Strategic Data Acquisition – a literature review

Seminar paper

Gorny, Christof, FH Wedel, Wedel, Germany, bwl105408@fh-wedel.de

Abstract

In the age of digital transformation, the topic of data driven decision making became increasingly more relevant. Empirics show that companies who use DDDM as part of both daily business operations and strategic decisions have an overall better financial performance. For companies to base their decisions on data, this data first needs to be acquired in the correct manner and cost efficiently. Although the topic on DDDM and data strategies have been sufficiently researched, the actual strategic data acquisition has very little holistic research papers. This review therefore summarizes the state-of-the-art literature on the components of the data collection process to give an overview on the topic as a whole. It covers the definitions, data sources and data usage possibilities. The review shows that there are a variety of sources and usage possibilities that are cost intensive and complex, so companies should carefully consider which goals they want to achieve with the data.

Keywords: Strategic data acquisition, data collection, data strategy, data driven decision making,

Table of Contents

1 Introduction

Data driven decision making (DDDM) is a topic that increasingly receives attention. Companies like Airbnb, Uber and Netflix have already centred their business model around acquiring, analyzing and basing strategic decisions on data (Sorescu 2017). The significance of these business models arises from the high market volatility as a consequence of the digital revolution. Companies constantly have to adjust to varying market conditions to stay competitive, especially during the current global COVID-19 pandemic (McKinsey 2021).

To be able to use DDDM as part of daily business operations, companies first need to implement a suitable data strategy. McKinsey (2013) describes three key aspects to build a data driven strategy, the first being to choose the right data to be acquired. Managers can get creative with external sources like social media conversations or focus on more traditional approaches, e.g. customer service data they already possess. Nonetheless, if the acquired data does not fit the criteria to give further insights into the business, DDDM becomes difficult. Therefore, a critical part of a data driven strategy and a digital business model lies in the first contact point, the data acquisition (Wilberg et al. 2017).

Despite the output of scientific research on data strategy and DDDM being quite rich, there is almost no state-of-the-art reference work that discusses different approaches on strategic data acquisition / collection in detail. To close this research gap, this literature review focusses on summarizing the main research findings in the literature on topics around data strategy and DDDM to gain further insights on how data acquisition should be strategically directed and how companies can use data long-term to build better products. In detail, I aim to synthesize a definition of the term strategic data acquisition, give an overview of common data sources as well as presenting how the big players in Silicon Valley adopted different usage possibilities to maximize business value through data.

2 Theoretical Background on Data Strategy

After analyzing the literature on data strategy briefly, it can be said that currently there is no general and widely accepted definition on the term. Many versions to the definition exist, with both scientific researchers and companies setting different focal points. Gür and Spiekermann (2020) from the Fraunhofer institute for Software and Systems Engineering therefore consolidated 9 different approaches to the definition and identified their similarities.

In conclusion, they summarized 6 key elements to a data strategy: (1) Clear vision, mission and business objective alignment, (2) long term benefits and competitive advantage, (3) constitution of a roadmap and objectives, (4) organizational and technological assessment and change management, (5) long-term and organization-wide data strategy establishment, (6) sets boundaries and objectives for data management (Gür and Spiekermann 2020). A data strategy can also be defined as organizing, governing, analyzing and deploying the generated data of an enterprise (DalleMule and Davenport 2017).

Another widely used term is „big data strategy“. Big data is usually characterized by 3 traits: volume (enormous quantity of data), velocity (real-time creation) and variety (structured, semi-structured or unstructured). Some further research has added more traits such as exhaustivity, another has defined big data as mainly being characterized simply by velocity and exhaustivity (Kitching & McArdle 2016). In this literature review, data strategy and big data strategy shall be treated as indifferent.

The relevance of data strategies become visible when we look at the different values they generate. An analysis conducted by Himmi et al. (2017) found out, that the usage of big data can both generate value in operative as well as strategic environments. In operative environments, data can be used to decrease uncertainty by automation and prediction (Himmi et al. 2017). One possible manifestation is the usage of predictive analytics (Waller and Fawcett 2013) or predictive maintenance (Nentwich and Reinhart 2021). In strategic environments, big data can be used to support decision making (Sorescu 2017). This is also known as data driven decision making. A survey from 2011 which included 179 large firms concluded that there is a strong connection between the usage of DDDM and an overall higher productivity and market value (Brynjolfsson et al. 2011). Another survey published by John Hurley (2018) in the Harvard Business Review identified a significant connection between the state of a company’s data strategy and its market performance. Based on these results, it can be seen that managing and deploying the collected data as part of a data strategy becomes more critical for success. Another article published by the Boston Consulting Group (2020) indicates that data is now not only creating competitive advantage, but rather has become the backbone of a company. Other Sources also imply that investing in data is now more valuable than ever (IBM 2020).

