Tampereen yliopisto

Tiia Wallin

PLANNING A DATA STRATEGY FOR A STARTUP COMPANY

Faculty of Information Technology and Communication Sciences
Master's thesis
August 2022

ABSTRACT

Tiia Wallin: Planning a data strategy for a startup company
Master's thesis
Tampere University
Master's Programme in Computer Science
August 2022

In both business and society, data and its use are one of the most urgent issues of the day. Data is one of the most valuable assets for companies, but it has to be strategically gathered and used to benefit from it. Thus every company nowadays requires a data strategy. The research question of the thesis aims to find out what data would be the most useful in a product launch for a startup company creating a C2C platform, and how the gathered data could be used most effectively.

This thesis consists of two parts: a literature review and a case study. The literature review provides a general idea of a data strategy as a concept, how it is studied in both academic and practical research, and how to create a data strategy for an organization. Based on the knowledge provided by the literature review, a case study is conducted. The case study is commissioned by a small IT startup company, and it creates an actionable short-term data strategy plan for a single business issue, with a focus on the time after the initial product launch and during the Proof of Concept phase. The plan will be used in the company to create a data strategy after the release of the thesis to gather and use relevant data for business and product improvements before launching into a new market area.

Keywords: data strategy, startup, data management, data-driven

TABLE OF CONTENTS

1 Introduction *1*

1.1 Research problem **2**

1.2 Research structure **3**

2 Research methods *5*

2.1 The research process of the thesis **5**

2.2 Information systems research and design research **7**

2.2.1 Design science research guidelines 9

2.3 Design research artifact **11**

2.3.1 Construction and evaluation of an IT artifact 11

2.4 Literature review and case study methods **14**

3 Research on data strategy *16*

3.1 Data, knowledge, and information **16**

3.2 Definitions of data strategy and related terms **18**

3.3 Academic research **20**

3.4 Practical literature **21**

4 Formation of a data strategy *23*

4.1 Why does a company need a data strategy? **23**

4.2 Elements and components of a data strategy **24**

4.3 Creating a data strategy **27**

4.4 Product launch phase strategy **29**

5 Case study *31*

5.1 Case study; a used mobile devices marketplace **32**

5.1.1 Product 33

5.1.2 Market positioning 34

5.2 Business strategy and goals **35**

5.3 Assessment of current state **37**

5.3.1 Expectations for the future 37

5.4 Proposed future state: Data from Proof of Concept **40**

5.4.1 Dashboards 43

5.5 Strategy roadmap for Q2/2022 – Q3/2022 .... **46**

5.6 Future iteration of the data strategy **47**

6 Discussion *49*

6.1 Limitations and future research **50**

7 Conclusions *52*

8 References *53*

1 Introduction

Data and its usage are one of the most pressing topics of the day both in business and society, and no wonder: nowadays the world is run predominantly by data. This became painstakingly clear to the public in 2018 after the Cambridge Analytica data scandal [Confessore, 2018]. To simplify, those who hold the data hold the power. But just holding the data is not enough: it needs to be thoughtfully gathered and purposefully used to benefit from it, and to prevent it from becoming a liability as it did in the Cambridge Analytica case. This is where the role of a data strategy is emphasized: with a systematic approach to data gathering and usage it is much easier to gain desired results.

Data can be a significant strategic asset when used right. Organizations function in much more efficient ways with the support of high-quality, timely information. Time is not wasted searching for information, and decisions are based on facts instead of assumptions. Creating a solid data strategy helps organizations plan and execute appropriate actions to gather data and turn it into useful information.

Over the last few years, there have been a lot of discussions – both academic and practical – about things such as big data, analytics, and data usage in business, and many organizations have started to purposefully gather and use data to improve their businesses. Big Data and other emerging technologies are increasingly being used to analyze and gain insights that are related to decision-making processes [Ranjan & Foropon, 2021]. Tabesh and others [2019] describe this current era of big data analytics as “a new gold rush”, referring to the 19th-century gold rush that significantly shaped American history. Today, the capacity to swiftly gather and analyze large amounts of data has spawned a new gold rush, with many firms and their analysts devoting significant time, money, and effort to gain business insights from the behavioral and structural aspects of their surroundings [Tabesh et al. 2019].

Despite significant recent improvements in big data analytics and the utilization of data in business, there is a lot of evidence that many businesses still fail to properly incorporate data and analytics into their decision-making processes [Tabesh et al. 2019]. There are also some areas and point-of-views that have received less attention in the debates and research, for example, small and medium-sized enterprises (SMEs) and platform businesses. There have been very few academic discussions about data strategies and data-related decisions specifically in platform businesses [Bhargava et al. 2020].

1.1 Research problem

The goal of this thesis is to create a plan for a solid, research-based data strategy base for a case study company, with the main focus on one narrower question or use case. The company has been operational for less than a year at the time of writing this thesis, and the product has not been launched yet. The idea is to create a knowledge base of data strategy and a high-level plan for creating one for a specific business issue that can be easily used and iterated once the product is launched, as well as be expanded into a more comprehensive data strategy.

The goal of the case study is to find out what kind of data collection would be most relevant from the perspective of product development and business goals, and how the collected data could be utilized most effectively. A special interest is in the months after the launch and before the expansion to a new market area. Based on this the research question is:

> *What data is relevant to gather and how to use the gathered data, in the first months after the initial product release of the C2C marketplace, to improve the business model and the product before entering a new market area?*

The question can be approached from three different sub-questions to create the full picture:

1. What data is relevant for data-driven decision-making in the big picture for the case company and why?
2. What data is the most relevant to focus on in the first months after the initial release of the product (C2C marketplace)?
3. How is the gathered data used most purposefully and efficiently to improve the product and the business model in the limited time before entering the next market area?

In addition to these three perspectives, there are several requirements to consider for the data strategy to be useful in the future:

1. Scalability: the high-level data strategy must be scalable when the company grows.
2. Simplicity: the data strategy must be simple enough to be widely adopted inside the company, both from process and technology perspectives.

3. Expediency: the data strategy is designed specifically for the company and with the market area of the company in mind to be as relevant as possible.

The theoretical framework of the thesis focuses on opening up the concept of data strategy and related terms, looking at how the topic has been researched in the past, and mapping out what the literature says about forming a data strategy. The theoretical framework aims to clarify for both the reader and the author the key concepts involved in designing a data strategy and to provide concrete guidelines for creating a data strategy. The case study will use the theories and knowledge from the theoretical framework to create recommendations and an actionable data strategy to answer the research question.

There are some limitations in the research question and goals of this thesis. Within this thesis, the goal is to create a high-level vision of what data should be focused on and how that data would be used to improve business and product development goals. Within this vision, the more detailed question of post-launch data gathering and usage is addressed. The timeline of the data strategy plan designed within the thesis is very short, and the plan will need iteration shortly after its creation. This is acceptable due to the fact that the product from which the data will be retrieved is not yet launched, and the iteration cycle can start only after the launch.

Another limitation is that the topic of this work focuses mainly on the C2C marketplace, and not the other components of the platform. This thesis also does not comment on precise technical solutions for data collection and management but focuses on the larger perspectives. With that said, it is desirable that this framework can be extended or combined with another to cater to the whole platform later on. It will hopefully create a good framework that can be used in the future to specify technical and other more specific needs.

1.2 Research structure

This thesis consists of several theoretical chapters based on literature and a case study for a small IT company. The theoretical chapters will introduce research methods, relevant terminology, previous research, and theoretical background for data strategy, and the case study will introduce the case company and focus on creating a high-level data strategy and data use case for the case company.

Chapter 2 introduces the research methods used in this thesis. The main research method is information systems design research with the end goal of creating an IT artifact. Chapter 3 introduces terminology related to data strategy as well as a scoping literature review of how the topic of data strategy has been addressed in the literature and research in recent years. In chapter 4, based on literature, it is justified why a data strategy is generally relevant to companies, what elements are often included in a data strategy, and what best practices the literature emphasizes when creating a data strategy. Chapter 5 focuses on the case company and the creation of the data strategy. Chapter 6 discusses the created data strategy, limitations of this research, and possibilities for future research. Chapter 7 concludes and summarizes this thesis.

2 Research methods

This chapter reviews all the research methods that are used in this thesis. This thesis approaches the research problem through the means of information systems design research. Both the research and the creation of the solution for the research problem aim to follow the information systems research framework created by Hevner and others [2004] as much as possible, as well as other widely accepted recommendations for the design and evaluation of IT artifacts. Design research is not limited to information systems research, but for clarification, in this thesis, the term "design research" will, from now on, indicate specifically information systems design research.

In chapter 2.1, the research methods regarding the case study are presented. The guidelines for a design research method and design research artifact are presented in chapters 2.2 and 2.3 respectively. In chapter 2.4, an overview of the literature review and a justification for choosing an approach is presented.

2.1 The research process of the thesis

The research in this thesis consists of two parts: one a literature review and the other a case study. The goal of the literature review in this thesis is to clarify the basic concepts, get an overview of data strategy literature, and establish a foundation of knowledge for the case study. The approach from which the subject is studied in the literature review is scoping review rather than a systematic review. The case study will use the results of the literature review as a knowledge base. Linking the case to the theories and research established in the literature review, a set of recommendations, in this case, a short-term data strategy or data use case will be created.

The literature review in this thesis is further divided into two parts: firstly, the aim is to provide an overview of terminology and what kind of research has been done on data strategy in recent years. Secondly, the materials are used to justify and establish best practices for the creation of a research-based data strategy. These best practices will be applied in the creation of the data strategy for the case company where applicable.

The research process for this thesis began with finding relevant materials. Various databases were used, but Andor (Tampere University library's search service), Scopus, Elsevier/ScienceDirect.com, IEEE, and Google Scholar have been the main databases and search engines used during the creation of this thesis.

The terms used when searching for relevant materials were as

- Data strategy
- Startup
- Business strategy
- Data management
- Data-driven development
- Data-driven decision making

With the terminology, searches such as the following were created

- ”Data strategy”
- ”Data strategy” AND startup
- Data-driven AND (development OR “decision making”)

The first and last of the search terms mentioned above resulted in tens of thousands of results, so the results had to be limited in various ways. The results were limited both according to relevance and the year of publication, excluding older releases. Also, as data strategy is nowadays related to nearly all professional fields, there was a lot of profession-specific literature in the reviews that were purposefully excluded from this thesis. One fascinating finding was that the search ”Data strategy” AND startup resulted in many articles but not ones that would handle the role or creation of a data strategy in a small, new company. In addition to finding materials with an open search with terms, the citation pearl growing strategy was also used for finding relevant materials through references. This method resulted in many relevant materials for this thesis.

With the case study, the focus was on building a data strategy base for the case company using the information systems design research method. In the case study, the focus was on a specific area of interest resulting in a data use case more than a fully designed strategy, and a more comprehensive data strategy needs to be created by the company later on.

During the case study, several unstructured interviews with different stakeholders (for example with management and with the person currently responsible for data gathering in the company) were conducted, and the expectations regarding both the general business strategy and the data strategy were discussed. Through the interviews and discussions, a better understanding of the business needs was formed, which are central components of the research.

The existing business strategy was familiarized through materials that the company provided, and its main points are referred to in this thesis. Together with the applicable knowledge gathered in the literature review, the assessment of business needs helps to comply with the guidelines of the Information Systems Research Framework created by Hevner and others [2004].

2.2 Information systems research and design research

The information systems (IS) research field is essentially multidisciplinary, with methods and theories deriving from fields such as computer science, psychology, and economics [Wade & Hulland, 2004]. IS research aims to increase knowledge that helps in implementing systems and processes in organizations and helps manage such organizations [Hevner et al., 2004]. Information systems design research can be seen as a category within IS research to create an innovative solution to a real-life problem. Hevner and others [2004] argue that collecting IS research knowledge requires two separate approaches that still complement each other: behavioral science and design science approaches.

The behavioral science model is based on natural science research and aims to establish and justify theories that describe and predict organizational and human events [Hevner et al., 2004]. The design science model is based on engineering, and it is a problem-solving model that aims to birth innovations to enable accurate analysis, design, implementation, management, and use of information systems [Hevner et al., 2004]. So to compare the two, the behavioral science model aims to new theories and attestation of truth by proving and overturning hypotheses, while the design science model aims to create new knowledge through application and problem-solving.

