CHALMERS
UNIVERSITY OF TECHNOLOGY

UNIVERSITY OF GOTHENBURG

Exploring Data Monetization to Customers Outside the Core Business

A Process Model from Ideas to Scalable MVPs in the Embedded Systems Domain

Master's thesis in Computer science and engineering

ROMEO RADEVSKI, DANIEL SANDS

Department of Computer Science and Engineering
CHALMERS UNIVERSITY OF TECHNOLOGY
UNIVERSITY OF GOTHENBURG
Gothenburg, Sweden 2021

MASTER'S THESIS 2021

Exploring Data Monetization to Customers Outside the Core Business

A Process Model from Ideas to Scalable MVPs in the Embedded Systems Domain

ROMEO RADEVSKI, DANIEL SANDS

UNIVERSITY OF GOTHENBURG

CHALMERS UNIVERSITY OF TECHNOLOGY

Department of Computer Science and Engineering
Chalmers University of Technology
University of Gothenburg
Gothenburg, Sweden 2021

Exploring Data Monetization to Customers Outside the Core Business
A Process Model from Ideas to Scalable MVPs in the Embedded Systems Domain
ROMEO RADEVSKI, DANIEL SANDS

Supervisor: Jan Bosch, Department of Computer Science and Engineering, Chalmers Universtiy of Technology
Co-supervisor: Helena Holmström Olsson, Department of Computer Science and Media Technology, Malmö University
Examiner: Christian Berger, Department of Computer Science and Engineering, Chalmers University of Technology

Master's Thesis 2021
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg
SE-412 96 Gothenburg
Telephone +46 31 772 1000

Typeset in LaTeX
Gothenburg, Sweden 2021

Exploring Data Monetization to Customers Outside the Core Business
A Process Model from Ideas to Scalable MVPs in the Embedded Systems Domain
ROMEO RADEVSKI, DANIEL SANDS
Department of Computer Science and Engineering
Chalmers University of Technology and University of Gothenburg

Abstract

Data is playing an increasingly important role in the embedded systems domain. Embedded systems companies are collecting vast amounts of data that hold significant value due to its uniqueness and irreplicability. This data is not being exploited to its full potential value. Currently, embedded systems companies are exploring how their data can be of value to their core business and current customers. However, the data can also be of value to stakeholders outside the boundaries of the core business. This study set out to explore how embedded systems companies can monetize their data outside of their core business, to new, secondary customers. Semi-structured interviews, as part of a multiple case study, showed that companies within the domain have an interest in monetizing data to secondary customers, but lack a clear way of finding and developing such business opportunities. The Data to Secondary Customers Exploration Model (DSCEM), developed in this study, aims to guide embedded systems companies in finding data monetization ideas worth pursuing with full-scale solutions. The DSCEM was validated through three means: validation interviews with practitioners, simulating use cases, and applying it to two real use cases in a workshop.

Keywords: Data, data monetization, data exploitation, embedded systems, agile, lean, multiple case study, data-as-an-asset, data-as-a-service, secondary customer.

Acknowledgements

First of all, we would like to thank our supervisors Jan Bosch and Helena Holmström Olsson for their continuous support, optimism, and sharing their network of industry professionals with us.

We would also like to thank all of the companies and participants who have taken part in the study. Without them, this thesis would not be possible.

Romeo Radevski, Daniel Sands, Gothenburg, July 2021

Contents

List of Figures

List of Tables

Glossary

Big data Data that is so large, fast or complex that it's difficult or impossible to process using traditional methods.

Cloud computing Delivery of different services (databases, software, analytics, etc.) over the internet.

Data as a service Cloud based data management strategy to leverage data as an asset.

Data monetization The act of exchanging information-based offerings for legal tender or something of perceived equivalent value.

Embedded system Computer systems embedded into either mechanical or electrical systems in which the computer system has dedicated functions.

GDPR General Data Protection Regulation.

Internet of Things (IoT) Physical objects embedded with technology to allow for connectivity.

Primary customer Customer of the core business and is essential for the business to exist.

Proprietary data Data that is owned and controlled by an organization, typically generated internally.

Secondary customer Customer outside the core business, who finds value in something that is generated by the company.

1
Introduction

The embedded systems domain is, like most other industries, undergoing a digital transformation. The increasing amount of smart and connected products is redefining the industry, radically changing all functions of organizations, and making companies fundamentally rethink their strategies [1]. However, digital transformation within the embedded systems domain has proved to be challenging due to hardware dependencies, safety-critical functionality, and strict regulations [2]. This has led to a slower transformation compared to other industries and many challenges remain to be solved. Previously, embedded systems companies focused on mechanics and electronics, but these functions are, to a growing extent, becoming commoditized [3]. As more and more new digital technologies have emerged, they have increasingly become complements to embedded systems companies' products as an opportunity to differentiate, gain competitive advantages, and generate new revenues [4]. Therefore, while mechanics and electronics in many cases remain the backbone of embedded systems offerings, it's software, data, and artificial intelligence (AI) that are critical for innovation [4]. Consequently, companies need to keep up with digitalization if they are to remain competitive and be a company of the future [5], [6].

A key dimension in digital transformation is data exploitation [4], and the ability to exploit the full value of data is increasingly becoming a source of competitive advantage [1]. Currently, the world is generating quintillions of bytes of data every day [7] and the amount of data generated is exponentially increasing. At the same time, a larger part of all data is becoming useful and accessible [8]. In addition to this, as companies are increasingly data-driven, the data fluency of organizations is improving, leading to an increase of receptible customers for data [8]. These factors have led to a great increase in possibilities to exploit data [9]. Firstly, data can help improve products by enabling optimization and giving a better understanding of customer usage and what delivers value [4]. Secondly, data is essential for developing AI models, which play a critical role in future technological innovation [4]. Lastly, data can act as a valuable asset for generating new revenues and enabling business model innovations [4].

Few organizations today have strategies for how to acquire, develop, and leverage proprietary data [10]. Instead, companies often tend to collect data accidentally and ad hoc [11]. Currently, most organizations are focusing on the internal use of data [10], e.g. using data to optimize internal processes within the company. When it comes to using data externally in the embedded systems domain, firms traditionally operate in one-dimensional value networks where they use data from one customer to that specific customer [4]. However, digitalized businesses tend to benefit from operating in multi-dimensional networks with many sources of revenue beyond their products [6]. This can be done by obtaining aggregated data from a primary customer group, and using it to create a monetizable service to a new, secondary customer base, i.e. customers who aren't part of the process in which the data was collected, but can benefit from it [4]. This is a viable opportunity because embedded systems companies' data, which is already being collected, can with low incremental risk and investment be leveraged into new revenue streams when it's useful to others [8]. Although this has been done in the online domain, there is currently no consensus as to how embedded systems companies can approach monetizing data to secondary customers [4], nor has the issue received sufficient attention from academia, leaving opportunities to guide practitioners in data monetization by providing more forward looking theories and models [12].

1.1 Research Goal and Questions

The purpose of the study is to shed light on how data can be used to create value in the embedded systems domain by researching how embedded systems companies can explore and develop opportunities of monetizing data to secondary customers. From this, the study intends to create a process model that assists embedded systems companies in the initial phases of developing data monetization offerings to secondary customers. The following research questions are investigated to achieve the research goal:

RQ1: *How are embedded systems companies collecting, processing and using data for monetization today?* The purpose of this research question is to understand the current state, challenges, and opportunities of embedded systems companies' data exploitation. This will be used as a basis for ensuring that the process model fits the context, and for understanding what the process model should cover.

RQ2: *What factors affect an embedded systems company's ability to monetize data to secondary customers?* The purpose of this research question is to understand what factors need to be considered when attempting to monetize data to secondary customers. The thesis will look at two sides of this issue. Firstly, factors affecting the embedded systems companies' maturity to monetize data to secondary customers. Secondly, how the dynamic with primary customers affects the ability to monetize data generated from them to secondary customers.