3 Methodical Approach

This literature review was conducted roughly following the method of Webster and Watson (2002). The approach consisted of 4 steps, which will be explained below.

Step 1: General information gathering

At first, I did a general research to gain an overview of the topic. Therefore, I performed basic google and google scholar search to understand how strategic data acquisition is received in general media. The keywords I used were “strategic data acquisition”, “data acquisition strategy” and “data strategy”. The search terms “data collection” or “strategic data collection” were not used, because it is a widely used term for every kind of data related topic, including the methodology of almost every quantitative research, therefore generating way too many results with no relation to the topic.

Blog posts by specialists and companies gave first insights and provided general information and scientific papers as well as conference papers were examined to get an impression of the current state of research. This first search resulted in the following key papers, which built the foundation of the following detailed database search:

Search word	Title	Author
“Strategic data acquisition”	Strategic data acquisition	Bohnhoff, Tobias
“Data acquisition strategy”	10 data acquisition strategies for start-ups	Mueller-Freitag, Moritz
“Data strategy”	Data Strategy Praxis Report	Gür & Spiekermann

Table 1. List of the key papers identified to each search word

What immediately became clear is that there are almost no results with the specific search term “strategic data acquisition”, especially when searching for peer reviewed papers in scientific journals. The few results of the articles are blog posts on different, primarily tech-based websites as well as self-published best practices from management consulting firms. The search term “data acquisition strategy” generated more results, but also close to no peer reviewed papers. I therefore started with an overview on the topic via blog posts. The “Data Strategy Praxis Report” from Gür & Spiekermann (2020) is a technical analysis with a focus on tools and approaches in the current data economy. Although the report does not directly revolve around the data acquisition itself, it provides a good overview on the topic as a whole, with acquisition/collection being a part of a holistic data strategy.

Step 2: Detailed and precise searching in databases

Due to the lack of specific literature on strategic data acquisition, a definition had to be synthesized first to be able to conduct further research on the key elements. In this case, the search terms had to be split in order to proceed.

At first, a definition for strategy had to be found. As it is a widely researched field with several accredited and supported definitions, a simple google search instantly generated sufficient results. The most accredited definitions were then used.

Following that, a specific database search with the search terms “data acquisition” and “data collection” was conducted. The libraries used were ScienceDirect, AiSel, Google Scholar as well as basic Google.

As data acquisition and data collection is also a common research field in natural sciences, the results were limited to business, economics and information technology based papers and articles. To finalize the definition, the blog posts and IT websites on strategic data acquisition and data acquisition strategies were amended.

The key statements of the papers were then searched in the databases. They were used for the fields title, abstract, keywords and in the full text.

Step 3: Forward and backwards search

As IT based websites and blog posts usually base their output on own experiences, they rarely cite any sources that can be checked. The backwards search therefore only worked for the articles and conference papers. That opened up more research fields of the large number of topics that were already identified through the detailed database search. Here, I had to narrow down on the literature that was used to support the key statements of the papers, as including even more papers at this stage would have diffused the findings. The same applies for the forward search. It was done using the “cited by” function of google scholar to identify relevant papers that also researched the topic of strategic data acquisition.

Step 4: Analyzing findings

The relevant literature was then further examined and analysed. The papers and articles were read and then clustered in an author – topic matrix. As there are many data sources and usages companies can consider, they were clustered accordingly. The clusters for data sources were internal data / external data and structured data / unstructured data. For the data usage, the literature was divided into data defense and data offense strategies.

Some articles included both sourcing and usage strategies as well as both internal / external data and data defense / offense methods. They were added accordingly. Next, the review was conducted by extracting the key statements of every paper and article as well as comparing some statements and approaches to identify differences in literature.

At the end, the findings were discussed with regard to similarities, differences, limitations and potentials of further research.

4 Findings

4.1 Definition of Strategic Data Acquisition

As previously stated, there is no general and widely accepted definition for the term strategic data acquisition. The goal of this chapter therefore is to analyze the different approaches by the authors, identify similarities and key aspects of strategic data acquisition and then synthesize a concluding definition.

To do that, we first need to find a baseline of what a strategy is. There is a wide array of often used definitions, with Porters (1980) being the most famous and critically acclaimed one. He defines a strategy as the “...broad formula for how a business is going to compete, what its goals should be, and what policies will be needed to carry out those goals.”. Furthermore, he defines it as the “...combination of the ends for which the firm is striving and the means by which it is seeking to get there.” (Porter 1980, p.150).