Hevner and others [2004] describe both the organizations and their supporting information systems to be purposefully designed, artificial and complex, consisting of people, technologies, work systems, and structures. The goal of design science is to create and evaluate solutions for identified organizational problems, whereas the work done by managers and IS practitioners focus on organizing resources purposefully to accomplish goals [Hevner et al., 2004]. Design science is a two-fold area with two different areas of development: processes and artifacts. The goals are both to create and improve processes as well as create tangible solutions for problems as a form of artifacts.

Hevner and others [2004] have created a conceptual framework of IS research that shows how design science and behavioral science are interconnected. This framework that is shown in Figure 1 can help to comprehend, carry out and assess the outcomes of IS research.

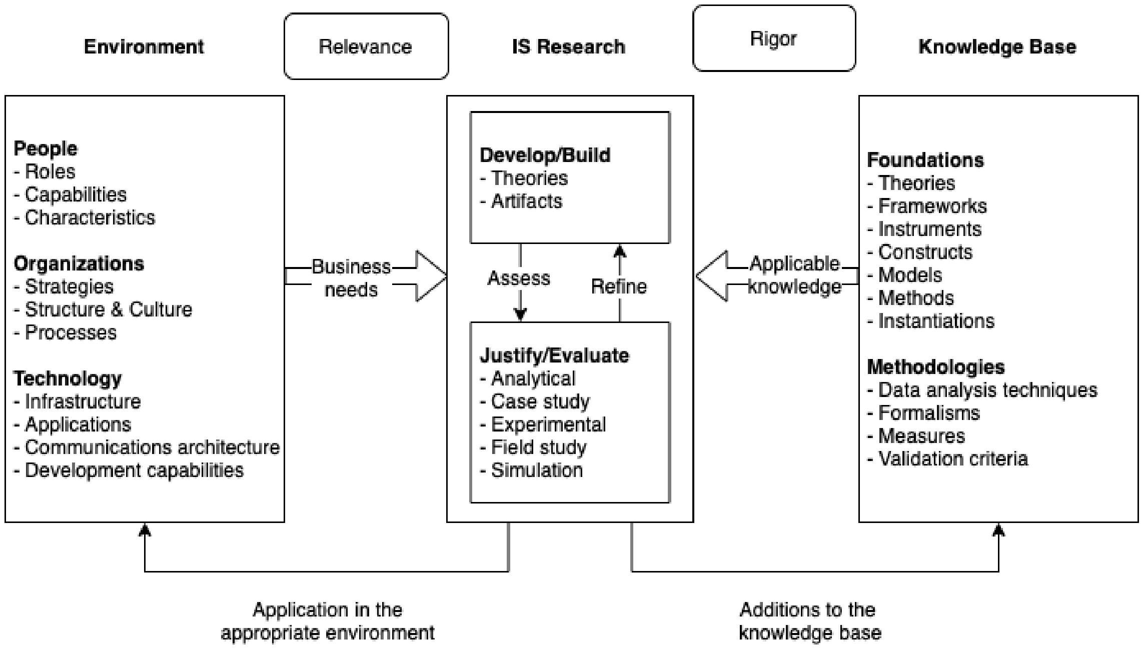

Figure 1. Information Systems Research Framework according to Hevner and others [2004].

The middle section, *IS research*, consists of two complementary stages of behavioral science and design science. Behavioral science approaches research by developing and justifying hypotheses relevant to previously identified business needs, whereas the design science approach creates and evaluates artifacts that intend to answer those previously identified business needs. Behavioral science aims to truth and design science to utility, and even though the outcomes of these two approaches differ, they are still inseparable. In both approaches, the evaluation through justify/evaluate procedures could lead to identifying flaws in theory or artifact, as well as the need to reiterate. [Hevner et al., 2004]

The left side section, *environment*, defines the area of interest, and in IS research the interests are multidisciplinary with a focus on people, organizations, and technology. Business needs are examined through the lens of organizational culture, people's roles and capabilities as well as organizational strategies and other aspects. Business needs are thus defined by the goals, opportunities, and problems of people inside the organization, and they are related to existing infrastructures, architectures, and development capabilities. The relevance of IS research is assured by addressing business needs. [Hevner et al., 2004]

The right side section, the *knowledge base*, is the base upon which the IS research is built. It consists of foundations and methodologies that help researchers in the justify/evaluate phase of research. These foundations and methodologies are also sources of rigor when they are applied to IS research. The typical methodologies in behavioral science are mostly empirical, whereas design science traditionally uses mathematical and computational methodologies but also applies empirical methodologies in some cases. [Hevner et al., 2004]

In IS research, several design activities have been widely researched and standardized, but as Hevner and others [2004] state, the theory of design in information systems is consistently in revolution. Design science in information systems research is addressing problems that are seen as difficult due to their nature. The characteristics of these problems consist of such as varying and poorly defined requirements, complexity, and reliance on human cognitive and social skills and abilities. These characteristics lead to innovative and creative design science procedures that result in significant technological advancements. [Hevner et al., 2004]

2.2.1 Design science research guidelines

In order for the design research goals to be formalized, Hevner and others [2004] have established a seven-step list of requirements and recommendations that every researcher in the field of design science research should consider when doing their research. They also stated that the goal of these guidelines is to help researchers, reviewers, and readers to have a consistent understanding of effective design science research requirements. The use of these is not mandatory, but Hevner and others [2004] believe that for design science research to be comprehensive, each of these requirements must be addressed in some way, and it is then up to the reviewers, editors, and readers to assess how well the study meets the goal of each of the requirements. The seven guidelines are listed in Table 1.

Guideline	Description
Guideline 1: Design as an Artifact	Design-science research must produce a viable artifact in the form of a construct, a model, a method, or an instantiation.
Guideline 2: Problem Relevance	The objective of design-science research is to develop technology-based solutions to important and relevant business problems.
Guideline 3: Design Evaluation	The utility, quality, and efficacy of a design artifact must be rigorously demonstrated via well-executed evaluation methods.
Guideline 4: Research Contributions	Effective design-science research must provide clear and verifiable contributions in the areas of the design artifact, design foundations, and/or design methodologies.
Guideline 5: Research Rigor	Design-science research relies upon the application of rigorous methods in both the construction and evaluation of the design artifact.
Guideline 6: Design as a Search Process	The search for an effective artifact requires utilizing available means to reach desired ends while satisfying laws in the problem environment.
Guideline 7: Communication of Research	Design-science research must be presented effectively both to technology-oriented as well as management-oriented audiences.

Table 1. Design Science Research Guidelines according to Hevner and others [2004].

These guidelines are not comprehensive or complete, and the researcher can use their own discretion in the application of them. Assessing the compliance of guideline 1 (design as an artifact) is quite straightforward for the researcher, but for example, compliance with guidelines 3 (design evaluation) and 5 (research rigor) might be harder for the researcher to assess themselves. Research needs to be able to withstand critical reviews by the community of other experts, and the researcher must demonstrate compliance with the guidelines. There must be a dialogue between the researcher and the target audience to show that guidelines such as 2 (problem relevance), 4 (research contribution), and 7 (communication of research) are complied with. Compliance with these three guidelines is not established until approval from the target audience. The 6th guideline (design as a search process) is often complied with during the relevance cycle (presented in chapter 2.2.1) that is also realizing the approval of the target audience.

2.3 Design research artifact

A *design artifact*, or an IT artifact, is a result of information systems design research that aims to create a concrete, utilizable solution to a research problem, for example, a prototype in the software development process. Hevner and others [2004] describe IT artifacts to consist of “constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems)”. Orlikowski and Iacono [2001] describe it as “bundles of material and cultural properties packaged in some socially recognizable form such as hardware and/or software”. Design research is a creative process with the goal of developing both the output (artifact) as well as the process of designing the output. The IT artifacts created in the design science process are often a study object for IS behavioral science, as the behavioral science theories attempt to predict or explain how the artifact works when implemented into the organization [Hevner et al., 2004].

The creation of a design artifact is a two-part iterative process consisting of constructing and evaluating the artifact while following design science research guidelines. The guidelines and process of creating an IT artifact will be presented in more detail in subchapter 2.3.1.

2.3.1 Construction and evaluation of an IT artifact

As mentioned, the creation of a design artifact is an iterative process. It starts with defining the problem area which is the environment in which the phenomena of research occur and where the artifact will be applied to. Figure 2 shows the information systems research framework created by Hevner and others in 2004, with overlays of three inherent research cycles added by Hevner in 2007 which he feels that need to be present and identified in any design science research process. These three cycles are the *relevance cycle* between contextual environment and design science activities, the *rigor cycle* between the knowledge base and design science activities, and the *design cycle* between research processes and core activities of the building and creation of design artifacts.

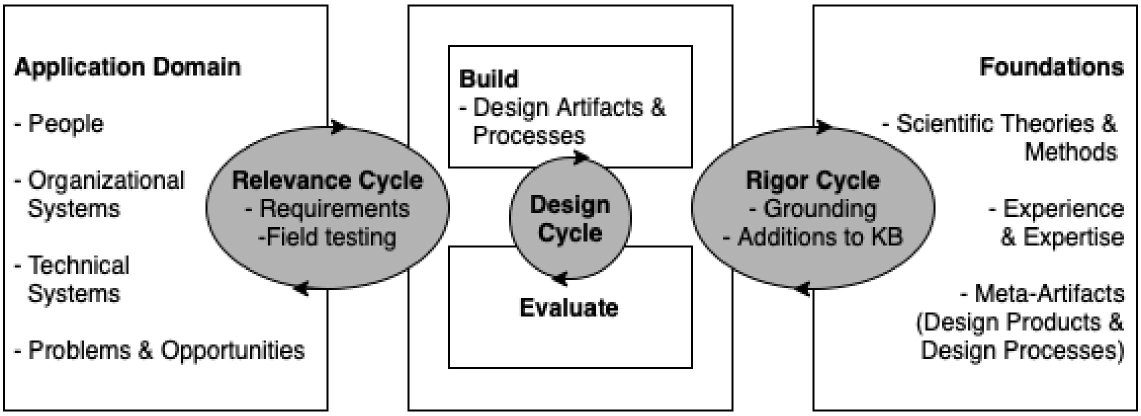

Figure 2. Design Science Research Cycles according to Hevner [2007].

The relevance cycle starts the design process by offering the requirements and acceptance criteria for the artifact. The rigor cycle offers information on previous research so that the researcher can make sure the research is innovative and creates new value. The researcher is expected to select and apply suitable methodologies and theories to their research for the research to be rigorous. The design cycle is at the heart of design science, in which alternatives to the created artifact are evaluated and iterated until an acceptable solution that meets the requirements is found. [Hevner, 2007].

The evaluation of an artifact is based on requirements that are established by business needs. According to behavioral science theory, an IT artifact must be evaluated with relevant metrics: the utility, quality, and efficacy of the artifact must be proven through methods that are predefined and precise [Hevner et al., 2004]. Hevner and others [2004] have compiled a list of evaluation methods to help assess a design artifact. This list can be seen in Table 2.

Type	Evaluation Method
1. Observational	Case Study: Study artifact in depth in business environment
	Field Study: Monitor use of artifact in multiple projects
2. Analytical	Static Analysis: Examine structure of artifact for static qualities (e.g., complexity)
	Architecture Analysis: Study fit of artifact into technical IS architecture
	Optimization: Demonstrate inherent optimal properties of artifact or provide optimality bounds on artifact behavior
	Dynamic Analysis: Study artifact in use for dynamic qualities (e.g., performance)
3. Experimental	Controlled Experiment: Study artifact in controlled environment for qualities (e.g., usability)
	Simulation: Execute artifact with artificial data
4. Testing	Functional (Black Box) Testing: Execute artifact interfaces to discover failures and identify defects
	Structural (White Box) Testing: Perform coverage testing of some metric (e.g., execution paths) in the artifact implementation
5. Descriptive	Informed Argument: Use information from the knowledge base (e.g., relevant research) to build a convincing argument for the artifact's utility
	Scenarios: Construct detailed scenarios around the artifact to demonstrate its utility

Table 2. Design Evaluation Methods according to Hevner and others [2004].

Design research is an iterative and incremental activity by nature, and it is only natural that the evaluation phase is a central source of feedback for the construction of the artifact [Hevner et al., 2004]. The artifacts entering the evaluation phase might be incomplete, and the evaluation phase helps to assess what to improve in both the design process and the artifact. As an iterative process, the artifact might go through the evaluation several times before it is complete, meeting the requirements set for it.