RQ3: *How can embedded systems companies develop services based on data from primary customers, to exploit these to secondary customer bases?* This research question aims to, based on the answers to RQ1 and RQ2, explore appropriate development processes for creating services that exploit data from primary customers, to secondary customers.

1.2 Scope and Limitations

To answer the research questions, an exploratory case study was conducted in six embedded systems companies in different industries. The study explored how the embedded systems companies are currently exploiting data within their organizations and what challenges and opportunities exist related to monetizing data from primary customers to secondary customers. Although the study considered the importance of data ownership, it did not intend to cover how legal aspects of the issue must be structured to enable data monetization to secondary customers. The same can be said about data valuation, although it's an important part of monetizing data, the study did not intend to investigate how embedded systems companies should value and price their data. The study is intended to benefit both researchers and practitioners.

A limitation of the study was that the it was unable to investigate aspects of data infrastructure related to data exploitation maturity due to limited access to the case companies. Furthermore, the lack of real use cases and limited access to the companies caused the last two steps of the model, assessing and allocating resources and building and testing MVPs, to not be as thoroughly validated.

1.3 Contributions

The contributions of the thesis are as follows.

C1: Identifies factors affecting how mature an embedded systems company is in regards to monetizing data to secondary customers.
C2: Identifies factors affecting how willing primary customers are to have their data monetized to secondary customers.
C3: Develops a structured process model guiding companies from data monetization ideas to scalable MVPs.

1.4 Report Structure

The remainder of the thesis is structured as follows. Section 2 introduces existing literature related to the study. Section 3 describes the research methodology conducted in the study and introduces the case companies involved. Section 4 presents the empirical findings of the study. Section 5 presents the model derived from the findings. Section 6 presents the key insights from the validation of the study. Section 7 discusses the model, threats to validity, and ethical implications of the study. Finally, section 8 concludes the thesis and suggests areas for future research.

2

Related Work

In this section, relevant literature is described to understand the problems investigated in this study. First, the section gives an overview of the embedded systems domain and its current state in relation to data exploitation. Secondly, the section covers related work on data exploitation and how data can be monetized. Finally, the section covers new product and service development methodologies.

2.1 The Embedded Systems Domain

The term ‘embedded systems’ varies in meaning depending on who’s using the term, what the application is, and what the context is [13]. What can be agreed upon is that embedded systems involve a combination of mechanical, electrical, and software components [13]. Embedded systems can be defined as computer systems embedded into either mechanical or electrical systems in which the computer system has dedicated functions [13]. An example of an industry operating within the embedded systems domain is the automotive industry. In the automotive industry, cars and trucks have gone from being purely electromechanical to relying heavily on software for functions such as autonomous driving. Other examples of industries operating within the embedded systems domain include the home appliance industry, mobile phone industry, and aviation industry.

2.1.1 Digital Transformation in the Embedded Systems Domain

Digital transformation is the cause of many of the challenges organizations are facing today [14]. It is changing how people are living their lives and how organizations are conducting their business [15]. The challenges of digital transformation are rarely to do with having access to technology itself [14]. Instead, it has to do with being able to rapidly implement new business models that are formed to create and capture value in new ways as a result of digital transformation [14]. Data is fundamental to the generation of new business models and utilising the data generated from different systems is essential for an organization to remain competitive [16]. Big data, analytics, and algorithms are enabling data-driven business models, in which data is viewed as an asset and monetized to open up new revenue streams [15].

The effects of digital transformations have been well documented in many domains, covering phenomena such as the complete restructuring of organizations, closer customer relationships, digital business models, and the use of digital technologies [14], [15], [16]. However, there exists very little research on the unique challenges and opportunities of embedded systems companies as they undergo digital transformation. Embedded systems companies are evolving from being hardware and product-oriented to focusing on software, data, and AI [6]. This is largely due to the increased amount of functionality taken over by software in embedded systems [17]. More research needs to be done in this area, as a challenge for embedded systems companies is that new digital value propositions are gradually becoming the primary value propositions [6]. Furthermore, the emergence of the Internet of Things (IoT) has had a large impact and is often described as a separator of winners and losers in the future of the embedded systems domain [8]. With big data and IoT being an essential part of the digital future, it means embedded systems companies need to develop their capabilities of generating, storing, analyzing, and exploiting data if they are to remain competitive [6], [8]. In a case study by Bosch and Olsson [6], it was seen that all participating embedded systems companies currently had a low maturity level when exploiting data, leading to few possibilities to expand into a multidimensional network with new revenue opportunities towards existing and new customers.

2.1.2 Data in Embedded Systems

Embedded systems companies collect several different types of data, but the data of interest in this study is data generated from the systems embedded into the products, i.e., asset data. This data is typically collected from sensors in the systems and can be accessed and stored in real-time with technologies such as cloud computing [18]. Examples of types of data that embedded systems companies are collecting include radar, lidar, sonar, accelerometer, GPS, and temperature data. Technology advancements, such as cloud computing, open up many new opportunities as to how data can be monetized.

2.2 Data Exploitation

Exploiting the value of data is becoming increasingly important across industries and domains, and the embedded systems domain is no exception [6]. This section first reviews current literature on the value of data to provide guidance in what to consider when attempting to understand the value of data. Next, the section presents current literature on monetizing data, different ways that it can be done, what is necessary, and what challenges it can involve.

2.2.1 The Value of Data

Traditionally, the value of data has been twofold [19]. Operationally, data has provided great support in decisions and business processes throughout organizations. Strategically, data has been vital to organization's operational backbone for customer engagement and digital solutions. However, data is starting to shift from being a secondary asset supporting decisions and processes, to a primary asset that can be monetized [19].

Figuring out how to put a value on data has long been a challenging issue [7]. This challenge becomes even more relevant in the context of monetizing data as an asset. Previous research has created guidelines and characteristics to help value data [7], [20], [21]. Some of the value drivers of data are as follows: exclusivity, timeliness, accuracy, completeness, consistency, usage restrictions, interoperability/accessibility, and liabilities and risk [7], [22]. While individual data-points can be of value, the true value is often in insights found in patterns from thousands of readings from many different products over time [1]. Furthermore, the data from products can be of value by itself, but it is exponentially more valuable when it is integrated with other types of data [1]. Proprietary data often holds the most potential value for monetization, as it is unique to a company, hard to replicate by other entities, and therefore, able to create sustainable competitive advantages [10]. Proprietary data often has positive network effects, where acquiring or integrating more data and curating it effectively increases the value of the data [10].

2.2.2 The Data Exploitation Dimension

Bosch and Olsson [4] present a model detailing the typical evolution path embedded systems companies take when transitioning towards becoming a more digital company. One of the dimensions of this model is the data exploitation dimension, which is critical for the transition to be successful. Figure 2.1 shows the steps in the data exploitation dimension when transitioning from a traditional to a digital company [4]. Companies can be at multiple steps at the same time and the data exploitation dimension shows in what ways companies can exploit their data. Bosch and Olsson [4] observed multiple embedded systems companies in steps one, two, and three, but steps four and five were only mentioned as ideas by at least one of the case companies. However, based on their empirical findings, Bosch and Olsson [4] foresee a future in which steps four and five will also be achieved.