Another commonly used description of a strategy comes from Mintzberg, known as the 5 Ps for strategy: companies need to have a plan, a ploy, a pattern, a position and a perspective (Mintzberg 1987).

A simpler definition was made by Watkins (2007). He claims the Goal of an enterprise is defined as the “what”, the vision as the “why” and the strategy as the “how”.

What all definitions have in common is that they define a strategy as a formula on how an enterprise has to act to reach its business goals.

The term (big) data acquisition also has different definitions and approaches that vary slightly based on the viewpoint of the authors. Lyko et al (2016) point out, that data acquisition describes the processes of “gathering data from distributed information sources with the aim of storing them in […] data storages.” In addition, they identified 3 main components: 1. Protocols that allow the data gathering, 2. Frameworks with which the data is collected and 3. Technologies that allow the persistent data storage.

Sagiroglu and Sinanc (2013) further added, that data collection includes the collection and managing of vast volume and different types of data in a way, that you can extract meaningful value from it. The

sources of which the data originates varies from case to case and also depends on the business goal – that will be discussed in the next chapter.
To further understand data acquisition as part of a big data strategy, we look at it as part of a big data value chain created by Lyko et al. (2016).

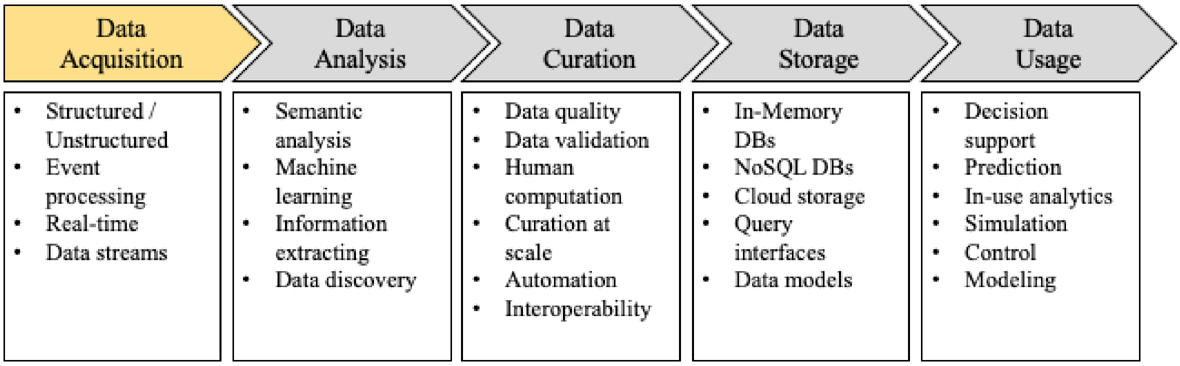

Figure 1. Big data value chain (Lyko et al. 2016)

As we can see here, the data we collect is then analyzed, stored and used for decision support, control or predictions. In return, that implies that we need to know what data to collect to then use it later in the process. That is also one of the key statements of the already existing definition approaches.

Bohnhoff (2019) hasn't stated a clear definition, but rather presented a framework. He divided strategic data acquisition in six core dimensions or questions: 1) Why – for what purpose the data should be used, 2) How – how you collect data, that suits the business goals, 3) What – what the data objective or data format is, 4) When – how regularly you have to receive new and current data, 5) Where – which data sources can or should be used, 6) Who – who else has access to this data.

In the digital age, strategic data acquisition is necessary to remain competitive. Gathering data about your business is crucial to identify needs and behavioral patterns of customers, on which you can then adjust or optimize your strategy (Websensa 2020). Strategic data acquisition is also "vital in collecting relevant data that corresponds to getting the desired outcome" (Websensa 2020).

In total, we can define strategic data acquisition as an approach to data collection, in which data is gathered from various sources (e.g. customer data, sales data, data from an enterprise resource planning system or external data) with a strategic approach, respectively with the goal to use the data long term to gain competitive advantages, discover new information of the enterprise or support product development. In any case, the source of the data is critical for the success of the project.

4.2 Data Sources

As part of a study that examined the complexity of data strategies, Kandogan et al (2014) conducted interviews with experts in the field of business analytics. One of the key aspects they elaborated is that the complexity arises from the data sources on which a data strategy builds its foundation. The more data sources are being considered and combined, the more complex it becomes, but also the more effective it is. This effect has also been described by Gür & Spiekermann (2020) in their technical analysis on data strategies. This insight supports the argument of the importance of strategic data acquisition that has been elaborated in the last chapter.

Fleckenstein and Fellows (2018) also point out, that the data collection part of the operation is critical, but also very investment intensive for it to be effective.

To get an overview on the wide array of data sources a company can chose from, we first cluster them according to the common ground of the current research, structured vs. unstructured data and internal vs. external data sources.