2.4 Literature review and case study methods

The purpose of a *literature review* is to demonstrate knowledge of academic literature on a specific topic, and the ability to critically evaluate it [The University of Edinburgh, 2021]. Literature reviews have several distinctive approaches, including systematic literature reviews and scoping reviews [CSU, 2022]. These two approaches differ both in methodology and goals, as is shown in Table 3. The systematic review is the more comprehensive and traditional academic review approach and aims to answer a specific research question, but the scoping review approach is useful when the research question is more general.

Systematic review approach	**Scoping review approach**
Uncover the international evidence	Identify the types of available evidence in a given field
Confirm current practice/ address any variation/ identify new practices	Clarify key concepts/ definitions in the literature
Identify and inform areas for future research	Examine how research is conducted on a certain topic or field
Identify and investigate conflicting results	Identify key characteristics or factors related to a concept
Produce statements to guide decision-making	Identify and analyse knowledge gaps

Table 3. Indications and purposes for systematic and scoping approaches according to Munn and others [2018].

Munn and others [2018] describe that the systematic review approach follows a structured, pre-defined process to make sure that the results are reliable and meaningful. Scoping review does not have to follow as strictly defined guidelines, and it can actually work as a precursor to a systematic review [Munn et al. 2018].

A case study is a method of exploring an issue within a real-world context, and it is usually used in business or healthcare research [Oxford Brookes University, 2022]. A case study typically involves case analysis, linking the case to theories established in research, and then creating recommendations based on theories and frameworks [Oxford Brookes University, 2022]. Yin [2009] explains that the process starts with a literature review of the topic and with research question positioning and continues with formal research procedures. Yin [2009] highlights this study method's relevance when the research question "seeks to explain some present circumstance".

Even though case study research method is sometimes questioned by others who perform research with other methods, Yin [2009] believes that there is a need for various research methodologies in social sciences as there is in, for example, in natural sciences. Yin [2009] goes on to explain that often research methodologies are unnecessarily put into a hierarchical order, and he points out that by candidly admitting the benefits and drawbacks of this research technique, it complements the advantages and drawbacks of other research types and methods. Even if case study research is often considered as a "soft" method, it is considerably hard and in order to be successful it requires the researcher to follow strict procedures [Yin, 2009].

3 Research on data strategy

Data strategy is not a new term, but it had gained a lot of attention in recent years, as many organizations have realized the potential business value of having one. What started as information systems research has evolved over the years to include several trends and subcategories, including data strategy research.

This chapter focuses on defining the basic terminology that is used in literature when talking about data strategy, and also finding what kind of research exists in the field, how much, and how recent. This will create a base for chapter 4 that looks further into what literature says about creating a data strategy.

3.1 Data, knowledge, and information

The terms *data*, *information,* and *knowledge* are at the core of information management, information systems, and knowledge management literature, but it is quite easy to be confused by the terminology, as the terms are sometimes used ambiguously. The problem is highlighted in the non-academic literature, where the terms are sometimes used synonymously even though they do not mean the same. In academic literature, there is a widely recognized and accepted model that helps determine the terms and their relations: the data-information-knowledge-wisdom (DIKW) hierarchy, also referred to as the 'Knowledge Hierarchy', the 'Information Hierarchy', and the 'Knowledge Pyramid' [Rowley, 2007]. The hierarchy is pictured in Figure 3.

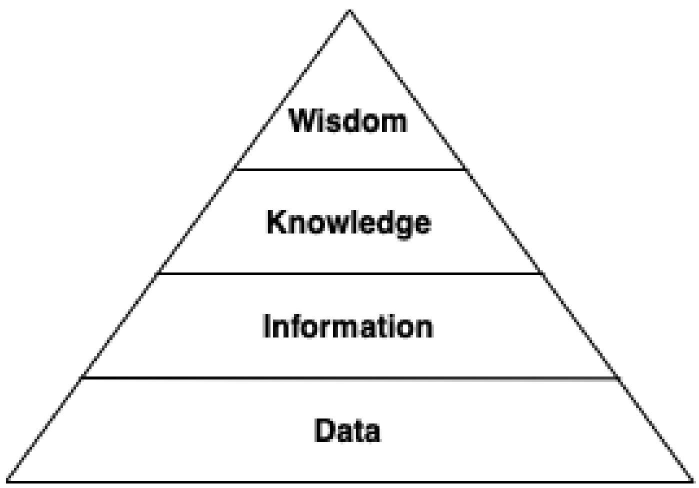

Figure 3. The DIKW hierarchy.

The hierarchy model is used to situate data, information, knowledge, and occasionally wisdom in relation to one another, as well as to recognize and characterize the underlying mechanisms in the conversion of a lower level to a higher level in the hierarchy, for example transforming data into information [Rowley, 2007]. The idea of DIKW hierarchy is not solely related to information systems or knowledge management, but in the context of IS, the model has been widely credited to Russell Ackoff and his 1989 paper titled ‘From data to wisdom’. In the paper, Ackoff [1989] also mentions another level between knowledge and wisdom: understanding. Also in the late 1980s, Zeleny [1987] published his version of the DIKW hierarchy with the top level of enlightenment. In addition to these two versions, several other versions have been published over the years, but the consensus nowadays is that the hierarchy model consists only of the four DIKW elements [Rowley, 2007]. The comparison of Zeleny’s [1987] and Ackoff’s [1989] versions of the DIKW hierarchy is presented in Table 4.

Element	**Zeleny [1987]**	**Ackoff [1989]**
Data	Know nothing	Symbols
Information	Know what	Data that are processed to be useful; provides answers to who, what, where and when questions
Knowledge	Know how	Application of data and information; answers how questions
Understanding		Appreciation of why
Wisdom	Know why	Evaluated understanding
Enlightenment	Attaining the sense of truth, the sense of right and wrong, and having it socially accepted, respected and sanctioned	

Table 4. Rowley’s [2007] comparison of Ackoff’s and Zeleny’s definitions of data, information, knowledge, and wisdom

Despite the differences and the additional levels proposed by Zeleny [1987] and Ackoff [1989], the four well-established elements have significant similarities in both definitions. With these definitions, it is fair to say that data is the original form from which information, knowledge, and wisdom are formed. Data is just symbols and does not provide value in itself. When data is structured, processed, and turned into information, it starts to create value. The information provides answers to factual questions, but does not answer how these facts should be leveraged. Information that is applied into knowledge is able to answer broader questions of how. Still, the last step of the pyramid, wisdom, is to understand why the the data, information and knowledge is needed. All these steps must be understood to leverage data in the best possible way.

3.2 Definitions of data strategy and related terms

To be able to assess the research and literature related to data strategy, there must first be a clear understanding of what the term means. There are also several terms that are closely connected to data strategy that are important to understand in order to understand data strategy as a term, as well as to understand the differences and connections between the terms. As Galliers [2009] mentions, the whole study field of information systems is full of terminology that is used diversely: some terms are used synonymously even though they have variation in the perspectives, and in other cases there are several terms that are used to describe single things. As mentioned in chapter 3.1, for example data, information and knowledge are highly related to each other, but do not essentially mean the same thing. Many of the terms and concepts introduced in this chapter work the same way: they are highly related to each other but do not mean the same thing.

Data strategy as a term is often used quite broadly and loosely, and although there is a consensus on the term in the big picture, a solid, well-defined and generally acknowledged definition of the term is still missing [Fleckenstein & Fellows, 2018]. Gartner's [2022a] definition says that a data strategy is "a highly dynamic process employed to support the acquisition, organization, analysis, and delivery of data in support of business objectives" and Fleckenstein and Fellows [2018] describe data strategy as "the coordinated approach of executing multiple data management domains to help manage revenue, cost, compliance, and risk". These definitions, to name a few, both highlight data strategy as a comprehensive process with the end goal of improving the business.

Data use case is a smaller section within a data strategy. Building a data strategy often starts with the creation of one or more data use cases based on the most pressing business issues [Fleckenstein & Fellows, 2018]. Having several well-thought data use cases helps to prioritize the data strategy, and clarifies the company's strategic decision-making [Marr, 2022].

Data management is the action of managing data: defining data elements, and deciding how they are structured, stored, and moved [Al-Ruithe & Benkhelifa, 2017]. Data management can be seen as a framework for clarifying the organization's data, and guiding the organization towards strategic data management [DalleMule & Davenport, 2017]. Data management is thus a base and a building block for data strategy.

Data governance is defined as the control and planning related to data management on a higher level and is complementary to data management [Al-Ruithe & Benkhelifa, 2017]. According to the Data Management Association (DAMA) International [2009], there are nine subcategories for data governance: data quality management, data architecture management, data development, database operations management, data security management, reference and master data management, data warehousing and business intelligence management, document and content management, and meta data management. Data governance and its subcategories are an essential part of data strategy as components that are making sure data is handled securely.

Information management, as defined by Jaeger and others [2005] is "the entire range of technical, operational, and social functions of a system handling information, affecting the creation, organization, storage, and disposition of information; the access to that information; the behaviors of users of the information; and the information policy environment in which it exists." Galliers [2009] describes that the term information management can refer to the "general field associated with the management issues concerned with information and ICT" but emphasizes that the term can be used in various ways and may not mean the exact same thing to all. This term can however be seen as an umbrella term for management policies related to information systems, and also for organizational behavior. Data strategy can be seen as a section within this umbrella term.

Knowledge management is more business than technical process, and it focuses mostly on organization's intellectual and informational assets. It fosters cooperative and integrative approach to information assets including unwritten information [Gartner, 2022b]. This term may occur in conjunction with data strategy, as data is part of organization's information assets. When integrated into knowledge management, a good data strategy can help in streamlining of knowledge management processes.

Information systems strategy (sometimes just *information strategy*) can be defined as "an organization's shared perspective of the development, use, management and investment in information systems" [Leidner & Milovich, 2014]. Chen and others [2010] note that in an organization, this definition includes elements such as data, IT infrastructure and personnel, and the consept consists of both human activities and technical components. In many ways, information systems strategy and data strategy are overlapping, but data strategy focuses more on achieving competitive advantage through the use of data whereas the information systems strategy focuses on a broader management of internal processes.

Knowledge strategy refers to the integration of knowledge management with a strategic aspect. It aims to create value by bringing knowledge into managerial decision making as a strategic asset [Bratianu & Bolisani, 2015]. This can also be seen as an umbrella term that includes data strategy as a section within it.

3.3 Academic research

Information systems has been a research area for decades now, with the first highly cited papers being from the late 1970's [Iivari, 2015]. The field of strategy research has also existed for decades, changing its form from 'planning' approach of 1960's, 'policy' approach of 1970's to the 'process' approach of 1980' and 'practice' approach of the 1990's and onwards [Whittington, 1996]. Iivari [2015] concluded that the number of highly cited papers in the IT management field increased significantly in the 1990's, and the knowledge management perspective established its place in the field. The first highly cited paper on the subject of data quality was published in 1995 [Iivari, 2015], and this can be seen as the beginning of the era that acknowledges the role of data as an integral part of the business. The link between data, IT and strategy has strengthened in recent decades, and the area of research has widened significantly, and with new technologies the number of subcategories is constantly increasing.

One big category related to data strategy research that has risen in the recent years is the developments of Big Data and its role in business. Grover and others [2018] name data and big data analytics as inimitable, valuable resources, but they need to be included into the long-term business strategy to generate competitive advantage.

In addition to different technologies and methods, the organizational roles in IS strategy development have been a subject of interest in the academic field. Leidner and Milovich [2014] recognize that, for example, the role of top management in the process of IS strategy development has been widely studied in the strategic IS planning research field.

Also, the role of big data in SMEs has emerged in literature in recent years. Wang and Wang [2020] note that SMEs face significant challenges in the adoption of big data usage compared to the larger organizations, due to the limited resources that SMEs have compared to the larger ones. Still, adopting a strategic plan for big data management and use can create a lot of value for SMEs.

Despite the ever-growing number of studies made in the field, there are still some areas that have not yet been researched extensively. Most of the literature in the field of big data analytics, data-driven decision-making, data science and analytics, business intelligence, and so on focuses on how to harness technology and analytics resources to enhance organizational competence and management processes [Medeiros et al. 2020]. Medeiros and others [2020] conclude that there are not many studies made on the strategic value and competitive advantage that the implementation of these technologies and methods in an organization might produce.

