cervo, 2015b)		Performance measurement model, including metrics to measure MDM activities - Connections across data governance, quality, MDM, reference data management and improvement of processes From three perspectives: strategic, tactical, operational MDM discipline areas: - Data Governance - Data Stewardship - Data Integration - Data Quality Management - Metadata Management ➔ All disciplines are explained further
Strategies for Master Data Management: A Case Study of an International Hearing Healthcare Company (Haug et al., 2023)		Analysis of applying centralised and decentralised MDM approaches, resulting in four data management strategies. Six roles in MDM projects →implemented in the discussion. Governance plays a central role in MDM. Considering historical data and master data is of great importance to IoT (especially for long-term analysis).
Scenarios for the Use of Master Data Management in the Context of the Internet of Things (IoT) (Krause & Becker, 2021)		Provides a definition for master data and MDM; good to introduce the topic.
Changes in roles, responsibilities and ownership in organising master data management (Vilminko-Heikkinen & Pekkola, 2019)		Useful definition of MDM. Showcases how MDM is becoming more of an organisational and social issue than a technical issue → strongly relates to the team pillar. Implement this in the discussion. Relates data ownership and data governance roles and responsibilities to MDM.
Master Data Management Maturity Assessment: Case Study of XYZ Company (Iqbal et al., 2019)		Briefly used source; useful for Master Data Management Maturity Model (MD3M) citing. The article researches to what extent Company XYZ is mature in the usage of MDM.
Master Data Management Maturity Assessment: A Case Study of A Pasar Rebo Public Hospital (Rahman et al., 2019)		Briefly used source; useful for Master Data Management Maturity Model (MD3M) citing. Article researches on master data management in a hospital for managing patient data.
Master Data Management Maturity Assessment: A Case Study of Organization in Ministry of Education and Culture		Briefly used source; useful for Master Data Management Maturity Model (MD3M) citing.

(Pratama et al., 2018)		
MD3M: The master data management maturity model (Spruit & Pietzka, 2015)		Snowballed source. This article defines the Master Data Management Maturity Model (MD3M) to which earlier used sources often refer. Five key topics and their 13 focus areas: - Data model (including 1) definition of master data, 2) master data model and 3) data landscape) - Data quality (including 1) assessing data quality, 2) impact on business, 3) awareness of quality gaps and 4) improvements) - Usage and Ownership (including 1) data usage, 2) data ownership and 3) data access) - Data protection (including data protection) - Maintenance (including 1) storage and 2) data lifecycle) Five maturity levels.
Master Data Management Maturity Model for the Microfinance Sector in Peru (Zúñiga et al., 2018)		Reflecting on different maturity models for MDM, I made my own version. MD3M model, together with other MDM maturity frameworks (by Oracle, IMN and Data Flux), have been analysed → a different matrix, made for the microfinance sector, has been proposed. The lesson of this article is that it (most probably) is the case that different sectors require different maturity models.