Data Type	Structured Data	Unstructured Data
Definition	Data is already formatted to a set structure before being stored	Data is stored in its native format and is not processed or formatted
Pro	• Easily understood by users • Easily used for business analytics • More tools currently available	• More data flexibility • Quicker and easier collection • Storage in data lake, more cost effective
Contra	• Predefined purpose limits use • Limited storage options • Change in requirements is laborious and expensive	• More complex, requires data science expertise • Requires specialized tools

Table 2. structured vs. unstructured data (IBM 2021)

Data Type	Internal	External
Definition	Data that is already owned and can be used by the enterprise	Data that is not yet owned, but has to be acquired or bought to be used
Pro	• Already owned, less acquisition cost • Easier to understand as the data is already formatted and used in other processes	• Richer on information • Data can be specifically acquired to fit the business needs
Contra	• Less information value for competitive advantages	• Acquisition process can be difficult and costly • More external sources result in higher complexity

Table 3. Internal vs. external data (Gür & Spiekermann 2020, Breakthrough Data 2022)

First, I want to discuss possible internal data sources and their benefit for the enterprise in context of strategy. The primary source for internal data is ERP Software. It usually contains customer data, supplier data and sales data (Holsapple and Sena 2005). In a survey conducted by McKinsey (2016) they identified that businesses, who deeply analyze existing sales data have an overall better market performance. Green (2005) points out the significance of customer data. With the information, an enterprise can better understand their customer base and target audience, learn of their behaviour and optimize the strategy accordingly.

Also, supplier data, e.g. capacities or historical delivery dates, gets increasingly more relevant, especially due to the current and upcoming supply chain disruptions (Zagorin, 2022). Companies should therefore consider implementing supplier data into their business to minimize the risk.

Mueller-Freitag (2016) also points out, that manual data input from employees can be also be an effective internal data source, for example for the training of chatbots.

Next to internal data sources, the literature also discusses the possibilities of external data sources. One of those is data generated via social networks. They are full of unstructured data, both qualitative (e.g. personal opinions, comments) and quantitative (e.g. user statistics, likes). Data from social networks is described as having high potential for companies as it is publicly available and free (Lyko et al. 2016, Rusitschka and Curry 2016).

Generally speaking, open data sets are often a valuable source. Market data for example is good to understand the market and the target audience (Hussain and Prieto 2016). Also, demographic data is publicly available through various sources. Companies like Yahoo also made huge datasets publicly available for free (Mueller-Freitag 2016)

Next to all these datasets that already exist, there are a lot of possibilities to obtain and generate new data sets. The benefit with this is a significant competitive advantage (Sorescu 2017). The goal therefore is to generate as much useful data as possible. On this account, Saura et al. (2021) address possible privacy concerns, so they differentiated between two types of user generated data: (1) Where the user is intentionally generating data and (2) when the data is generated non-intentionally.

One of the intentionally generated data methods is crowdsourcing. In crowdsourcing, a company uses a crowd, e.g. the user base of their platform or product, to get information or perform specific organizational tasks (Howe 2006). Typical cases are small survey between Facebook posts or google captcha (Mueller-Freitag 2016). Buettner (2016) already executed a systematic literature review on the topic of crowdsourcing research and identified that the most discussed advantages are cost, speed, quality, flexibility, scalability and diversity. Some companies, especially digital natives like amazon, eBay or Google, heavily rely on crowdsourcing as a data source to improve their products or platforms (Brynjolfsson et al. 2011).

Another method to collect user generated data are loyalty programs. They classify as both intentional and non-intentional (Saura et al. 2021). Generally speaking, loyalty programs are considered a win-win situation for both the customer and the company, as the customer gets bonusses in exchange for personal data the company can use and further improve on. This effect has already been described relatively early by Reichheld and Schefter (2000). On the other hand, the negative effects are privacy concerns, inequality concerns and sustainability concerns. The privacy concerns are also addressed by Jai and King (2016) in a study on the effect of loyalty programs on the customers willingness to share personal information with third-party advertisers.

Another topic of interest in modern data acquisition is user in the loop. It describes the possibility of the product or platform itself generating data by usage of the user (Mueller-Freitag 2016). An example for this is tracking the behaviour of a user on a website (Lotame 2019), e.g. in the case of Netflix. The recommendation engine runs on machine learning algorithms, the data to train these algorithms are generated by the user via their viewing history, the time spent on a video or their personal MyList (Kaushal 2022).

Other collection methods discussed in the literature are sensors as a commonly used way to generate internal data at little cost (Lyko et al. 2016, Becker et al. 2016, Kakatkar and Spann 2019). But as authors generally use a more technical approach or don’t go further in detail about which specific data is acquired and how it is used, it has a relatively little place in the context of digital transformation.