Another important area that is still mostly missing in academic research is the role of data strategy in platform firms. There has been very little study connecting data-driven decision-making to business strategy in the context of platform businesses, even though developing a platform data strategy is an important part of platform governance, especially given the significance of data to platform firms [Bhargava et al. 2020].

3.4 Practical literature

As data strategies are essentially built for practical use instead of being just academic theories, it is only natural that there is a lot of literature about the subject outside of academic and peer-reviewed publications. From an academic perspective, this non-academic literature might not be as convincing as the academic one, but from a practical perspective, these non-academic texts might help immensely in the creation of a data strategy. Many of these non-academic texts are written by professionals working in the field of data management, and they have hands-on experience in the field.

In this thesis, to support academic research there are some references to non-academic books that are written by professionals for other professionals. These include books from Bernard Marr [2017], Mike Fleckenstein and Lorraine Fellowes [2018], and Kristina Powers [2020]. These writers may have some academic research background in addition to their professional ones, but most of their work is done outside of academic circles. Still, their work is often cited in academic research and seen as a valuable part of literature related to data strategy. These non-academic references are included in this thesis due to being cited in the academic peer-reviewed texts.

In addition to officially published texts, the internet is full of content about data strategies and data management. It is quite typical that companies that sell strategic services, data handling tools, or something related to the area of data strategy, publish their whitepapers and blog posts about topics related to the area. This is partly a marketing gimmick to show that they are professionals in the topic that they are trying to profit on, but also a way to increase the overall knowledge of the professional field.

Many blog posts and whitepapers, in particular, focus on defining the basic concept of data strategy, as well as laying out the basic idea of why every company should have one. These texts are quite practical, focusing on defining how a data strategy is created, how it is implemented in the organization, and how its benefits can be measured. These blogs and whitepaper texts include companies of all sizes and industries. As they are not academic texts, the veracity of the information is sometimes difficult to prove, and some of the texts are clearly more of the views of individuals than credible research information.

4 Formation of a data strategy

Nowadays firms are actively collecting, developing, and sharing an increasing amount of data in order to explore their potential, boost income, decrease expenses, and manage risks [Medeiros et al., 2020]. In this chapter, a more close look is taken into what literature says about different data strategies, the formation of data strategies, and the organization's overall need for one. First, in section 4.1 the need for a data strategy is established through examples of the benefits of having a data strategy and the disadvantages of not having one. In section 4.2 the core components of a data strategy are identified and categorized. The last section 4.3 looks into the creation process of a data strategy and some recommendations on what to do or not to do during the process.

4.1 Why does a company need a data strategy?

Data is nowadays a valued asset that helps with processing and decision-making rather than being a byproduct of business operations. But with data, quantity does not necessarily correlate with quality. Having tons of data does not automatically mean that actual insights can be derived from it, and even having insights does not generate much value without a strategic plan of what to do with the insights. This is where the role of a data strategy comes to play.

The goal of designing a data strategy is to ensure that all data resources are organized in a way that allows them to be readily utilized, moved, and shared. A data strategy aids in the management and utilization of data as an asset. In order to ensure that data is used effectively and efficiently, data strategy creates a standard set of goals and objectives for all projects. A data strategy develops repeatable methodologies, strategies, and processes for managing, manipulating, and sharing data across the company. [Levy, 2016]

Marr [2017] highlights three core areas in that data has the most value for business: better decision making, improved operations, and the possibility to monetize data. This breakdown emphasizes both the improved development of a product for the end-user and the streamlining of the organization’s internal operations, both of which are central goals when creating a data strategy. These first two areas focus on improving existing products and processes, but monetization is actually about creating new business models, products, and opportunities. Data monetization provides a huge business potential as it enables the selling of refined data and information to third parties, in addition to enabling data as a service (DaaS) business models.

Medeiros and others [2020] concluded in their study that data strategy, regardless of its positioning in the defensive-offensive spectrum, endorses data-driven decision-making and increases the analytical maturity of the organization, ultimately resulting in competitive advantage. Yoshikuni and Albertin [2020] also found in their study that in the bigger picture, a knowledge management strategy has positive effects on business process and firm performance as well as on information strategic alignment and benefits of IT use.

Medeiros and others [2020] also note that in the fast-paced world that we are living in, the organizations that recognize changes and react to them quickly and purposefully will have the greatest competitive advantage. The lack of data strategy might not directly and instantly have a negative impact on processes and overall performance, but companies that are missing a data strategy are bound to be left behind. Conclusively, organizations must develop a data strategy that is based on their strategic positioning and the analytical skills they possess to stay ahead of the game [Medeiros et al., 2020].

4.2 Elements and components of a data strategy

A purposeful action plan to collect, integrate, and then use data to promote an organization's mission and goals is known as a data strategy. According to Hosch [2020], there are several ways to articulate the basic components and approaches to building a data strategy. Levy [2016] suggests that the organizational data strategy can be broken down into five distinctive core actions: identifying, storing, provisioning, integrating, and governing data. Carruthers and Jackson [2018] emphasize the change management perspective when they suggest ways that the chief data officers should create a data strategy.

Hosch [2020] acknowledges the usefulness of the approaches of Levy [2016] and Carruthers and Jackson [2018], but notes that these might not be enough to provide actionable or sectorally suitable information in creating a data strategy. He proceeds to suggest a data strategy framework that consists of seven key elements: data acquisition, data governance, data quality, data access, data literacy and usage, data extraction and reporting, and data analytics [Hosch, 2020]. The framework's descriptions are shown in Table 5.

Element	Description
Data Acquisition	How the organization obtains its data.
Data Governance	How people make decisions and behave with respect to how data will be defined, produced, used, stored, and destroyed.
Data Quality	How data will be maintained to be complete, valid, consistent, timely, and accurate to make it appropriate for a specific use.
Data Access	How authorized individuals can obtain and use data while maintaining privacy and security.
Data Usage & Literacy	How data users understand and use data.
Data Extraction & Reporting	How data will be queried and retrieved from storage and delivered to users.
Data Analytics	How data will be used through dynamic and visual deployment for benchmarking, exploratory and causal analysis, and prediction and forecasting.

Table 5. Key Elements of a Data Strategy according to Hosch [2020].

Data strategies as a whole can be categorized into two distinctive categories: defensive data strategy and offensive data strategy [Medeiros et al., 2020]. DalleMule and Davenport [2017] have established a framework for building a modern, business-oriented data strategy, and this framework addresses two central issues: It assists businesses in clarifying the core purpose of their data and supports them in strategic data management. The framework helps organizations to create suitable strategies by presenting “either or” type of decisions between defensive and offensive objectives.

The two strategies, defensive and offensive, have different business goals and ways to achieve them: the defensive strategy focuses on playing safe and minimizing risks, whereas the offensive strategy focuses on flexible data management. Organizations that are working with highly regulated business environments (such as insurance or healthcare) where data control is critical, are often opting for defensive strategy. Offensive strategies are more typical for organizations that operate in competitive, dynamic and less-regulated environments, such as retail and marketing. The central difference and trade-off between the two strategies is wether to standardize data or keep it as flexible as possible. [DalleMule & Davenport, 2017]. The key differences of the two strategies are shown in Table 6.

	Defense	**Offense**
Key objectives	Ensure data security, privacy, integrity, quality, regulatory compliance, and governance	Improve competitive position and profitability
Core activities	Optimize data extraction, standardization, storage, and access	Optimize data analytics, modeling, visualization, transformation, and enrichment
Data management orientation	Control	Flexibility
Enabling architecture	SSOT (Single source of truth)	MVOTs (Multiple versions of the truth)

Table 6. The elements of data strategy according to DalleMule and Davenport [2017].

It is important to understand that these two strategies are the two ends of a spectrum, and even though strategies usually emphasise one or the other, there there may be features of the other as well. It is also good to understand that regardless of the industry that the organization is functioning in, its position on the spectrum is rarely static [DalleMule & Davenport, 2017].

One might be attempted to think that one strategy would have better results than the other. Especially in the dynamic world that we live in today, it would be easy to assume that the offensive strategy would have better results regarding competitive advantage. This is, however, not automatically the case, and empirical data shows that both strategies have direct, positive impact on gaining competitive advantage [Medeiros et al., 2020]. Having some data strategy that is gradually updated is better than not having one because it is not yet perfect. Data strategy will evolve over time and the positioning in the defensive-offensive spectrum might also change.

4.3 Creating a data strategy

The creation of a data strategy should not start with the idea of "what data is easily available", but it should begin with determining what the business goals of the company are, and how data can help achieving those goals [Marr, 2017]. And as Powers and Weiner [2020] emphasize, the goals are more important than the time to get there: it is better to create a good data strategy slowly than just any data strategy in a hurry.

Data strategy should also not be created just for the sake of it but with thoughtfulness and purpose, fitting into the bigger picture inside the organization. Data strategy should not be independent of existing strategies or practices, but to be built on top of the existing ones and intertwined with them so that it is possible to iterate them all in synchronization. As Fleckenstein and Fellows [2018] put it, the purpose of a business strategy is "to maximize profit, mimimize cost, and manage risk", and they underline that is it of utmost importance to have data strategy and business strategy aligned to be able to address the most critical organizational needs.

Marr [2017] highlights that a data is the most useful when it addresses a specific business need: the data stategy can have quite narrow and specific questions such as how to target a small customer segment. When creating a data strategy, it is possible to start with a specific business need and then extend the strategy from that. Gathering and analysing correct and purposeful data for a specified business need ensures that limited resources are used in the most efficient way. In his book, Marr [2017] actually recommends to his clients and readers this tactic of designing a strategy firstly from a perspective of more specific question and then extending the strategy to a broader context. Fleckenstein and Fellows [2018] also emphasize that data strategies often focus on specific business issues to solve the most urgent problems, and then expand towards a more comprehensive data strategy.

When creating a data strategy, there are many classic strategic frameworks that can be utilized in the process, such as Porter's Five Forces framework, SWOT analysis, or Balanced Scorecard. Like many strategic methods and frameworks, the Balanced Scorecard has received criticism over the years, but it still remains as one of the most popular strategic tools: a survey made by 2GC [2020] suggests that 88 percent of the participating companies use Balanced Scorecard with their strategic planning. The four perspectives that the framework measures are *financial*, *internal business*, *customers*, and *innovation and learning* [Kaplan & Norton, 1992]. These four perspectives are also central when creating a data strategy: data strategy aims at using data to make better, data-driven business decisions as well as creating new financial opprotunities based on data.

Data can be used for improving internal processes and operations, improving customer satisfaction and raising the levels of innovativeness and competence inside the organization.

Hosch [2020] mentions that there are many approaches to data strategy that might be useful, but many of them are only high-level ideas and consepts, not actionable instructions on creating a data strategy. Most academic studies have resulted in these high-level concepts, and the more tangible and detailed instructions and best practices come from professionals working in the field, sharing their knowledge in, for example, blogs or webinars.

The steps to creating a data strategy might vary depending on the company and the overall situation, but there are some steps that are mentioned in several sources. These steps and instructions can be used in the context of a specific business question as well as with a more comprehensive, high-level strategy. Some of these might overlap in time, but the order is usually as follows.

1. *Business strategy and goals*
   This step is about establishing a clear understanding of current business stategy and goals. During this phase it is important to establish what is the problem that the data strategy aims to solve. [Fleckenstein & Fellows, 2018; Maguire, 2019; Marr, 2017; Marr, 2019]

2. *Assessment of current state*
   This step consists of assessing the current data maturity level and existing data management infrastructure. Concluding what data the organization currently has and in what form is also to be established. [Fleckenstein & Fellows, 2018; Maguire, 2019; Marr, 2017]

3. *Proposed future state*
   After creating a comprehensive idea of what the current situation regarding data and its management is, the organization should define its goals regarding the gathering and usage of data. Also the scope of the data strategy initiatives should be determined, and the actions to achieve the goals should be laid out. This phase includes determining what skills are needed inside the organization and how to improve them. The overall vision of how the change management related to the strategic changes is done. [Fleckenstein & Fellows, 2018; Maguire, 2019; Marr, 2017; Marr, 2019]

4. *Implementation roadmap*
   In this phase, the base for the data strategy is already laid out. This step focuses on creating a timeline and setting milestones for the activities that are required in order to achieve the set goals. It is also important that the roles and responsibilities of the participants are clearly stated, and a plan to create a data culture inside the organization is established. [Fleckenstein & Fellows, 2018; Maguire, 2019; Marr, 2017; Marr, 2019]

As any other business strategy, data strategy needs to be reassessed and iterated regularily. This includes the assessment of how well the current goals are met. To determine the successfulness of a strategy, there needs to be some comparable metrics such as key performance indicators (KPIs). The success of a data strategy is not only about metrics, however, but about how well it is implemented into the organization. A successful data strategy has a strong buy-in at every level of the organization, and everyone understand the importance of data in operations and decision making [Marr, 2017]. For example, KPIs are a way to show where the management focuses its attention, but in order for a metric to "be in use" in a company, it has to have wide buy-in and be actively used [Gordon et al., 2018].