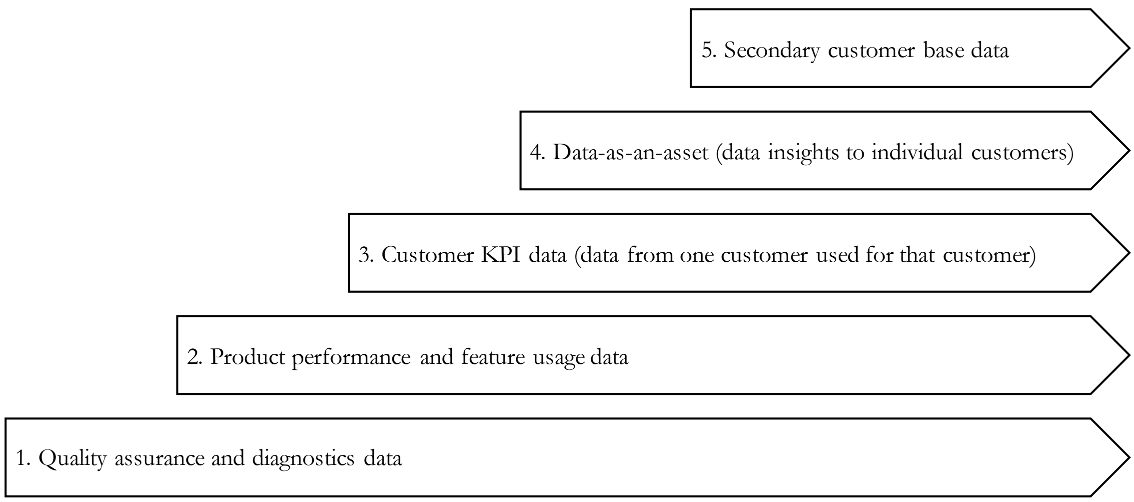

Figure 2.1: The Data Exploitation Dimension

2.2.3 Data Monetization

Wixom and MIT Center for Information Systems Research define data monetization as "*the act of exchanging information-based offerings for legal tender or something of perceived equivalent value*" [23, p. 1]. The ability to monetize data is being led by a few forces. Firstly, the cost of storing data is decreasing, and real-time big data processing and analytics is improving [12]. Secondly, businesses are recognizing the amount of unrealized potential value of data and therefore making business intelligence and analytics a top priority [12]. According to research, big data, business intelligence and analytics, and the cloud, are the current trends allowing for data monetization [12].

2.2.3.1 The Information Offerings Consumption Path Model

There are many ways in which data can be monetized by a company. Buff et al. [19] present a model of the information offerings consumption path, which covers the range of data monetization offerings that companies can create to generate revenue. It is a continuous spectrum, containing three phases: data, insights, and actions. Companies can choose to position themselves in any parts of the spectrum of information offerings depending on its environment, resources, and capabilities. No matter where along the spectrum the company chooses to position itself, it is vital to have a deep understanding of the end customers' problems, as well as how the offering influences the customers' processes and decisions. Not only is this understanding necessary to align the offerings with worthwhile use cases, it also increases the chance of successfully adjusting to the end customers' changing demands. Furthermore, it is vital to be aware of the fundamental importance the data quality has on all types of offerings, as they all depend on the underlying data.

The first phase of the information offerings consumption path [19], data offerings, includes raw and processed data. This phase is the furthest from the end customers' consumption of insights. Therefore, it is critical for companies offering data offerings to understand what happens with the data after the point of sale, to understand both the value and the risks the data monetization activity entails. This understanding is especially important when selling customer data, as it generally tends to be more sensitive.

The second phase, insights offerings, includes reporting and analytics [19]. These types of offerings are often centered around presenting information in a way that is easy to understand and use. Therefore, there is often an emphasis on data visualization and user interfaces. Reporting offerings present business intelligence through visualizations, while analytics offerings take it a step further and apply mathematical algorithms, statistical modeling, and machine learning techniques to find deeper insights in the data.

The third phase, action offerings, includes process design and process execution [19]. This category of offerings helps the customers act on the insights, often related to decision-making or solving a problem. These types of offerings mean a close relationship between the company and the customer, where the company contributes not only with data and insights, but also specialized expertise such as technical or managerial skills and deep industry knowledge. Process design offerings involve using insights from reporting and analytics to provide recommendations to the customer regarding decisions and change-processes. Process execution goes further and makes decisions on behalf of the customer, commonly taking on some of the risks and costs, as well as the value created.

2.2.3.2 Business Capabilities for Successful Data Monetization

Buff et al. [19] have identified three sets of business capabilities vital for successful data monetization: people, platform, and perception. Firstly, successful data monetization requires several types of roles depending on the context and type of information offering. Data management professionals are necessary to ensure data quality and compliance. Developers and designers are important when the information offering requires good user interfaces and user experiences. Technical engineers are needed when secure, cost-efficient networks and infrastructures have to be built as part of the information offering. When modeling, analytics, statistics, and machine learning are required for the information offering, the company needs data scientists, analysts, and statisticians.

Secondly, an appropriate platform for the data monetization offering is essential for both user compatibility and cost efficiency [19]. The requirements of the platform depends on the type of information offering, but the six main types of attributes are integration and openness, performance, cost-efficiency, automation, repurposability, and continuous improvement.

Thirdly, perception of the customers existing, evolving, and future needs are key for successful and sustainable data monetization [19]. The primary reason behind this is that it allows organizations to develop information offerings that continually address important problems and decisions for the customer. Two of the main ways of obtaining perception identified by Buff et al. [19] is through deep domain expertise and customer interaction. Most information businesses successful in data monetization have an intimate understanding of the domain that they develop services for. Common ways for these companies to obtain deep domain knowledge is through hiring employees who have previously worked in the domain, or even for the customer, or training current employees on the domain. In customer interaction, it is important to have several customer touch-points and frequent interaction in order to continuously learn about the customer. Furthermore, it is important to have the information obtained from the touch-point, for instance through sales, to permeate other parts of the organization such as product management and development to be used to shape existing and new offerings.

The culture of a company significantly affects the probability of successful data monetization initiatives [19]. Some of the key traits in culture associated with successful information offering companies, identified by Buff et al. [19], are experimentation-driven and risk taking, and having credibility and trust. The two main factors for trust and credibility, according to Buff et al. [19], is transparency and fairness. Transparency can be achieved by helping stakeholders clearly understand the data monetization transaction, as well as ensuring that the affected parties are comfortable with limitations and controls set in place. Making sure that all stakeholders feel fairly treated is mainly based on how the value-sharing is set up in the network of actors.

2.2.3.3 Issues and Challenges with Data Monetization

Bataineh et al. [22] identify challenges of data monetization from both the data consumer's and the data provider's perspective. Three challenges for the data consumer when attempting to acquire data are: (1) to target the right providers who are able to deliver the required quality of data, (2) limited budgets, and (3) to collect the right amount of data to satisfy their needs [22]. Three challenges from the data providers' point of view are: (1) to find customers who are interested in the data, (2) to know what the monetary market value of their data is, and (3) to maximize profits generated from data.

2.3 New Product and Service Development

This section first presents related work on new service development for manufacturing firms, such as embedded systems companies. This is followed by sections on methodologies for developing new products and services, with a focus on software development. The purpose of this section is to act as a foundation for determining an appropriate development process for developing data as a service to secondary customers initiatives.

2.3.1 New Service Development for Embedded Systems Companies

In many industries, embedded systems companies are increasingly focusing on developing services in addition to their products [24]. However, as most embedded systems companies have had a product-oriented focus, adopting a more service-based logic is challenging as it requires a set of new resources and capabilities [24]. Furthermore, the conflicting service and product cultures in a company that is trying to expand more into services is often a significant challenge [24].

When adopting a service-based logic, companies must change their view of customer relationships from a transactional view with a passive customer to a more long-term relationship view [24]. Service suppliers should therefore look at the value delivered in terms of long term value-in-use instead of the more traditional short-term value-in-exchange [25]. Furthermore, when developing new services it is important to focus on the customer's practices and identify customer needs beyond traditional marketing research [25]. The customers expressed needs can sometimes be misleading, but by understanding the customers everyday practices the service provider can understand what is truly valuable to the customer and how they best can support the customers [25].