This also partly applies to web mining (Mueller-Freitag 2016, McKinsey 2016). Web mining is not a clearly defined term and therefore will not be further elaborated in this literature review.

4.3 Data Usage

As there are a lot of different data sources companies can extract data from, the literature also discusses potential usage scenarios. First of all, as described previously, many systems already generate and collect large amounts of data. But only a small fracture is actively used in business processes. In addition, many of these systems lack the possibility of real time usage of the data (Lyko et al. 2016).

The literature generally distinguishes between defensive and offensive data strategies / data usages. Defensive strategies try to minimize the risk of or through data. Examples for that are detecting and / or limiting fraud (Davenport and DalleMule 2017). Offensive strategies are more focused on the customer side like sales and marketing. Most papers and articles revolve around offensive strategies, especially when researching the literature on strategic data acquisition specifically, so this literature review will therefore also focus on the offensive data usages.

A general discussion is the topic of fully digitized business models or business model innovation through data. Sorescu (2017) pointed out several ways to utilize data to innovate a business model by using DDDM. The company HelloFresh for example doesn't have a technological end product, but the platform customers use to order the food boxes generate the necessary data. Uber also uses DDDM and disrupted the cab market by presenting a platform which revolves around data. They connect drivers and passengers and calculate the fee via an algorithm. The algorithm is based on data they captured previously (distance travelled, time needed, fuel used) (Pereira 2022). Triosi et al. (2020) showed in an empirical study, that B2B Marketing can benefit from data driven modelling as well, generating multiple advantages like economical, marketing- and knowledge-based.

Next to that, Parvinen et al. (2020) research about directly monetizing data. Part of a digital, data-based business model can be to sell the data that is being generated. It could be either raw and unstructured or aggregated and structured data, with latter increasing the value of the data set. To do that efficiently, companies need capabilities accordingly, which many don't have. Data monetization is also addressed by Nicolau and McKnight (2006). Their research focusses on perceived data quality in data exchanges and also point out the importance of structuring.

Machine Learning is also a topic of interest. Li et al (2021) have suggested to obtain user generated data specifically to improve machine learning algorithms and therefore enhance their predictions. This is done by Netflix as part of their recommendation engine (Kaushal 2022). Bakthiani et al (2018) suggested several different machine learning approaches that rely on big data acquisition, e.g. in combination with different databases or data sets. Vatres & Masetic (2022) proposed to use customer data and customer segmentation with a machine learning model to predict shopping behaviour. With the predictions, specific actions to adapt to the circumstances can then be derived.

There are also indirect ways to monetize data. Mueller-Freitag (2016) suggests, that a large and specific dataset can be a way to jumpstart data network effects. He describes the process of users generating data, therefore improving the quality of a product or platform. When the quality increases, the chance of more people getting into the product or onto the platform increases. That way more users generate even more data, creating a network effect. This effect is also addressed by Wilberg et al. (2017), who point out that it is critical for companies to adjust their processes, organizational structure and product design accordingly. These kinds of devices, that are based on user in the loop or user generated data principles, are transforming the competition and will be a source of major competitive advantages (Porter and Heppelmann 2014). Provost and Fawcett (2013) claim, that there currently is a revolution in online advertising due to users spending more time on platforms generating data. An example for this is Amazon's recommendation engine, that is progressively improved by data.

As mentioned, products can not only be used to generate data, but data can also support product innovation. Wilberg et al. (2017) for example show how there are a large range of possibilities for using data in the product innovation process. This is strongly linked to the market the company operates in (Porter & Heppelmann 2014). Some use cases of data in product development are product viability, informed product decision making, product progress measurement, user experience insights and product development inspiration (EMLV 2020). Also, a company that is rather inexperienced with using data in product development should carefully consider and define a clear data strategy before starting the process (Wilberg et al. 2017). On this matter, Kayser et al. (2019) presented a data collection map, where companies that are starting a data driven innovation process can strategize their data acquisition process beforehand to have a common understanding of possible sources and raising overall data awareness.

Deriyenko et al. (2015) suggest a model in which the customer data from product interaction gets connected with social media posts to identify the customer needs and therefore support the product optimization process.

Referring to data acquired via loyalty programs, Loyalty Science Lab (2020) shows 6 use cases of user generated data: Customer segmentation, customer lifetime value analysis, customer attrition risk analysis, shopping basket analysis, customized marketing communications and testing and experimentation.