In their 2018 study, Gordon and others found out that the companies using "product-related" metrics, such as product volume, time-to-market and unit cost, performed better in the short term than the companies that did not. But in the long term the picture changes: focusing on product-related metrics is actually causing weaker results [Gordon et al., 2018]. Even though the study is not comprehensive and this topic still needs more research, it makes sense that in the short term, product metrics have high value for the company, but in the long run companies should rather focus on overall metrics such as customer satisfaction and loyalty.

4.4 Product launch phase strategy

The success of a newly launched product is a sum of many factors, both before and after the launch. These factors include, among others, the use of external data, competitor intelligence, supply chain intelligence, and external organizational communication.

It is important to gather and leverage external data throughout the product launch, both in the the pre-launch and post-lauch phases. Especially technological knowledge gathered from customers, competitors, suppliers, and other stakeholders is widely considered to have positive impact on the product launch in the form of better products, fitting commercialization plans, and greater adaptability [Schoenherr & Swink, 2015]. Shi and Weber [2022] also note that the age of the organization affects the product launch and activities related to it, such as data gathering. New, early stage organizations are more at risk of failing due to limited resources, and issues related to the legitimacy of the organization.

In addition to data gathering, it is important to consider what data and information about the company and the product is given out, especially pre-launch. External strategic communication can be seen as a dynamic capability, and strategic silence as a way to protect organizations resources. Shi's and Weber's [2022] study confirms that companies that formally communicate about their product pre-launch are at risk of limiting product development flexibility and at risk of exposing too much to competitors. Companies that practice strategic silence in the external communication pre-lauch have more momentum when entering the market, and more often a positive outcome of product launch. It is also noteworthy that the strategy of practicing strategic silence pre-launch but then changing the external communication to highly active post-lauch is highly associated with better product performance. [Shi & Weber, 2022]

5 Case study

The case study of this thesis is done as commissioned by a small IT company located in Finland. The central problem and research question in this thesis is to find out what kind of data collection would be most relevant from the perspective of product development and business goals, and how the collected data could be utilized most effectively. This case study focuses on building a data strategy for a specific business question: what data should be gathered right after the initial product launch, and how should it be used so that all needed improvements are done when entering the next market area a few months after the initial launch. As the product is lauched only after this thesis is released, the thesis provides a plan for data gathering and use, but the effectiveness of the data strategy cannot be established within this thesis.

The company does not yet have an organized data strategy, so in practice the goal is to both answer the specific question and to create some base for a wider data strategy. The goal is that the future business decisions and product development of the company are heavily based on properly collected and structured data (*data-driven development and decision making*).

With the case company, literature may need to be applied in many ways: much of the research and literature on the subject seems to focus on changes in the data strategies of established companied, but in literature it is rare to deal with a newly established company that does not yet have the weight of history. As the literature does not handle startups directly, this thesis makes notions on how the research can and cannot possibly be used in the startup context such as the case company. The basic idea of creating a data strategy, however, is applicable to any company. The creation process of the data strategy base for case company follows the recommended steps established in chapter 4.3.

The creation of a data strategy starts by establishing business strategy and goals. It continues with looking at the current state of the company: by finding out what data is already being collected, and how that data is possibly being utilized. Next, after having both several free-form discussions and unstructured interviews with the management about what the goals and expectations of the company are, and what the management believes the data could be utilized for, the list of goals was created. These goals involve both larger business entities and smaller areas related to product development. After assessing the current situation, business objectives and goals regarding the data strategy, a narrower business question to be answered within this thesis was chosen. When creating the data strategy, a high-level roadmap is created, and the data strategy was planned in a way that it will be easy to use and develop in the future. The strategy created is mainly business-driven, not technology-driven.

The chapter is divided into five subsections. First, we get to know the case study company, its market positioning, and its motivations. In the second section, the business plan and business-related goals are presented. In the third section, we proceed to assess the current situation of the data handling and usage of the company as well as current expectations for the future data strategy. After that, the fourth section opens the plan made for the case company, answering the research question of what data the company should focus on during the Proof of Concept phase of the product launch, and how that data should be used to improve the product and business model before launching in the next market area later this year. The fifth section lays out the plan as a roadmap, and the sixth chapter focuses on how to iterate and improve the plan in the future.

5.1 Case study; a used mobile devices marketplace

The market for used mobile devices has been quickly rising in recent years, with industry research predicting even greater growth in the next years. Shipments of old smartphones are expected to surpass 350 million units by 2024, with the market value of $65 billion [IDC, 2021]. As the flagship models are growingly more expensive, consumers have started turning to second-hand phones. Also the global microchip shortage, caused by many reasons including the Covid-19, has started to affect the production and sales of new smartphones [Browne, 2021]. These trends have created a lot of business prospects for organizations who deal with used mobile phones. The majority of businesses in the industry buy back used devices from clients who are upgrading to a newer model. The business then refurbishes or sells these used devices directly to another customer. There are several refurbishing companies, but the market is currently missing a trusted consumer-to-consumer (C2C) marketplace.

The case study company is a small Finnish IT startup that has been operational for about a year by the time of this thesis. The mission of the company is to revolutionize the trade-in market by providing the easiest and most trusted platform for consumers to sell and purchase second hand mobile devices. The basic idea for the marketplace is that the company ensures a safe sales between two consumers, making sure that the seller receives the money from the sales and the buyer gets the device that they paid for in the expected condition. By the time this thesis is written, the marketplace has not been launched yet. It is expected to be launched briefly after this thesis is published.

In addition to making C2C trades easier and more trustworthy for consumers, the case company has strong ecological values and aspires to make a change in people's consumption habits. There is a great demand for any circular economy solutions to reduce the amount of electronic waste: it is estimated that the total of electronic waste in 2021 was 57.4 million tonnes, a weight that exceeds that of the Great Wall of China [WEEE, 2021]. This ammount is expected to increase by 3% annually, mostly due to limited repair options and short product lifecycles [WEEE, 2021].

With this in mind, the case company aims to also help reduce CO2 emissions. For example, iPhone 12 Pro Max creates approximately 86kg of CO2 emissions, and 82% of that is created during production [Thorne, 2021]. The amount of CO2 has increased with the newer phones, and also the percentual role of manufacturing in the total CO2 emissions of a phone has increased [Thorne, 2021]. By expanding the lifecycle of mobile devices, the trends of increasing e-waste and CO2 emissions can be turned around.

The company has approximately 20 employees located in several cities, mostly across Finland. Currently, the organization has two teams: The business and managerial team, and the research and development (R&D) team. As the company is quite small, the communication inside the organization is highly cross-functional.

5.1.1 Product

The case company will create a one-of-a-kind marketplace for used mobile devices. By delivering advanced and accurate MMR (Make Model Recognition) with counterfeit identification, as well as insured and protected transactions, the marketplace will reduce mistrust between the two parties on direct C2C trades. Additional marketplace value may be added by providing additional services such as insurance, accessories, and financing programs when acquiring a used mobile device.

The process is quite simple but highly trustworthy for the consumers: a consumer wants to sell their device, so they list it to the marketplace. The MMR detects the device and recommends a price for the seller (this recommendation is not binding to the seller). Another consumer, a buyer, sees the listing and wants to buy the device. They buy it with additional accessories and pay the total price. The money is transferred to a customer asset account to wait, and the seller is notified that the device is sold and that they are expected to deliver the device to a checkpoint within a certain time period. Once the device arrives at the checkpoint, it is confirmed by professionals that the device is correct and in the expected condition. After that, the device is packed with accessories and sent to the buyer, and the money from the customer asset account is released to the seller. If there would be an issue with the device (for example the condition is worse than expected), the deal would be canceled, the buyer would get their money back and the device would be returned to the seller. This way both parties can be sure that the deal is right. The company is providing a service to a C2C trade but not being an active participant in the trade. This is a sort of an escrow service similar to ones that are used in many other fields such as banking, or mergers and acquisitions.

To give the greatest performance and most reliable MMR and counterfeit detection, mobile applications will be created natively for both iOS and Android. The visual state will be determined by the user. Artificial Intelligence and device hardware such as cameras might be used in the future to detect the exterior status of the device.

Market value data for used mobile devices will be provided via the Artificial Intelligence (AI) engine (patent pending). The information will be utilized in mobile and online apps to give consumers accurate device market value information and insight. The AI engine gives the users dynamic information about the correct pricing of a certain device. This AI engine has great potential in other aspects of the platform, too, but in this thesis, the focus will be on how it works in the C2C marketplace.

The possibility to have additional accessories and services, such as insurance or financing, in the C2C business is also something new that the platform will provide. Buyers will have the option to buy new accessories, such as chargers or earbuds, with the used device. Also having the financing option in C2C trade is something that has high demand but currently quite limited presentation.

Each sold device in the marketplace will be subject to a commission/service fee. The charge will cover the price of shipping, escrow service, and other necessary fees. In this manner, the buyer will pay a different amount of money than the seller will receive. It is expected that this is not a problem for customers since the platform can still offer better pricing for both sides than in the typical second-hand equipment market, where firms buy devices and resell them to customers.

5.1.2 Market positioning

Currently, the company does not have any major competitors in the specific field of trusted C2C device sales. Some existing marketplaces have a similar business idea, but they operate in other market area such as designer clothing or furniture. There is a wide market for second-hand mobile phones, and the case company is positioning itself into a niche that currently has little to no competition. The positioning concerning the overall competitive situation is good, although there are some companies that are operating in similar market areas such as:

1. Refurbishing companies
2. Marketplaces for refurbished devices
3. Established C2C channels
4. Operators and other businesses with trade-in programs.

The main difference between these business models compared to the model of the case company is that these companies usually offer only B2C or C2B trades. Also, the refurbishment of the device affects the prices: typically, the seller gets a lower price in C2B sales than they would with C2C, but the customer buying the refurbished phone in B2C trade is going to pay more than they would with C2C. Of course, refurbishment may increase the value of the device, but at the end of the day, the refurbishers and marketplaces are the ones benefiting from the pricing system, not the consumers. Another factor that increases the price when purchasing a secondhand device from a business, is that the consumer needs to pay the VAT, whereas direct C2C trading is VAT free.

There are currently some established C2C channels, such as eBay. For the reasons of security and trustworthiness, consumers often prefer B2C sales to C2C sales when buying used devices, even though it is more expensive. There is currently a need for a C2C platform that ensures the trustworthiness of the trades. This is something that the new marketplace aims to provide: a consumer can sell their device anonymously to another consumer while both parties being secured a safe and trustworthy transaction.

Concerning the product, things that clearly distinguish the case company from other businesses in the field are the AI-based pricing engine, and additional services provided. Usually, C2C trades do not enable consumers to have financing, additional accessories, or insurance, but the platform of the case company will provide such services to add value to C2C trades.

5.2 Business strategy and goals

The major objective for the case company is to develop a one-of-a-kind used mobile device marketplace that will become renowned as the "Über" of the used mobile device industry. In the long run, the marketplace could be expanded to include other electronic devices such as Macs, PCs, and other electronic devices with high second-hand market value.

The main short-term business goal of the company is to start the Proof-of-Concept (POC) phase in Italy. The company has planned several channels for consumers through which to use the platform. For the POC phase, the company focuses on providing native applications for iOS and Android. In the future, there are several possible partner models for expansion into new markets, so there are several possible options to provide the platform in different market areas.

After the initial launch and POC phase, the company aims to scale up in Europe, heading first to the Spanish market. The data strategy created within this thesis focuses on the data gathered from POC and applied to the business and product development before entering the Spanish market. The timeline is presented in Figure 4.