Kowalkowski and Kindström [24] propose a circular, continuous four-stage model for guiding companies in their new service development: (1) Market sensing, (2) Development, (3) Sales, and (4) Delivery. They raise critical aspects and main organizational challenges for each stage. For market sensing, a critical aspect is for firms to be able to balance exploiting their current business and exploring new means of revenue [24]. A key organizational challenge in this stage is that many embedded systems companies lack the capability to sense market opportunities, particularly in the cases where services tend to be given free of charge to support product sales [24]. Similar situations were seen in a case study by Gremyr et al. [26], which saw an initial scepticism of selling services as their traditional services had been free of charge to support product sales. In the development stage, a critical aspect raised by Kindström and Kowalkowski [24], is to design the new service development process to involve the customer. They further raise the importance of increased cross-functional coordination during development. In their case study [24], they also saw that most services developed were highly contextual, and thus a challenge for embedded systems companies is to find ways to industrialize the service offering so that it can better be sold to more customers. In the sales stage, critical aspects are related to giving the sales teams means to sell services instead of products, which often means more customer-centric and relationship-based transactions [24]. This requires that the sales personnel have a good understanding of the customers operations and revenue logic [24]. Finally, in the delivery stage, it is crucial for the firm to be able to show monetary benefits achieved for the customer [24].

2.3.2 Agile Development

The agile methodology is an iterative team-based methodology in which requirements and deliverables evolve through cross-functional collaboration [27]. Instead of sequential phases with different tasks, the team defines deliverables to be completed within a time-boxed phase, and thus fixes the time and resources to determine the scope for the iteration. The typical agile workflow can be seen in Figure 2.2 below.

The advantages of agile methodology include that it's highly focused and accelerated, customers are highly involved, and requirements are communicated during development when there is more knowledge [27]. Also, the iterative nature of the agile methodology spreads the risk to each increment of the development instead of to one large development. The disadvantages include that the scope could be reduced and quality of deliverables compromised due to the time boxing, and interdependent deliverables can result in suboptimal design and a system with parts that don't work well together [27].

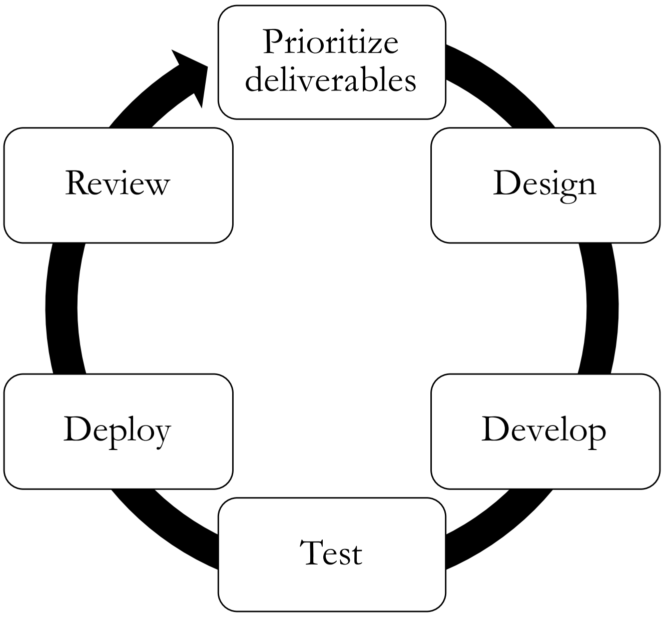

Figure 2.2: The Agile Workflow

2.3.3 Lean Software Development

Lean Software Development is about applying the principles of Lean Manufacturing to software development. Poppendieck and Poppendieck [28] have summarized Lean Software Development to the following seven principles:

1. **Eliminate waste.** In Lean philosophy, anything that doesn't add value to the customer can be viewed as waste. Examples of waste include extra features, waiting, management activities, relearning, and switching people between tasks. In a software development context, waste can include collecting requirements that are not used and coding extra features that aren't immediately needed. The goal in Lean Software Development is to find out what the customer wants, and develop this as quickly as possible.

2. **Amplify learning.** Amplified learning in Lean Software Development is about continuously learning while developing the software. This is achieved with short iteration cycles during development with feedback from the customers each iteration. With short feedback loops, customers and developers learn more about the problem and where efforts need to be put for further improvement. Short feedback loops help the customer understand their needs by having knowledge of the current development efforts, and in turn, helps the developers learn what the customer values so that the right efforts can be put into development.

3. **Decide as late as possible.** When there is uncertainty, late decision making is effective because it provides an option-based approach. Making decisions as late as possible is valuable because the decisions will be based on more facts, due to the increased knowledge from the customers understanding their

needs better. To make decisions as late as possible, the software needs to be developed to allow for change into the system.

4. **Deliver as fast as possible.** It's essential to be able to develop software rapidly to be able to delay decisions and have short iteration cycles that allow for feedback to be received sooner. Delivering software as fast as possible thus improves the software through learning because of shorter iteration cycles.

5. **Empower the team.** Because decisions are made as late as possible and the software is developed as fast as possible, it's impossible for a central authority to manage the developers. Instead, software developers should be given access to the customers and be trusted to get the job done.

6. **Build integrity in.** This principle is about providing and maintaining system integrity, which includes perceived integrity and conceptual integrity. Perceived integrity refers to how it is delivered, maintained, priced, solves the problem, etc. Conceptual integrity represents how the system's components work together as a cohesive whole.

7. **See the whole.** Experts in different areas tend to put all their focus on their area instead of focusing on overall system performance. It's even more so in the case when two organizations work together, due to people trying to maximize the performance of their own organization. Seeing the whole is important to avoid suboptimization and it's particularly difficult when contracts are involved due to several organizations being involved.

2.3.4 Lean Startup

Eric Ries, the author of The Lean Startup, defines startups as "*a human institution designed to create new products and services under conditions of extreme uncertainty*" [29, p. 7]. From this definition, which implies that entrepreneurs can be everywhere, Ries states that the Lean Startup approach "*can work in any size company, even a very large enterprise, in any sector or industry.*" [29, p. 7]. The Lean Startup methodology is based on many previous management and product development ideas, and Ries [29] explicitly lists: lean manufacturing, design thinking, customer development, and agile development as foundations for the Lean Startup. Ries [29] extends on agile methodology by creating a methodology that aims to remove the waste of building a product or service that customers don't want to use, by learning more before building the product or service. Ries further argues that this is a flaw in the agile methodology, and states the following: "*Agile is an efficient system of development from the point of view of the developers. It allows them to stay focused on creating features and technical designs. An attempt to introduce the need to learn into that process could undermine productivity*" [29, p. 124].

2.3.4.1 Validated Learning

The main purpose of the Lean Startup methodology is effective validated learning about how to build a sustainable business from a vision. Validated learning refers to learning that is scientifically validated through experiments based on hypotheses [29]. It is done through one of the most fundamental activities of the Lean Startup methodology, the Build-Measure-Learn (BML) cycle [29]. Its purpose is to build something with minimum resources that the customers can try, measure customer feedback, and from that learn if the venture should persevere, pivot or perish. It is worth noting that while the BML-cycle is carried out in the order of build, measure and learn, the planning of each cycle is done in the reverse order [29]. First the company must plan on what needs to be learned, then what needs to be measured in order to learn it, and from there decide on how to build something that can be used to measure it.

2.3.4.2 The Hypotheses

Using the Lean Startup methodology means being hypothesis-driven [29], [30]. The first stage in developing a new business idea is to develop a vision. It is from that vision that hypotheses will be formulated. The purpose of the hypotheses is to through experimentation find a business model from the vision that achieves product/market-fit. The questions that need to be answered about the business model can be divided into four categories: customer value proposition, technology and operations plan, go-to-market strategy, and cash flow formula [30].