5 Discussion

As already mentioned in the methodical approach to this review, the literature that is focussing on strategic data acquisition as a holistic topic is scarce, especially when searching for definitions. To address this issue, the search terms were split up into different categories in order to structurally analyze the research topic components individually. During the review of the literature, it was identified that the strategic data acquisition process essentially consists of two major steps: data acquisition/collection from different sources and data usage. There are still various intermediate steps that deal with data storage and data preparation. However, these are technical in nature and therefore not included in the scope of this review.

In conclusion, I focussed on a three, non-peer reviewed articles that discuss the topic and bring up different forms of definitions, data sources and data usages. The detailed research to the topics were then conducted on the individual methods. Subsequently, it is not very useful to identify all individual differences in the literature, as only a few authors deal with several methods. Nevertheless, there are some differences concerning the literature in general.

First, different focal points in the definition of the term data acquisition and data strategy were identified. That is not unusual as every researcher has different focal points himself and therefore a different view on the topic. With the joint merging of the approaches, a definition was created. For further research, it would be useful to weave the topics of data acquisition and strategic alignment more closely together to further clarify the definition.

Second, the wide range of available data sources was summarized. Again, it is difficult to identify individual differences because the authors each cover different topics. The data source selected by the company varies depending on which data (internal, external, customer-specific, supplier-specific, market-specific, demographic) is required in which structure (structured or unstructured, qualitative or quantitative). A common feature of the literature reviewed is the mentioning and discussion of the high cost that can result from data acquisition. This has been demonstrated in some studies. The wide range of data sources for different application areas supports this. Therefore, in the case of a practical application, companies should clearly define in advance which data is needed in which form for their concern in order to avoid additional effort as far as possible. Although Bohnhoff (2019) has already presented a framework on how to conduct a strategic data acquisition, the results were not yet empirically validated. Future research could address the question of the extent to which the application of a framework can provide the company with a significant cost advantage. Furthermore, it should be thoroughly investigated which data should be explicitly collected for which purpose.

Third, data usages were as varied as the sources the data originates from. Some authors address the question of how collected data can be monetized directly. In doing so, some authors address the privacy aspect, while others focus on the data quality implications. Another focus of the literature is the indirect monetization of data. Again, many different approaches are taken by the authors. Since there is little holistic research available, detailed differences can only be identified with much more extensive research.

Overall, this research is limited by the fact that too few peer-reviewed papers were used. The quality of the articles dealing with strategic data acquisition cannot be proven flawlessly. Care was taken to ensure that articles from management consultancies, best practices or industry specialists were predominantly used. However, in order to represent the complete state of the art, unknown sites also had to be included. The results they elaborated seemed conclusive in context but were not empirically or qualitatively verified due to my lack of expertise on the topic.

For the detailed search of the methods, peer reviewed papers and conference papers were included. Despite their higher quality standard, many of the papers used were not directly revolving around researching the topic of data acquisition itself, but data acquisition was rather part of the conclusion.

Another limitation comes from the limited time that has been invested in this review. As it was created as part of a university seminar in part time over a few months, it is not assured that all relevant papers have been identified. The lack of access to more scientific databases to find more paper add to this.

6 Conclusion

This literature review was conducted as part of a seminar on IT Management in the digital age. The starting point were several articles revolving around the subject of DDDM and data strategy to get an overview on the topic as a whole. It became clear that there is no universally accepted definition or framework on strategic data acquisition. What was also found is that strategic data acquisition has to be viewed in the context of data strategies, as the wide range of possible data sources and data usages varies depending on the chosen data strategy. Also, DDDM gets increasingly more relevant. For DDDM to be effective, companies need sufficient data regarding their business processes, customers, suppliers and products. Based on this knowledge, this literature review was conducted in 3 steps.

First, based on the initial review of the literature, a general information gathering was carried out to identify 3 key papers (see chapter 3). Based on these key papers and some supplementary literature, a definition for strategic data acquisition was formed.

Next, I did a detailed database search for the identified sources and usage possibilities to discuss the different approaches. In total, I identified nine possible data sources for DDDM: (1) ERP data, (2) manually input data from employees, (3) social networks, (4) publicly available datasets, (5) crowdsourcing, (6) loyalty programs, (7) user in the loop data, (8) sensors, (9) web mining. The decision for a certain data source has to be made depending on the required data. User in the loop data for example is useful in product development, as data collected from the user provides insight into usage and clues for optimization. To identify the users' needs as part of product optimization, qualitative data from social networks can provide the necessary information.

This also brings us to the third step, researching the topic of data usage. While the literature research was conducted, countless possibilities to use generated data were found. The results were narrowed down to strategically relevant fields of action. In total, nine use cases were presented: (1) detecting and / or limiting fraud, (2) digitizing business models or business model innovation, (3) DDDM, (4) direct monetization, (5) machine learning algorithms, (6) jumpstarting data network effects, (7) support product innovation, (8) support product optimization, (9) various customer specific analysis. As with the data sources, companies need to elaborate a plan before they start collecting or using the data, otherwise they could face extra work due to the high complexity and novelty of the topic.