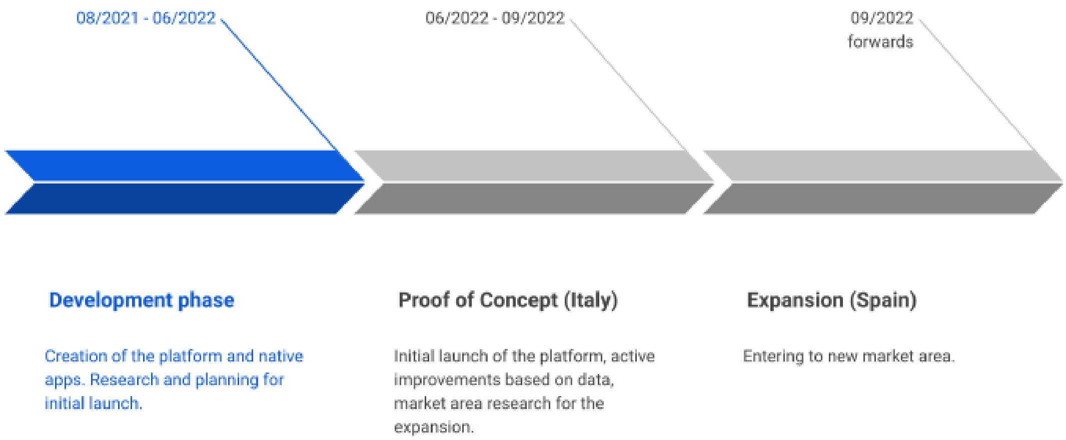

Figure 4. Timeline for POC and expansion.

Italy was chosen as a POC market area due to having a larger potential customer base and shorter logistic distances than Nordic countries, and also due to existing business relations that the founders of the company had. Spain is planned to be the second area mostly due to the same reasons. Once the company has established the platform in the first market areas, the options for the next expansion areas are wider.

As the company is new, launching its first product and there is no existing customer base, it is vital to improve the team's understanding of the customers and how they react to the product. The planned time between the first launch in the POC area and the expansion is quite short, but there are most likely some changes that need to be made based on the data gathered from the POC. The purposeful gathering and usage of POC data is the most pressing topic related to business and product development, so it was natural to choose the issue for the focus area of this thesis.

5.3 Assessment of current state

As mentioned, the product – C2C marketplace for second-hand mobile phones – is not yet launched at the time of this thesis. So currently, the user data is not gathered as there are no users yet. The current state assessment focuses mostly on discussions with the management about future requirements.

At this point, several stakeholders are involved in the POC stage of the business and product.

1. The case company: Finnish company that has developed the product.
2. Joint venture company: Finnish-Italian company that runs operational tasks in the Italy market area.
3. Anchor investor: There is a single anchor investor that has invested in the case company. They are highly interested in the POC stage and keeping a close eye on the process.

There are also other stakeholders that do not have an active role in the product launch and POC stage, such as small investors. The main stakeholders are the ones that are highly interested in the data that the POC provides. Especially the case company and the joint venture company need the POC data to make improvements to the business and product as real-time as possible, before entering the new market area.

5.3.1 Expectations for the future

In January 2022, the first unstructured interview with the management of the company was held about their thoughts and requirements regarding the data strategy and the things that are expected from this thesis project. All interviewees in this initial interview were top management and co-founders of the company. Interviewees are listed in Table 7 with brief information of their current role and field of their professional background.

Interviewee	Current role	Background
1	Chief Executive Officer (CEO)	Technology
2	Chief Financial Officer (CFO)	Financial services
3	Chief Commercial Officer (CCO)	Financial services
4	Chief Operating Officer (COO)	Business
5	Chief Technology Officer (CTO)	Technology

The scope of the discussion was quite wide at that point, and there was no clear idea of what business questions or product development issues are the most pressing ones. These requirements, however, create an idea of what the management thinks are the important goals and factors for data-driven decision-making in the bigger picture. The requirements that were raised in the unstructured interview are listed in Table 8 under the Requirements column.

Requirement	KPI (or other metrics) to measure success	Business or R&D	Relevance
Better user satisfaction	Net Promoter Score (NPS), Customer satisfaction score (CSAT), Customer retention rate (CRR), Churn rate	Both	Yes
More users	Churn rate, Customer aquisition cost (CAC), Conversion rate	Both	Yes
Larger one-time purchases	Average revenue per user (ARPU)	Both	Yes
Customer buys or sells again	Stickiness	Both	No
Recommendations as accurate as possible	How many users buy a recommended device	R&D	No
Prioritizing feature enhancement	Roadmap scoring	R&D	Yes
Helping users in a timely manner	Pain points, Number of support tickets created	R&D	Yes

Targeting of marketing activities	Brand recognition, Conversion rates	Business	Yes
Allocation & optimization of the company's resources (related to the problems, employees in the right positions, etc.)	Resource availability	Business	No
Market-specific effects on operations	Market area research	Both	No
Employee satisfaction	Employee surveys	Business	No
Sales & optimization of add-ons	Revenues	Both	Yes
Pricing machine (how accurate the price is)	Ratio of how many take the proposed price	R&D	Yes
How many searches do you make before you buy, how long does it take to make a decision	This could be measured in the future with analytics and heatmapping	R&D	No
Buyer profiles / target groups	Customer data to create profiles	Both	Yes

Table 8. Requirements raised in the initial discussion in January 2022.

There are dozens of key performance indicators (KPIs) that can be used to analyze and utilize gathered data. Table 8 includes some examples of possibly suitable KPIs to measure the success of each requirement presented in the interview, and an estimation if the requirement is targeting more business or R&D teams. Many of the requirements listed are high-level or long-term requirements, and in the Relevance column, it is assessed if the requirement is relevant to the specific research question for the data strategy created within the thesis, or if the requirement can be focused on in later stages. It is noteworthy that the requirements mentioned in Table 8 are all based on the interviews with the management of the company, and thus the list might be incomplete. The KPIs and other fields are only examples and estimates and should not be considered as something set in stone.

Often the data that is relevant to product development is also relevant for business development, and vice versa. The case company is no different, especially because the product is directed at consumers. In a market where there is a lot of competition, the product must be as perfect for the consumers as possible to be successful business-wise.

As can be expected, the Return on Investment (ROI) KPI was mentioned in the discussion several times. In the big picture, all improvements that are made need to be profitable and improve the business, delivering continuous ROI. The CEO of the company reminded, however, that the data should not be the only thing that is looked at, and data should never be looked at from a too narrow perspective. Looking at just one data section and making decisions based on that might have larger, adverse side effects.

In addition to the discussion with the management, the person that is currently working with data gathering was interviewed to see how he thinks about the current situation, and what he thinks is the most needed and valuable data for the product development, but also for the business.

User data and activity (sales, purchases, returns, additional services) is the data that is really needed to do everything AI-related in the future, and that in itself is certainly our most valuable data.

– AI and data engineer of the case company

The company does not yet have its own customer data, but external data from partners has been used for AI programming, market research, and so on. There has also been a test stage use of different data platforms to find out which would be the best for data management and visualization.

One strategic goal that came up in the discussion with the management was the possibility to monetize data. The process of harnessing data to gain quantifiable economic value is known as data monetization [Gartner, 2022c]. For a platform company such as the case company, there are a lot of possibilities in data monetization. After the product is launched and lots of data starts to pile up, it is possible to use it for more than just the development of the current product and business. With all data privacy laws and requirements kept in mind, it might be possible to, for example, sell some data to third parties, creating a new product out of the gathered data.

5.4 Proposed future state: Data from Proof of Concept

A data strategy can start with a focus on a single, specific business problem and a data use case, as previously stated. This helps to prioritize the most pressing issue while also considering a more comprehensive data strategy for the future. For the case company, one of the most pressing business issues is to gather and use data from POC to improve the product and business models before expanding to new market areas.

The actions towards reaching this goal are:

1. Defining the metrics
    a. What do we want to achieve?
    b. How do we measure, and what are the right metrics?
2. Tracking the progress compared to the set metrics
    a. Are we reaching the set goals?
    b. If not, how to improve the situation?
3. Collecting customer feedback and analyzing it with
    a. session replays: heat-mapping, recording tools, etc.
    b. data from customer service: customer service tickets: number of tickets, categories of issues, resolve times, etc.
    c. customer surveys, focus groups, online reviews, etc.
4. Implementing data-based decisions into action

Data-driven value creation is at its highest in an organization with high analytical capabilities [Madhala et al., 2020]. These include all analytics skills from technical analytics to business analytics and are the most valuable organizational capabilities to have. The actions mentioned above need to be implemented throughout the organization, and all stakeholders should understand the relevance of analytics and data to organizational decision-making. From the change management perspective, a startup with a small group of like-minded employees and a highly flexible culture (both technically and interpersonally), is quite easy as changes are often taken in with enthusiasm and interest instead of resistance and suspicion.

A good way to concretely track and use the gathered data is to visualize the data in a business intelligence (BI) tool. A tool enables the creation of visualizations, statistical charts, and dashboards for all parts of the organization, and to easily gain insights into the product and users. There can be several different dashboards for tracking different things or aimed at different users. For the case company, it starts with two separate dashboards: one for management, mostly for business development, and one for the technical team to use for product development. The dashboard items are inspected more closely in chapter 5.4.1.

Prior to the launch, it is important to plan and implement tracking into all key parts of the funnel. This means that for the case company, it is not just tracking the customers: the checkpoint, the logistics, and the customer service, for example, need to be tracked as well. Even if the product seems perfect in all other ways, if there are issues with these areas, the product is set for failure. These parts of the funnel need to be tracked from day one and if issues emerge, they need to be fastly reacted to.

Application analytics is a big source of valuable data, especially for the development team to gain knowledge of things such as user actions and pain points. The company has plans to use analytics tools in the applications, but the original idea of using Google Analytics is on hold, as Google has had issues with GDPR compliance [Lubowica, 2022]. There is still a possibility that Google Analytics is used, especially if Google releases a new, GDPR-compliant version, but the company wants to explore other options as well. For a platform company handling lots of personal data, GDPR compliance is of the utmost importance.

Overall, the biggest challenges to data gathering and usage for the case company could be

1. *Limited data*

   This is most likely the biggest challenge in the beginning: how to gather enough data for it to be useful. Data is not very useful if the amount is low, and a startup that does not have any data from the past is fully dependent on the gathered data. On the other hand, it is positive that there is no irrelevant old data mixing up with the newly gathered data.

2. *Irrelevant or low-quality data*

   Data in itself is not valuable if it cannot be used to make conclusions. This could result from gathering the wrong kind of data or having incomplete data. This risk is highlighted in the case company, as the initial decisions to gather data are based on assumptions rather than knowledge of what is needed. The end result of gathering irrelevant or low-quality data is, at the very least, that the data cannot be utilized, but at worst, collecting the wrong or poor quality data can lead to wrong conclusions and result in major problems with product development and overall business.

3. *Taking the full potential into use*

   Data as itself is not valuable, it has to be turned into information to be relevant. In the beginning, there is the challenge of finding the best ways to use data efficiently in both business and product development. This is an iterative process that cannot be fully designed before the product launch.

5.4.1 Dashboards

Dashboards are created based on the gathered data and used to visualize trends and information. Different teams need to focus on different metrics, so separate dashboards are created for each team. In the future, there will be several other dashboards for different stakeholders, but in the beginning, the most important dashboards are the ones made for the management team and technical (R&D) team. The most important metrics and how to use them will be laid out next, Table 9 focusing on management team requirements and Table 10 focusing on technical team requirements.

Measured metric	How to utilize data
Number of users	This helps to track the number of single users and make sure that the trend is growing. If there is a significant drop in the numbers, do a more close inspection on why this is happening.
Number of listed devices	This helps to track the number of listed devices and make sure that the trend is growing. If there is a significant drop in the numbers, do a more close inspection on why this is happening.
Number of sold devices	This helps to track the number of sold devices and make sure that the trend is growing. If there is a significant drop in the numbers, do a more close inspection on why this is happening.
Ratio of listed and sold devices	This ratio helps to understand how many of the listed devices are actually moving through the platform. If there are lots of devices listed but not sold, the buyers need to be activated. If there are not much listing but most of them are sold, the sellers need to be activated. These actions can be changed dynamically based on data.
Total sales and revenue	This helps to track the total business and make decisions related to the business model.
Average (device) selling price	What is the worth of an average device sold through the platform. This can be used in marketing to focus more on interesting customer segments (for example sellers with more expensive, higher end devices).
Average shopping cart price	What is the worth of an average transaction in total, so how much an average buyer is willing to use.
Ratio of average selling price and total shopping cart price	The ratio of the device price and additional services/accessories price. This helps to improve the business model and focus marketing interests (for example if not many are taking additional services, marketing could focus more on that).