The hypothesis must be falsifiable, meaning that it can be verified decisively through an experiment. Furthermore, it must be a hypothesis that can fail, because if it is defined as something that is very unlikely to fail, then it won't lead to learning [30]. Furthermore, while it is wasteful to initially develop detailed hypotheses on all elements of the business model, and should instead be done iteratively, it is still important for the set of hypotheses to be comprehensive [30]. This means quickly passing through all elements of the business model early on, mainly to identify potential deal-breakers or inconsistencies [30].

2.3.4.3 The Experiments

First, the experiment must be planned and structured. In this stage, it is important to be aware of the trade-off between effort and the level of believability of the information [31]. This trade-off is illustrated in Constable's [31] truth curve in figure 2.3, which shows different experimentation archetypes with their respective effort required and the believability of the information they generate. Constable [31] recommends prioritizing the quicker and easier experiments, and progressing to the higher effort experiments when more is known.

Following the planning of the experiment is the build phase. The experiments should have constrained functionality and only contain the necessary features to answer the hypothesis [30]. Furthermore, it should have constrained operations, meaning that

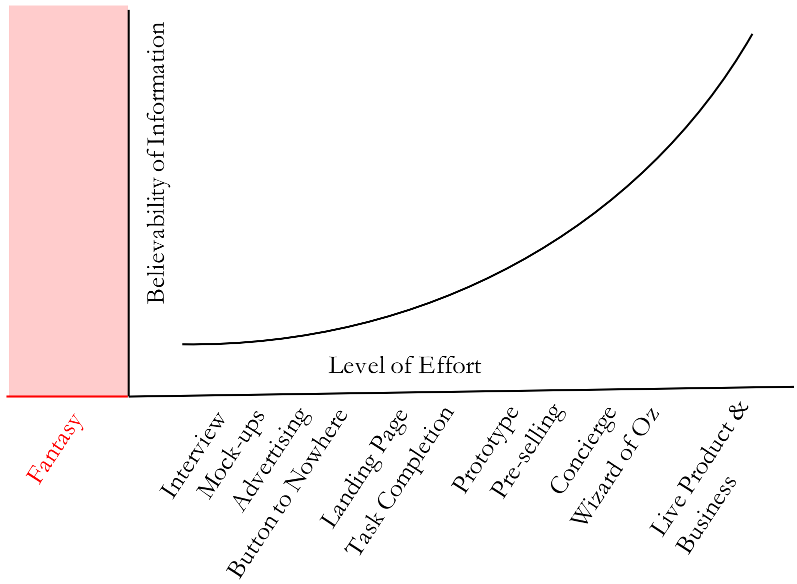

Figure 2.3: The Truth Curve

it can be a makeshift and temporary, unscalable solution, just to test customer demands [30]. Finally, as the experiment has constrained functionality and operations, it should be tested to a constrained customer set. The customer set should only just be large enough to validate the hypothesis, more than that would require unnecessary resources and could potentially increase the risks of damaging the brand [30].

To be able to reject or validate hypotheses after an experiment and measure progress of the venture, good metrics are needed. It can be difficult to figure out what metrics are key in a new venture, however the overall purpose of analytics remains the same: to find the right offering and market before the budget runs out [32]. According to Croll and Yoskovitz [32], a good metric is: (1) comparative, for instance in relation to time, (2) understandable, to enable people within the organization to discuss it and learn from it, and (3) a ratio or a rate, as they are easier to act on, inherently comparative and are good for comparing opposing factors and their relation. However, the most important quality of a metric is that it should be able to change the way you behave [32].

Finally, the outcome of the experiment should lead to validated learning. Based on the validated learning, a decision must be made for how to proceed. Ries [29] frames it as a decision with three options: to persevere, pivot or perish. Constable [31] frames it similarly, with four options: (1) The experiment gave insufficient information to confirm or reject the hypothesis and needs to be further tested, (2) you can confirm that the hypothesis is true and move forward, (3) you change the hypothesis based on the information received from the experiment, and (4) you reject the hypothesis and terminate the initiative. Constable's second option can be seen as persevere, the third as pivot, and the fourth as perish.

2.3.4.4 Running Lean

Running Lean is a book written by Maurya [33] that takes a practical approach to the principles of Lean Startup. Maurya [33] presents a process of three key steps to maximize the chances of creating a successful new product or service:

1. Document your Plan A.
   This step is about writing down the initial set of business model hypotheses. To do this Maurya [33] recommends the Lean Canvas, which can be seen in figure 2.4, to create a fast, concise, comparable, and portable business model. The Lean Canvas divides the business model into nine key parts, which can then be systematically tested.
2. Identify the riskiest parts of your plan.
   The next step is about identifying and prioritizing risks. There are three categories of risk when creating a new business or service: product risk (getting the product right), customer risk (building a path to customers), and market risk (building a viable business).
3. Systematically test your plan.
   After having a documented plan and prioritized risks, the final step is to systematically test the plan using BML-cycles. To test the nine key parts of the business model, Maurya recommends the following process: Understand the problem (is the problem worth solving), define the solution (how customers would like the solution to be), validate qualitatively (build an MVP), and verify quantitatively (launch refined product to a larger audience).

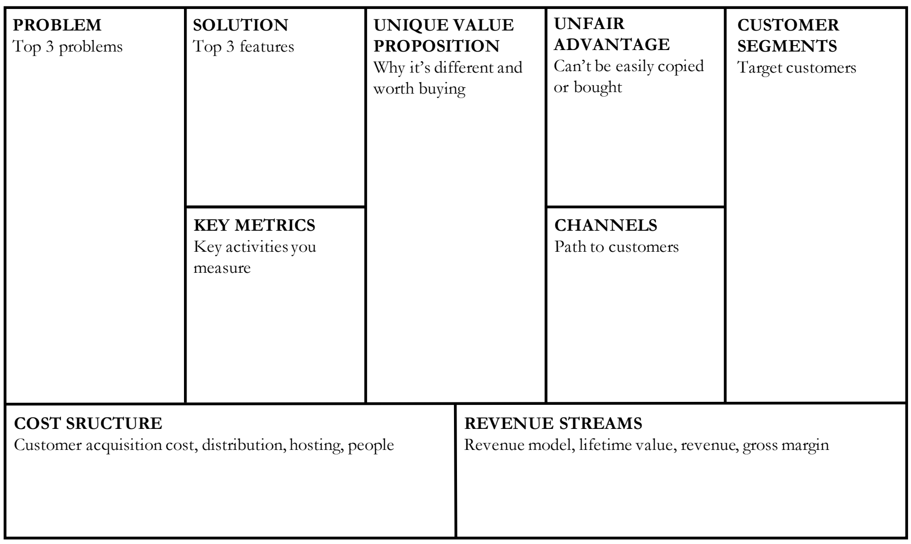

Figure 2.4: The Lean Canvas

2.3.4.5 The Early Stage Software Startup Development Model (ES-SSDM)

The ESSSDM model extends existing Lean Startup principles by offering practical support for investigating multiple ideas in parallel [34]. It's purpose is to find an idea worth scaling and to provide guidelines on when to abandon ideas. The model consists of three steps:

1. Idea Generation.
   The model supports working with several ideas in parallel. This step is about using techniques to find ideas for potential products.
2. The Backlog.
   Startups don't have the capacity to work on all ideas simultaneously. The purpose of the backlog is to write the ideas in a comparable format so that they can be prioritized.
3. The Funnel.
   The final step of the model is the funnel. In this step, multiple ideas from the backlog are fed into the funnel. These ideas are systematically tested using BML-cycles, as described in Running Lean [33], with guidelines on how to work with multiple ideas in parallel.