In the discussion, key similarities and differences of the papers were identified. However, due to the lack of holistic research and resources for this review, not all differences may have been revealed. In total, this review should give a sufficient overview on the topic as a whole and provide useful information for practice as well as avenues for future research.

References

Bakhtiani, R., Gandhi, M., Churi, P., Gupta, P., 2018. Big Data Strategies – A Review and Survey, in: 2018 4th International Conference on Applied and Theoretical Computing and Communication Technology (ICATccT). Presented at the 2018 4th International Conference on Applied and Theoretical

Computing and Communication Technology (iCATccT), IEEE, Mangalore, India, pp. 287–293.

Bohnhoff, T., 2019. How to set up a corporate data strategy? URL https://medium.com/appanion/how-to-set-up-a-corporate-data-strategy-b61df1c329c5 (accessed 5.3.22).

Boston Consulting Group. 2020. Data Dominates. Available on: www.bcg.com (accessed 5.9.2022)

Breakthrough Data 2022. Data Strategy Canvas. Available on www.prometheusresearch.com (accessed 5.25.22).

Brynjolfsson, E., Hitt, L.M., Kim, H.H., 2011. Strength in Numbers: How Does Data-Driven Decisionmaking Affect Firm Performance? SSRN Journal.

Buettner, R., 2015. A Systematic Literature Review of Crowdsourcing Research from a Human Resource Management Perspective, in: 2015 48th Hawaii International Conference on System Sciences.

Presented at the 2015 48th Hawaii International Conference on System Sciences (HICSS), IEEE, HI, USA, pp. 4609–4618.

Cavanillas, J.M., Curry, E., Wahlster, W. (Eds.), 2016. New Horizons for a Data-Driven Economy. Springer International Publishing, Cham.

DalleMule, L., Davenport, T.H., What’s your Data Strategy? Harvard Business Review May-June 2017,

Deriyenko, T., Hartkopp, O., Mattfeld, D.C., 2017. Supporting Product Optimization by Customer Data Analysis, in: Dörner, K.F., Ljubic, I., Pflug, G., Tragler, G. (Eds.), Operations Research.

Proceedings 2015, Operations Research Proceedings. Springer International Publishing, Cham, pp. 491–496.

EMLV, 2020. How Can Data Analytics Improve Product Development? Available on: www.emlv.fr. (accessed on 06.13.2022).

Fleckenstein, M., Fellows, L., 2018. Modern Data Strategy, 1st ed. Cham.

Green, A., 2012. Understanding the value of customer data. J Direct Data Digit Mark Pract 13, 221–233. https://doi.org/10.1057/dddmp.2011.47

Gür, I., Spiekermann, M., 2020. Data Strategy Praxis Report, Fraunhofer ISST

Himmi, J., Guan, P., Zhou, W., 2017. Value Oriented Big Data Strategy: Analysis & Case Study. Presented at the Hawaii International Conference on System Sciences 2017. https://doi.org/10.24251/HICSS.2017.124

Holsapple, C.W., Sena, M.P., 2005. ERP plans and decision-support benefits. Decision Support Systems 38, 575–590.

Howe, J., n.d. The Rise of Crowdsourcing. Available at: Wired.com (accessed 06.21.22)

Hurley, J., n.d. Why your data strategy is your B2B growth strategy. Harvard Business Review analytic services.

IBM. 2020. Why investing in your data is now more important than ever. Available on: www.ibm.com. (accessed 5.9.2022).

IBM 2021. Structured vs. unstructured data – what is the difference. Available on: www.ibm.com (accessed 06.21.22)

Jai, T.-M. (Catherine), King, N.J., 2016. Privacy versus reward: Do loyalty programs increase consumers' willingness to share personal information with third-party advertisers and data brokers? Journal of Retailing and Consumer Services 28, 296–303.

Kakatkar, C., Spann, M., 2019. Marketing analytics using anonymized and fragmented tracking data. International Journal of Research in Marketing 36, 117–136.

Kandogan, E., Balakrishnan, A., Haber, E.M., Pierce, J.S., 2014. From Data to Insight: Work Practices of Analysts in the Enterprise. IEEE Comput. Grap. Appl. 34, 42–50.

Kaushal, V., 2022, Evolution of ML Fact Store, available on: https://netflixtechblog.com (accessed 06.21.2022).