Table 9. Dashboard for the management team

The management team dashboard obviously focuses on business related number such as sales and revenue, but the goal is also to understand consumer behavior better through the data. The management team wants to make data-driven decisions in the future and have data to make dynamic business decisions. In addition to the visuals and analytics in the dashboard, the management team has asked for a weekly summary of the metrics to be sent via email.

Measured metric	How to utilize data
Average listing price and average selling price	The ratio indicates how close to truth the listing prices are. If there are big differences in listed and sold prices, it might indicate that the platform might need some redesigning.
Pricing engine suggested price compared to asking price and sold price	This ratio indicates how accurate the price engine suggestions are. If the suggestions are significantly higher than sold prices, it might indicate that the price engine needs to be adjusted.
Number of listings accepting offers	How many users are trying for a higher price than they would expect to get, and on the other hand how many users are unsure of the price they can ask for a device. This also helps to assess if the recommended asking prices are valid or if the recommendations need adjustments.
Number of accepted offers and the ratio compared to made offers	This partially indicates the same things as the previous one, but could also help to understand if users are, for example, actually accepting the offers or just using the offer possibility as a trick to peak interest in the buyers.
Accepted offer prices compared to asking prices	This could indicate that users are unsure of the asking price, or unsure if the device would be sold with asking price.
List of the most sold device models (inc. average selling price)	These models are in the category "hottest finds" and can be recommended to users when the personalized recommendations are not available. These can be used to gain insights and make decisions on which models need to be supported in the future.
Number of registered users buying or listing devices	This indicates how interesting the service is to users. If most users register but never make purchase or list their devices, the platform fails to communicate the value to customers, and that can be seen as a usability issue that needs to be tackled.
Success rate for device detection	Especially in the beginning, it is important to detect issues with device detection early on. This is one of the central features in the platform, and could cause major issues is there are lots of misidentifications.

Condition grading accuracy	Percentage of cancellations due to condition (out of total sales). Way to determine the margin of error, and assessment of the difference in perspectives between consumer and checkpoint employee. High cancellation rate might indicate that there is an issue between condition assessment instructions for users and the condition standards at the checkpoint. This metric is very important in the beginning to detect usability or standardization issues early on.

Table 10. Dashboard for technical team

Technical team wants to use the data to make sure that the product works as intended, but also to gain insight on customer expectations, so the improvements are as timely and relevant as possible. Product owner and the development team will use the gathered data to help determine requirements and place them on the product roadmap.

5.5 Strategy roadmap for Q2/2022 – Q3/2022

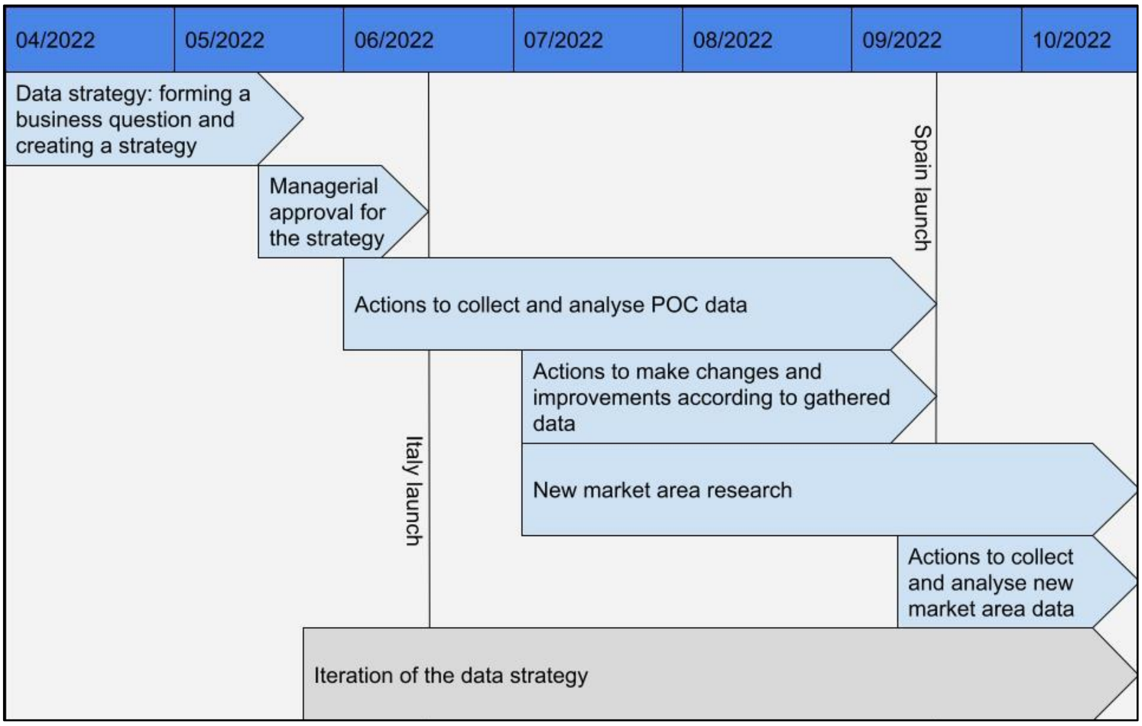

Figure 5. High-level roadmap for POC phase data strategy

The first step is to create a data strategy based on a single business issue: in this case, the creation of the strategy is done within this thesis and the issue is the research question. The managerial approval for the data strategy is required to put the strategy into action, and this happens when the thesis is ready to be released. Once the managers give the initial approval for the strategy – meaning that they see it to be fitting for the issue and actionable - it also means that the iteration of the data strategy starts. Data strategy needs to be continuously iterated, especially in the case of launching the first product: it is not possible to create a one-size-fits-all solution, and the iteration is really possible only after the data starts to arrive. Once the initial approval for the planned actions is received, it is time to start the actions. The preparations for the ways to show, analyze and use the data need to be in place before the launch, and some weeks need to be reserved for that.

During the first days and weeks after the launch, the R&D team needs to focus on addressing possible bugs and issues fast to avoid roadblocks to user adoption. User requirements need to be collected, as there might be a need for use cases that no one from the company and stakeholders had thought of during the development phase. Marketing and business teams need to react swiftly if customer acquisition and retention are not as expected.

After some weeks, once there is some actual data from the POC, it is possible to start listing, prioritizing, and making the first data-based improvements to the product and the business model. At the same time, it is important to understand which requirements are market area-specific and which are universal, so it might be a good idea to compare and combine the new market area research with the data from the POC.

When the launch of the new market area approaches, it is important to be ready for data gathering and use insights from the POC launch. New market area data needs to be accessed on its own and also compared to the previous data from the other market area to see similarities and differences in the market areas. This insight might provide some valuable information in terms of product and business development.

5.6 Future iteration of the data strategy

The current data strategy created within this thesis focuses on a single business issue. This data strategy is actually more of a data use case and a starting point for a more comprehensive data strategy. In the future, several factors need to be considered that are missing from the current one.

Most importantly, the current data strategy created within the thesis lacks the technical side, and that needs to be considered when iterating the strategy. In addition to combining the current data strategy issue with technical solutions, it is important to include the technical side of all future data strategy issues. For example, different APIs, both internal and external, and their roles in data strategies need to be laid out, and also stakeholders and data coming from them need to be counted in the strategies. The security and privacy of the user’s data need to be confirmed, and all actions need to be GDPR compliant.

In the future, both this existing strategy and other strategies should have more KPIs and other metrics defined. At this point, when there is no data yet, the scope of this thesis was to focus on the gathering of relevant data from the perspective of what the team needs to know right away. Once there is more data, the iteration needs to focus more on selecting suitable KPIs. In the future, it is also important to use existing internal data and external data to define the requirements of a data use case and data strategy more closely: currently, the requirements are based on estimates, professional knowledge of the management team, and literature, but not on existing data as there is none.

For a company in the first stages of its lifespan, the position of the future data strategy on the defensive-offensive scale should start from the defensive side. The position is never static and is often a combination of both, but a good offensive strategy cannot be built if the foundation – objectives, and activities related to defensive strategy – is not in place. For the case company, starting from having the emphasis on the defensive strategy and then actively moving toward the offensive side might be the best option. The customer-centric focus related to offensive strategy must be in place from the start, but the emphasis should move from defensive to offensive when the company establishes the basic foundations for data strategy.

The future data strategies also need to focus on several other factors that are not yet taken into account in this issue due to their nature as a short-term, pre-launch strategy. The future strategies need to be more scalable to work once the company and the amount of data grows. The company has plans for several business models in the future including C2C, C2B, B2C, and B2B, and in the future strategies, all these need to be considered. Joint ventures and anchor investors are also factors that might influence the formation of data strategies in the future. Another goal in the future is to bring the sustainability factor into all parts of the business, and even though it might not directly affect a data strategy, it certainly will have comprehensive spillover into all parts of the business.

6 Discussion

The research question was set to find out what is relevant data to gather and use right after a C2C platform launch and to produce an actionable, high-level plan for data strategy creation based on the findings. In practice, the created case study artifact is more of a data use case plan than a proper data strategy plan, but with the knowledge of the literature review and the acknowledgments of missing parts, it can be expanded into a data strategy plan. The data use case plan, however, responds to the research question of data relevance and usage, and it can be seen that the case study had a favorable outcome.

The literature review of this thesis, as well as requirements set by the management in the case study interviews, set the frame for what data is seen as valuable. Chapter 5.4 and its sub-chapter mapped out what seems to be the most relevant data for business and product development right after launch and before entering a new market. Especially in chapter 5.4.1, it was described how the collected data could be interpreted and utilized. With these results, the research question and its sub-questions were largely answered. Still, there is a need for improvements and iterations of the created data strategy or data use case plan in the near future: especially when the product is launched, the plan needs to be re-evaluated, iterated, and expanded.

The thesis was done with design science research methods and guidelines, with the goal of producing a design artifact. The goal was mostly met with the creation of an actionable, high-level data use case plan that is a part of a data strategy plan. The use case results consist of actionable plans to create dashboards for both management and technical teams, as well as a roadmap and plans for future iterations.

As the case study answers a business issue set by the management of the case company, it proves the artifact's problem relevance. The literature review done prior to the case study can be seen as a search process, and this thesis communicates the findings to both the case company and anyone interested in the topic. The utility of the designed data use case was confirmed by the case company, but the efficacy remains to be seen due to the nature of the case study.

On a larger scale than just the case study, the findings of the literature review can be somewhat applied in any organization where it is desired to build a data strategy. The literature review gives a quick but concrete idea of what one should do when starting to build a data strategy, such as the steps of creating it.

6.1 Limitations and future research

This thesis was conducted mostly following the design research guidelines, but due to the nature and timing of the case study, the number of research cycles, for example, was very low. Overall, there were some limitations to this research that were known beforehand, and some that were revealed during the writing of the thesis. Some limitations are related to the case study, others to literature and previous research.

Concerning the case study, it was already known to the case company before the thesis project that the time window for writing the thesis is very limited. This means that the iteration cycles done within the project were minimal, and the case company will continue the iterations after the thesis is released. The assignment of the project was quite vague at the beginning, and it was narrowed down a lot during the process. Still, the scope was quite large and it caused the case study to be done on a very high level. It also caused some parts, for example, the technical side, to be purposefully left out of the scope.

Even though the research and the case study were done within Hevner and others [2004] design science research guidelines, the research cycles and the evaluation of the design artifact are not as extensive as the guidelines suggest. This was due to the nature of the case study: the iterations can happen only after the product launch, and that happens after the release of this thesis.

Some parts of the information gathered in the literature review was referring to data strategy from a wider and higher-level perspective, and not all parts of the literature review were thus used directly in the case study. Still, the literature review was relevant in terms of future iterations of the data strategy: it can be used as an information kit for employees that are less familiar with data strategy as a concept, and for anyone interested in the topic.