2.3.4.6 Limits to the lean startup methodology

Finally, there are limits to the Lean Startup methodology worth mentioning. It is less appropriate in situations where: (1) mistakes must be limited, as the Lean Startup methodology uses the philosophy of fail fast and often, and learn from it, (2) demand uncertainty is low, because then launching early and often in order to learn about the customer demand becomes unnecessary, and (3) demand uncertainty is high, but development cycles are long, as this makes it impossible to launch early and often [30].

3
Methodology

This section describes the design of the study, including how the research questions are to be answered. As there is still very little done within the research area of monetizing data to secondary customers in the embedded systems domain, the research area is viewed as nascent, and therefore studied exploratively and qualitatively to ensure methodological fit [35]. While qualitative research methods originally stem from social sciences, it is increasingly used in software engineering to understand problems related to human behaviour, humans performing software engineering tasks such as development processes, and humans using software [36]. Furthermore, the study takes an applied research approach, as it aims to produce outcomes useful for practitioners to improve their conduct of business and use of technology [37]. This is done by exploring the value of basic knowledge in an applied setting, as described by Easterby-Smith et al. [37]. An overview of the research sequence is shown in figure 3.1.

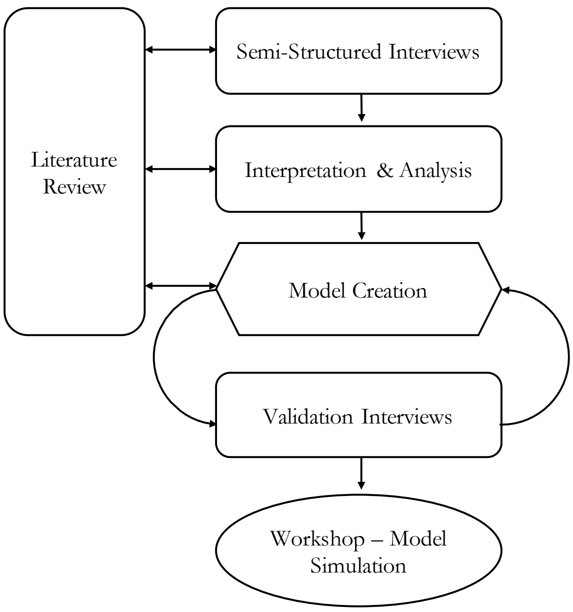

Figure 3.1: Overview of the Research Sequence

3.1 Case Study Design

Due to the exploratory nature of this study, the case study research methodology was deemed most suitable to answer the research questions proposed in section 1. Yin [38] describes that case studies are preferred when "how" questions are being posed, when the investigators have little control over events, and when the focus is on a contemporary phenomenon within a real-life context. This is all applicable to the context of this study and supported by Runeson et al. [39], who claim that case study methodology is a perfect match for exploratory research questions. The thesis conducted a multiple-case study with six companies operating in different business areas within the embedded systems domain. The multiple-case study design is likely to be stronger than single-case study designs, because it makes the results more generalizable [38]. The companies were selected through convenience sampling, primarily through a collaboration of universities and companies that work together to accelerate the adoption of novel approaches to software engineering. The study focused on the perspectives of people involved in the process of data exploitation, both on the business side and technological side.

3.2 Data Collection

Ten semi-structured interviews, with the six case companies, were conducted to collect a deep empirical understanding of how these companies are collecting, processing and using data for monetization today. The interview questions were designed with the purpose of identifying and learning about key challenges and opportunities related to monetizing data for embedded systems companies, what ambitions they have, and what issues they are running into. The case company interviews were semi-structured due to its flexibility, depth, and room for exploratory discussions [37]. This is seen as appropriate for the exploratory nature of the study. The semi-structured interviews followed a topic guide based on a thorough review of the relevant literature, in order to have a clear outset for the areas of interest [37]. The topic guide was also useful for making sure that all areas of interest were covered in each interview, as well as maintained some degree of structure [37]. As recommended by Easterby-Smith et al. [37], the topic guide is organized into three sections: opening questions, questions around key topics, and closing questions. The topic guide can be found in Appendix A. All interviews were recorded, after getting permission from the interviewee. Each interview was then transcribed.

3.3 Interpretation and Analysis

Thematic analysis was used to identify, analyse, and report themes within the data. The thematic analysis followed the methodology of Brown and Clarke [40] and their six steps of thematic analysis: (1) familiarization with the data, (2) generating initial codes, (3) searching for themes, (4) reviewing themes, (5) defining and naming themes, and (6) producing the report. Table 3.1 describes how each step was achieved.

Step	Action
Familiarization with the data	The data gathered from the semi-structured interviews was transcribed as a first step to get familiarization with the data and to conduct thematic analysis. It was read several times in an active way, i.e. searching for meanings, patterns etc., as Braun and Clarke [40] suggest to prepare for coding in the next step.
Generating initial codes	Initial codes were created by systematically identifying features of the data that seem interesting and may be the basis of repeated themes across the data set, as Brown and Clarke suggest [40]. The coding was done with the help of NVivo and completed by both researchers individually to reduce the risk of bias [37].
Searching for themes	Once all data had been individually coded by the researchers these were combined and different codes were sorted into potential themes.
Reviewing themes	In this step the themes were analysed and refined so that the data within themes cohered together meaningfully and so that there was a clear distinction between themes.
Defining and naming themes	This step consists of identifying the essence of what each theme is about and how they fit in relation to the research questions.
Producing the report	Finally, once the overarching themes had been defined, section four of the report was written to tell the story of the data. Data extracts, i.e. quotes, were used to demonstrate the prevalence of the themes.

Table 3.1: Actions performed during thematic analysis

Based on the empirical data analyzed from the interviews, a gap between where embedded systems companies are today, and where they want to be, was identified. As companies expressed interest in monetizing data to secondary customers, but none had launched a successful venture yet, the findings showed a need for a structured process to assist and increase the chances of success when starting out with monetizing data to secondary customers. Therefore, a model for the initial phases of developing data monetization offerings to secondary customers was created. The model is inductively derived from the findings of the case company interviews together with existing literature.

3.4 Validation

After inductively deriving the model, a validation phase was conducted to obtain increased validity to the model. The validation phase involved collecting evidence from multiple sources, to validate the model through triangulation [41].

Firstly, the validation was done through interviews by returning to some of the case companies. In the validation interviews, the model was first applied as far as possible with each case company in a workshop-like manner. Afterwards the model was discussed and feedback was gathered from the interviews. The interviews gave useful insights from a practitioner's point of view. The main insights from the validation interviews were regarding their impression of the model, its applicability to their organization, if it could be improved for their context in any way, and a discussion about the order of the steps in the model. The feedback from the validation interviews were analyzed iteratively by continuously adjusting the model and validating the new version in following validation interviews.

Secondly, the model was validated by using it to simulate use cases. Due to the nascent nature of the research, where no case company has yet to monetize data to secondary customers, as well as the researchers limited access to companies, the researchers simulated the use cases by acting as the embedded systems company, primary customer, and secondary customer. The simulated use cases involved obtaining publicly available datasets from embedded systems, and based on that data, simulate the process of the model. The simulated use cases were selected with the findings from the case interviews in mind, to ensure that relevant and useful cases were used. The simulated use cases both contextualized the model and tested the logic, steps, and techniques within the model.

Finally, the model was validated by applying it, in a workshop, to a real use case that was in progress at one of the case companies. As the case company had not yet gone through the entire process, the model was applied to the extent of their progress. The real use case tested the model in reality, and allowed the researchers to compare the outcome using the model compared to the case company's way without the model.

3.5 The Case Companies

Six embedded systems companies were interviewed as part of the study. The case company selection covers several different industries and varies in data exploitation maturity.

3.5.1 Case Company A

Case company A is operating in the automotive industry. They are working with developing autonomous driving and safety features. The company is active globally with offices around the world. They collect various sorts of datasets in large amounts through sensors and cameras on vehichles.