Kayser, L., Mueller, R.M., Kronsbein, T., 2019. Data Collection Map: A Canvas for Shared Data Awareness in Data-Driven Innovation Projects. Proceedings of the 2019 Pre-ICIS SIGDSA Symposium Winter 12-2019

Kitchin, R. and McArdle, G. (2016) 'What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets', Big Data & Society.

Li, Y., Yu, X., Koudas, N., 2021. Data Acquisition for Improving Machine Learning Models. Proceedings of the VLDB Endowment 14, 1832–1844.

Loyalty Science Lab, 2020. 6 use cases of loyalty program data, available on: www.medium.com (accessed on 06.21.2022).

McKinsey 2016. Unlocking the power of data in sales. Available on: www.mckinsey.com (accessed 5.3.2022).

McKinsey 2022. Digital strategy in the postpandemic era. Available on: www.mckinsey.com (accessed 5.3.2022).

Mintzberg, H. (1987) 'The Strategy Concept I: Five Ps for Strategy', California Management Review, 30(1), pp. 11–24.

Mueller-Freitag, M., 2016. 10 Data Acquisition Strategies for Startups. Medium. URL https://medium.com/@muellerfreitag/10-data-acquisition-strategies-for-startups-47166580ee48 (accessed 5.7.22).

Nentwich, C., Reinhart, G., 2021. Towards Data Acquisition for Predictive Maintenance of Industrial Robots. Presented at the 54th CIRP Conference on Manufacturing Systems, Athens.

Nicolaou, A.I., McKnight, D.H., 2006. Perceived Information Quality in Data Exchanges: Effects on Risk, Trust, and Intention to Use. Information Systems Research 17, 332–351.

Parvinen, P., Pöyry, E., University of Helsinki, Gustafsson, R., Aalto University, Laitila, M., Futurice, Rossi, M., Aalto University, 2020. Advancing Data Monetization and the Creation of Data-based Business Models. CAIS 47, 25–49.

Pereira, D., 2020. Uber Business Model. URL https://businessmodelanalyst.com/uber-business-model/ (accessed 5.28.22).

Porter, M. E. Competitive Strategy: Techniques for Analyzing Industries and Competitors. New York: Free Press, 1980.

Porter, M.E., Heppelmann, J.E., 2014. How Smart, Connected Products Are Transforming Competition. Harvard Business Review.

Provost, F., Fawcett, T., 2013. Data Science and its Relationship to Big Data and Data-Driven Decision Making. Big Data 1, 51–59.

Reichheld, F.F., Schefter, P., 2000. E-Loyalty: Your Secret Weapon on the Web. Harvard Business Review.

Sagiroglu, S., Sinanc, S., 2013 Big data review. International Conference on Collaboration Technologies and Systems (CTS), 2013, pp. 42-47.

Saura, J.R., Ribeiro-Soriano, D., Palacios-Marqués, D., 2021. From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. International Journal of Information Management 60, 102331.

Sorescu, A., 2017 Data-Driven Business Model Innovation. Journal of Product Innovation Management 34, 692–696.

Triosi, O., Maione, G., Grimaldi, M., Loia, F., 2020. Growth hacking: Insights on data-driven decision-making from three firms. Industrial Marketing Management 90, 538–557.

Waller, M.A., Fawcett, S.E., 2013. Data Science, Predictive Analytics, and Big Data: A Revolution That Will Transform Supply Chain Design and Management. Journal of Business Logistics 34, 77–84.

Websensa 2020. Strategic Data Acquisition, available on: www.websensa.com (accessed 06.11.2022)

Webster, J., Watson, R.T., 2002. Analyzing the Past to Prepare for the Future: Writing a Literature Review. MIS Quarterly 26, xiii–xxiii.

Wilberg, J., Triep, I., Hollauer, C., Omer, M., 2017. Big Data in Product Development: Need for a Data Strategy. Presented at the 2017 Portland International Conference on Management of Engineering and Technology (PICMET), Portland.

Vatreš, A., Mašetić, Z., 2022. Exploring Customers' Behavior – Analysing Customer Data, Customer Segmentation and Predicting Customers' Behavior on Black Friday, in: Ademović, N., Mujčić, E., Akšamija, Z., Kevrić, J., Avdaković, S., Volić, I. (Eds.), Advanced Technologies, Systems, and Applications VI, Lecture Notes in Networks and Systems. Springer International Publishing, Cham, pp. 15–26.

Zagorin, E., n.d. Council Post: The Emerging Role Of Supplier Data Exchanges For Digital Procurement Teams [WWW Document]. Forbes. URL: https://www.forbes.com/sites/forbestechcouncil/2022/02/08/the-emerging-role-of-supplier-data-exchanges-for-digital-procurement-teams/ (accessed 5.25.22).