With literature, it was expected before the thesis project that data strategy as a concept is widely researched, and the expectations were right: data strategy is handled a lot both in academic and practical literature. Still, most of the literature focuses on established organizations and larger companies. The most significant limitations in the literature were that SMEs, especially startups, are not very present in the current literature, even though it is often mentioned that data strategy concerns companies of all sizes and ages. Also, the point of view of platform companies is not widely researched, although the role of platforms has increased rapidly over the last few years.

Another noteworthy finding was that it was significantly hard to find literature considering data strategy around the time of a product launch: there is lots of literature about many other product launch related themes such as supply chain management or marketing strategies, but the data perspective does not seem to be well researched in the product launch context. Product launch strategies as a whole have been considered in literature over the last 20-30 years. But over the recent years, with emphasis on the data gathering and usage around the product launch, only some blog texts related to single companies' experiences around the subject were found.

These findings are not a result of a systematic literature review, but some conclusions could still be made. Based on the literature review of scoping style, it seems that there is a need for future research on data strategy from the perspective of SMEs, startups, and platform businesses. The role of data strategy in the different life cycle stages of a product is also a topic from which there is not as much relevant literature as one would hope, and it might be beneficial to make more research and research-based recommendations on the subject.

7 Conclusions

This thesis provides a scoping literature review of data strategy, with a special interest in the perspective of building a data strategy for an organization. The results of the literature review show that the field is somewhat heterogeneous between academic and practical literature. It is also shown that there is room for more research in the field, especially on specific points of view such as having data strategy in an SME or platform business.

The knowledge gathered within the literature review was applied to a case study, and a data use case plan was created by applying the knowledge from the literature review. The case study included recommendations on future iterations to use when creating a more comprehensive data strategy and some limitations of the current data use case plan were also mentioned.

The thesis provides basic knowledge of data strategy theory for professionals, as well as some concrete recommendations on how to create a data strategy for an organization. The case study is an example of creating a data use case and data strategy plan for an SME in the field of the platform business.

8 References

2GC. 2020. Balanced Scorecard Usage Survey 2020. Retrieved from: *https://2gc.eu/media/resource_files_survey_reports/2020_Survey_Report-compressed.pdf* (Accessed: 23.4.2022)

Ackoff, R. 1989. From data to wisdom. *Journal of Applied Systems Analysis*, 16, 3-9.

Al-Ruithe, M. & Benkhelifa, E. 2017. Analysis and Classification of Barriers and Critical Success Factors for Implementing a Cloud Data Governance Strategy. *Procedia Computer Science*, 113, 223-232. DOI: 10.1016/j.procs.2017.08.352

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*; New York 31, 4, 323-334. DOI: 10.1007/s11002-020-09539-3

Bratinau, C. & Bolisani, E. 2015. Knowledge Strategy: An Integrated Approach for Managing Uncertainty. *The 16th European Conference on Knowledge Management, University of Udine, Italy, 3-4 September 2015, Proceedings*, 169-177.

Browne, R. 2021. The global chip shortage is starting to hit the smartphone industry. Retrieved from: *https://www.cnbc.com/2021/07/29/the-global-chip-shortage-is-starting-to-hit-the-smartphone-industry.html* (Accessed: 23.4.2022)

Carruthers, C. & Jackson, P. 2018. *The chief data officer's playbook*. Facet.

Chen, D., Mocker, M., Preston, D. & Teubner, A. 2010. Information Systems Strategy: Reconceptualization, Measurement, and Implications. *MIS Quarterly*, 34, 2, 233-259.

CSU. 2022. Literature Review: Types of literature reviews. Charles Sturt University. Retrieved from: *https://libguides.csu.edu.au/review/Types* (Accessed: 20.4.2022)

Confessore, N. 2018. Cambridge Analytica and Facebook: The Scandal and the Fallout So Far. New York Times. Retrieved from: *https://www.nytimes.com/2018/04/04/us/politics/cambridge-analytica-scandal-fallout.html* (Accessed: 15.4.2022)

DalleMule, L. & Davenport, T. 2017. What's your data strategy? *Harvard Business Review,* 95, 3, 112-122.

Dama International. 2009. *DAMA-DMBOK: Data Management Body of Knowledge*. Technics Publications, LLC.

Fleckenstein, M. & Fellows, L. 2018. *Modern Data Strategy*. Springer International Publishing. DOI: https://doi.org/10.1007/978-3-319-68993-7

Galliers, R. D. 2009. Conceptual Developments in Information Systems Strategy – Reflections on Information Systems Strategizing. In: Dorothy E. Leidner, and Robert D. Galliers (eds.), *Strategic Information Management: Challenges and Strategies in Managing Information Systems*, Chapter 1.

Gartner. 2022a. Data Strategy. Gartner Glossary. Retrieved from: *https://www.gartner.com/en/information-technology/glossary/data-strategy* (Accessed: 26.3.2022)

Gartner. 2022b. Knowledge Management (KM). Gartner Glossary. Retrieved from: *https://www.gartner.com/en/information-technology/glossary/km-knowledge-management* (Accessed: 26.3.2022)

Gartner. 2022c. Data Monetization. Gartner Glossary. Retrieved from: *https://www.gartner.com/en/information-technology/glossary/data-monetization* (Accessed: 26.3.2022)

Gordon, M., Kowski, M. & Smits, S. 2018. Taking the measure on product development. McKinsey & Company. Retrieved from: *https://www.mckinsey.com/business-functions/operations/our-insights/taking-the-measure-of-product-development* (Accessed: 15.4.2022)

Grover, V., Chiang, R. H. L., Liang, T. P. & Zhang, D. 2018. Creating strategic business value from big data analytics: a research framework, *Journal of Management Information Systems*, 35, 2, 388-423.

Hevner, A. R., March, S. T., Park, J., & Ram, S. 2004. Design Science in Information Systems Research. *MIS Quarterly*, 28, 1, 75–105. DOI: 10.2307/25148625

Hevner, A. 2007. A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, 19, 2, Article 4.

Hosch, B. 2020. Key Elements of a Data Strategy. In Powers, K. (eds.). *Data Strategy in Colleges and Universities: From Understanding to Implementation*, 14-36. Routledge.

IDC. 2021. IDC Forecasts Worldwide Market for Used Smartphones to Reach 351.6 Million Units with a Market Value of $65 Billion in 2024. Retrieved from: *https://www.idc.com/getdoc.jsp?containerId=prUS47258521* (Accessed: 23.4.2022)

Iivari, J. 2015. Making Sense of the History of Information Systems Research 1975-1999: A View of Highly Cited Papers. *Communications of the Association for Information Systems*, 36, 25. DOI: 10.17705/1CAIS.03625

Jaeger, P., Thompson, K. & McClure, C. 2005. Information Management. *Encyclopedia of Social Measurement*, 277-282. DOI: 10.1016/B0-12-369398-5/00531-4

Kaplan, R. & Norton, D. 1992. The Balanced Scorecard—Measures that Drive Performance. *Harvard Business Review*, January–February 1992.

Leidner, D. & Milovich, M. Jr. 2014. Middle Management and Information Systems Strategy: The Role of Awareness and Involvement. *2014 47th Hawaii International Conference on System Sciences*, 4396-4405. DOI: 10.1109/HICSS.2014.542

Levy, E. 2016. The 5 Essential Components of a Data Strategy. SAS. Retrieved from: *https://www.sas.com/content/dam/SAS/en_au/doc/whitepaper1/five-essential-components-data-strategy.pdf* (Accessed: 15.4.2022)

Lubowica, K. 2022. Is Google Analytics GDPR-compliant? Retrieved from: *https://piwik.pro/blog/is-google-analytics-gdpr-compliant/* (Accessed: 23.4.2022)

Madhala, P., Li, H. & Helander, N. 2020. Organizational Capabilities in Data-driven Value Creation: A Literature Review. *Proceedings of the 12th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, 3, 108-116. DOI: 10.5220/0010175601080116

Maguire, J. 2019. Creating a Data Strategy. Gartner Research. Retrieved from: *https://www.gartner.com/en/documents/3975595* (Accessed: 15.4.2022)

Marr, B. 2017. *Data Strategy : How to Profit From a World of Big Data, Analytics and the Internet of Things*. 1st Edition. Kogan Page.

Marr, B. 2019. How Do You Develop A Data Strategy? Here're 6 Simple Steps That Will Help. Retrieved from: *https://bernardmarr.com/how-do-you-develop-a-data-strategy-herere-6-simple-steps-that-will-help/* (Accessed: 3.4.2022)

Marr, B. 2022. How To Define A Data Use Case – With Handy Template. Retrieved from: *https://bernardmarr.com/how-to-define-a-data-use-case-with-handy-template/* (Accessed: 3.4.2022)

Medeiros, M., Maçada, A. & Freitas Jr, J. 2020. The effect of data strategy on competitive advantage, *The Bottom Line,* 33, 2, 201-216. DOI: 10.1108/BL-12-2019-0131

Munn, Z., Peters, M.D.J., Stern, C., Tufanaru, C., McArthur, A. & Aromataris, E. 2018. Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. *BMC Med Res Methodol*, 18, 143. DOI: 10.1186/s12874-018-0611-x

Orlikowski, W. & Iacono, C. 2001. Desperately Seeking The "IT" In IT Research - a Call to Theorizing the IT Artifact, *Information Systems Research,* 12, 2, 121-134.

Oxford Brookes University. 2022. Case studies. Oxford Brookes University. Retrieved from: *https://www.brookes.ac.uk/students/academic-development/online-resources/case-studies/* (Accessed: 26.3.2022)

Powers, K. 2020. *Data Strategy in Colleges and Universities: From Understanding to Implementation*. Routledge.

Powers, K. & Weiner, S. 2020. The Value of Creating a Data Strategy. In Powers, K. (eds.). *Data Strategy in Colleges and Universities: From Understanding to Implementation*, 3-13. Routledge.

Ranjan, J. & Foropon, C. 2021. Big Data Analytics in Building the Competitive Intelligence of Organizations. *International Journal of Information Management*, 56. DOI: 10.1016/j.ijinfomgt.2020.102231

Rowley, J. 2007. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33, 2, 163–180. DOI: 10.1177/0165551506070706

Schoenherr, T. & Swink, M. 2015. The Roles of Supply Chain Intelligence and Adaptability in New Product Launch Success. *Decision Sciences,* 46, 5, 901-936.

Shi, W. & Weber, M. 2022. Communication Strategies of Entrepreneurial Organizations in Mobile Apps Industry: Hidden Communication Prior to Product Launch. *International Journal of Knowledge Management*, 18, 1, 1-15.

Tabesh, P., Mousavidin, E. & Hasani, S. 2019. Implementing big data strategies: A managerial perspective. *Business Horizons*, 62, 3, 347-358. DOI: 10.1016/j.bushor.2019.02.001

The University of Edinburgh. 2021. Literature review. Retrieved from: *https://www.ed.ac.uk/institute-academic-development/study-hub/learning-resources/literature-review* (Accessed: 15.3.2022)

Thorne, M. 2021. The carbon footprint of your phone – and how you can reduce it. Retrieved from: *https://reboxed.co/blogs/outsidethebox/the-carbon-footprint-of-your-phone-and-how-you-can-reduce-it* (Accessed: 23.4.2022)

Wade, M. & Hulland, J. 2004. Review: The Resource-Based View and Information Systems Research: Review, Extension, and Suggestions for Future Research. *MIS Quarterly*, 28, 1, 107–142. DOI: 10.2307/25148626

Wang, S. & Wang, H. 2020. Big Data for Small and Medium-Sized Enterprises (SME): a Knowledge Management Model. *Journal of Knowledge Management,* 24, 4, 881–897.

WEEE. 2021. International E-Waste Day: 57.4M Tonnes Expected in 2021. WEEEForum. Retrieved from: *https://weee-forum.org/ws_news/international-e-waste-day-2021/* (Accessed: 5.5.2022)

Whittington, R. 1996. Strategy as practice. *Long Range Planning*, 29, 5, 731-735. DOI: 10.1016/0024-6301[96]00068-4.

Yin, R. K. 2009. *Case Study Research: Design and Methods.* SAGE Publishing.

Yoshikuni, A. & Albertin, A. 2020. Leveraging firm performance through information technology strategic alignment and knowledge management strategy: An empirical study of IT-business value. *International Journal of Research -Granhaalayah*, 8, 10, 304-318. DOI: 10.29121/granthaalayah.v8.i10.2020.2088

Zeleny, M. 1987. Management support systems: towards integrated knowledge management. *Human Systems Management,* 7, 1, 59–70.