3.5.2 Case Company B

Case company B is in the power controls solution industry. They are offering environmentally friendly power control solutions to decentralized power production units on land and sea. The company is operating globally. They are able to collect asset data from the power production units in operation.

3.5.3 Case Company C

Case company C is in the automotive industry. They are producing trucks and offering services supporting their products, such as efficiency optimization and fleet management services. The company is operating globally. They are collecting data through various technologies, like sensors and connectivity components in their trucks.

3.5.4 Case Company D

Case company D is offering process and packaging solutions to many different industries related to consumer goods and has a large market share of the global market. They are collecting a large volume of asset data as part of the process solutions, as well as a large amount of production data as part of their packaging solutions.

3.5.5 Case Company E

Case company E is in the military defense industry. They are offering a wide range of products and services for civil security. Case company E is operating globally with customers all over the world. Their products are collecting large volumes of asset data in operation. However, given the industry that case company E operates in, the data is often highly sensitive and well protected.

3.5.6 Case Company F

Case company F is offering bearing and lubrication products and services, as well as adjacent services like education and mechanical maintenance. As bearings and lubrication are fundamental in many different types of products and machines, they are operating in multiple different industries. Case company F is a large and mature company. As part of their product, and their optimization services, they are able to collect a lot of asset data used for analytics and services such as predictive maintenance.

3.5.7 Practitioner Roles

The following list presents the roles of the practitioners who participated in the interviews. The roles are presented in the following list without disclosing what case company they represent to improve anonymity:

- Chief Digital Officer with data, security, and IT as their responsibilities
- Chief Technology Officer, head of all research and development
- Head of Data Governance
- Business Owner of third party developer environment
- Technical Director for software product enablers
- Product Manager for data service initiatives
- Machine Learning Specialist
- Head of Data Services
- Enterprise Architect for Data Management and Analytics
- Software Innovation Specialist

4

Empirical Findings

This section reports the findings of the study from the semi-structured interviews held with the case company practitioners. First, a summary of the findings is given. Following this, it's reported how the case companies currently collect and exploit data. Next, factors affecting the maturity to monetize data to secondary customers are identified. Finally, factors affecting primary customers willingness to have their data monetized are presented.

4.1 Summary

The case companies are all collecting asset data from their product offerings, which can be valuable to other companies besides their primary customers. Currently it's the core business and the case companies' domain knowledge therein that dictates what data to collect. In terms of how the data is used to deliver value to the case companies, it was used primarily for quality assurance, diagnostics, product performance, information about feature usage, and customer KPI data. This is equivalent to steps one, two, and three in Bosch and Olsson's [4] Data Exploitation Dimension, seen in section 2.2.2. The data is typically offered as a free complement to the embedded systems companies' products to meet customer needs. For the case companies that have reached step three in Bosch and Olsson's [4] Data Exploitation Dimension, the data was offered as a free or charged complement and involved providing services or insights to the primary customers.

None of the case companies are currently monetizing the data they collect to customers outside of their core business. However, all but one of the case companies expressed an interest to do so, and some were in the initial phases of exploring opportunities. The biggest concern to monetizing data is the negative impact it can have on their core business and primary customers. Furthermore, the case companies are unsure of how to go about pursuing such opportunities because it lies outside of their expertise and domain knowledge. Factors identified, from the empirical data, that are important to consider when monetizing data to secondary customers include: available resources, data ownership, anonymization, secondary customers, data valuation, and how such an initiative can affect the primary customer. The main factors identified in the case study affecting how willing primary customers are to have data generated by them monetized is: impact, trust, incentives and legal agreements.

4.2 Data Collection in Embedded Systems today

Data collection is an essential part of a company's data exploitation and all of the case companies collect data from their primary customers. The findings of the study indicate that the challenges and opportunities lie within the business area of data collection. The case companies expressed few challenges with their technological ability to collect and store data, but more so with the strategic business decisions surrounding the data.

The majority of data being collected by the case companies is asset data, i.e. data from the products that the case companies are offering to their customers. Asset data are collected with the help of technologies such as sensors, transmitters, and the cloud, allowing data to be measured and captured by the companies. Popular cloud computing platforms used to handle the data amongst the case companies include Amazon Web Services (AWS) and Microsoft Azure. These technologies have enabled the case companies to collect more data in a more convenient way and to freely decide what data they wish to collect from their assets. The case companies all expressed that the amount of data points they collect has increased in the past few years.

4.2.1 How Companies Decide what Data to Collect

The case companies' approaches to deciding what data to collect were mostly done in an ad hoc fashion. All of the case companies used an approach where the core business and domain knowledge decided what data to collect. The following quote indicates how domain knowledge was used to decide what data to collect:

> *We have [redacted for anonymity] years of knowledge of machinery where things spin. So we know that in an electric motor, for example, the most common failure is either its current passing through the bearings, or it's overheating of the grease. So that means that you need to measure the temperature of that bearing and you also need to measure the power.*

The following quote indicates how the core business would decided what data to collect:

> *What data to collect is determined by our feature teams, based on their specific needs. So if the car has to stay in the current lane that it's going in, what do they need to test and what kind of closed-loop open-loop re-simulation needs to happen on the data that they collect?*

The core business approach includes both an internal perspective, in which data is collected to improve internal operations in the companies, and an external perspective, in which data is collected to deliver something to the primary customer based on their needs and wants. All of the case companies focused on both the internal and external perspective when deciding what data to collect. One case company however, is unique because they used domain knowledge to decide what data to be able to collect, but then only collected the data points that the customer was interested in, meaning some sensors would go unused.

4.3 How Data is Exploited today

The case companies all exploit data both internally in the company and externally to their primary customer groups. The main focus of data exploitation in the case companies has been on using it to optimize product performance and new product development. In more recent years, exploiting data for services complementing their products for their primary customers has gotten increasingly more attention, but it is still, to a large extent, new territory for the case companies.

4.3.1 Exploitation of Data to Support the Embedded System

The case companies were all mature in exploiting data in ways that supported their products. The most common use of data from embedded systems was related to diagnostics and product performance. With this data, companies were optimizing the usage of embedded systems through means such as optimizing energy consumption. In addition to this, some companies were able to use this type of data to offer predictive maintenance to their customers. In new product development, all case companies collected product performance data together with feature usage data to improve functionality and meet customer demand. In the case companies' current efforts of exploiting data, they expressed challenges related to identifying opportunities in the data, data governance issues such as assigning data owners, the large volumes of data, and establishing uniform formats and structures throughout the organization.

4.3.2 Exploitation of Data as a Service

Most case companies had initiatives to start exploiting data as a service, but the large majority expressed it as very nascent. Half of the companies were exploring opportunities but had yet to launch any data as a service. The other half had some data as a service offerings and were actively exploring to further expand their data exploitation capabilities.

Amongst the case companies that were offering data as a service, all of them were offering it as a complement to their products for their primary customer base. In these cases, the embedded systems companies had two different ways to offer data to their customers. The first, more common approach, was to provide it as a free complement in order to increase sales of their products and to increase customer satisfaction. In two case companies, the customers even considered it as a necessity for them to continue purchasing their products, which can be seen in the quote below:

> *We were in a situation where customers were telling us, if you don't have a remote monitoring solution; if you don't have a way to collect data from my device, I don't even want to buy your equipment.*

In the second, less common approach, data-based services were provided as a paid complement to their embedded systems, which was paid through a subscription fee. These types of complementary data-based services were often insights, i.e. conclusions or KPIs provided to the customer. One key challenge for the embedded systems companies who were offering data as a service was to determine in what condition and level of refinement to deliver the data, as shown in the following quotes:

> *A feedback we get from customers as we are building a solution, let's say a vendor solution with an API mobile app. Their company will also have mobile app developers, naturally, there will be some software people. So a lot of customers also want to do it themselves. So those customers will say, 'Okay, I will pay for API, that is enough for me'.*

> *And what are the differences [between delivering clean and processed data vs. raw data]? And how much monetary uptake is there between the two models? Is this something that's an order of magnitude between it or is it something where customers don't really care because they'll do it themselves anyways?*

No case companies were currently selling data as a service as a standalone offering. Furthermore, the findings show that none of the case companies were exploiting their data to other customers than those who purchased their products. However, when asked, all case companies thought that there were other potential secondary customers who would be interested in the data that they collect. Five companies showed interest in pursuing it and two of them were currently exploring opportunities for it. For these two, the secondary customers had some kind of value offering to the embedded systems companies' primary customer. The findings also show that one of the main barriers of monetizing data to secondary customers for the case companies is the risk of negatively affecting their business with their primary customers. All case companies raised it as the most important issue to first resolve when discussing monetization of data to secondary customers. Furthermore, as long as monetizing primary customers' data to others was perceived as a risk to their core business, few case companies were willing to contemplate it as a way to create additional revenue streams. The reasoning of those case companies can be exemplified in the quotes below:

> *No, no, no. No, because we think that [monetizing data to secondary customers] is not our core and we find it will be dodgy and not so ethical.*

> *But anything you do with a third category [monetizing data to secondary customers] requires serious investigation from a commercial point of view, which we don't do, because the customers are giving the data based*

on trust. So we don't want to breach that trust for a business opportunity. Anything we do, we very transparently communicate with customers, if customers are not convinced we will not do it.

Other case companies however, saw that there were ways to offer data as a service to secondary customers without harming the primary customer in any way, through means such as anonymization.

4.4 Factors affecting Maturity to Monetize Data to Secondary Customers

The study found several aspects that affected how mature an embedded systems company is to monetize data to secondary customers. Although data monetization to secondary customers varied in the case companies' different contexts and often was case specific, the study found common factors that had an effect in all case companies.

4.4.1 Available Resources

One of the most vital factors for the maturity of embedded systems companies' data monetization to secondary customers was the extent to which they had available resources to pursue it. As the degree of available resources largely determines what competencies a company can involve in the project, how much time can be spent, and what technologies can be used, it is a prerequisite to most other factors identified in the study.

The lowest level of maturity for monetizing data to secondary customers found in the case study was when the core business uses all resources, leaving no available resources to explore how to monetize the data to secondary customers. Four of the case companies were in this stage, making it the most common stage. These case companies were still focusing on utilising data, however only to improve the core business offering through means such as developing AI models for predictive maintenance.

The second level of available resources for monetizing data to secondary customers was identified to be when a company had some involvement from departments with spare capacity. One case company was at this level, and had advanced to this stage relatively recently. At this level, the resources and focus on monetizing data to secondary customers are still scarce. Not being able to have access to the competencies from all departments lead to challenges of not obtaining a cross-functional perspective, thus lacking critical aspects necessary to realize a monetization initiative. At this level, the case company emphasized the importance of having a well-made business case to get traction for data monetization initiatives. However, for the case company at this level, the lack of resources for the initial phases of data monetization to secondary customer initiatives was not always seen as a problem. It was rather about having a strong focus on the company's main objectives and not wanting to pull resources away that would risk affecting their core business. This can be exemplified in the quotes below:

> *We're building this on the side because we do not want to disrupt that [the main] workflow.*
>
> *I think it's a fine balance that I'm walking here. You could have people saying that I'm doing something secretive, which I'm not. I just don't want to come with additional tasks on them, who already have a task and an urgent, very hard timeline to fulfil. So I think at this point, it is a bit secretive, even though it's not. But I think as soon as we have something in place, that will be the time then when I anchor it with the organization.*

The third and most advanced stage observed in the case companies was having a dedicated cross-functional data exploitation team involving data scientists, R&D, a sales group familiar with monetizing data, and support from top management. Among the case companies, one company was currently at this stage. While this team mainly focused on monetizing data to primary customers, they were also exploring monetizing data to secondary customers.

4.4.2 Data Ownership

A fundamental part of monetizing data to secondary customers is data ownership. The extent to which the embedded systems company owns the data they collect from the primary customer greatly affects how much it can be used for other purposes, such as monetizing it to secondary customers. The topic of data ownership was ambiguous, with the view of it differing both between the different case companies, as well as among different employees within case companies. Some case company representatives thought that they owned the data, some thought they had shared ownership of the data with their customers, some thought that the customer owned the data, and one thought that no one can own data. However, all case companies, no matter their view on data ownership, were of the opinion that they owned the insights produced from the data. The different views of data ownership can be seen in the quotes below from three different case companies, in the order of owning the data, the customer owning the data, and finally that none can claim ownership of data:

> *Aah data ownership, the eternal conundrum. We own the data. Possession is two thirds of the law as they say, right? We have the data in our server so we own the data.*
>
> *No, we don't own our customers' data at all. We cannot. We may own the conclusions. It depends. Usually, this is consent that the customer will have to allow.*

This is an interesting legal question. But actually, no one owns data, it will be a very hard legal process to prove that you own data. In Europe at least, you have different privacy rights on data. So I would say no one can really claim ownership of data, you just have different privacy rights on data. So in that regard, you can say you own it, but you can never own it like a physical thing.

Although the view of data ownership showed to be ambiguous, legally unclear, and varied between companies, the most common view of it from the perspective of the embedded systems companies was that they owned the data that they collected, but that they needed customer permission to do so. In relation to this, some of the case companies were seeing that their customers are increasingly valuing data and its potential, and therefore becoming more protective of it. This can be exemplified in the quotes below:

But you know, we have not seen any customer who goes so far as to just say 'take and use my data'. Actually, we get other questions like, are you sure you will use it for this purpose? What purpose? So we see the customers are increasingly protecting and valuing the data. That's where we see the trend.

I think it's also a culture [in Sweden], right? 'I don't want you to use my data'.

They [primary customers] would say, 'Yeah I like your service. I Like all the stuff, but I want the data to be hosted on my servers' because they understand that possession is the key element here.

At the same time, a trend seen in the case companies is that many of them are moving from a data collection "opt-in" approach, to an "opt-out" approach, as a way to legally get data ownership from more customers.

4.4.3 Ability to Anonymize Data

Depending on the nature of the data, the ability to anonymize it can be critical to have a chance to monetize it to secondary customers. The main priority of data anonymization in the case companies was to be compliant with the General Data Protection Regulation (GDPR) and other legal contracts. When the case companies went beyond anonymizing it for legal reasons, it was most often due to the impact on the primary customers' trust and privacy. All case companies believed that if they were to sell data to secondary customers that wasn't adequately anonymized, they would lose that primary customers' business. This reasoning of the case companies can be exemplified in the quote below:

We wouldn't be able to say 'customer XYZ has 30 machines of this model and performing this way with so much availability'. We wouldn't be get-

> *ting much customer trust that way. But if you anonymize the data and say look, you have a ton of data on B47 machines, that's anonymized without telling you which customer and we can show you what availability they are on average, then this might be a business model for data that you can then provide to somebody that can then make use of that data.*

The findings show that the most common way among the case companies to anonymize data is through aggregated averages. In almost all mentioned scenarios, it was considered anonymous enough to monetize it to others, according to the case companies. A few companies also used data masking as a way to make it more anonymous when necessary. All the case companies that were currently offering some kind of insights or data as a service to primary customers, used anonymous aggregated averages unless the data only came from the customer who received the data service.

4.4.4 Knowledge of Secondary Customers

While the majority of case companies expressed an interest in monetizing their data to secondary customers, they differed in how aware they were of who potential secondary customers could be. The large majority of case companies were still in the very early stages of identifying and learning about potential secondary customers. Furthermore, it was seen that the type of secondary customer was highly contextual and varied among the case companies. However, the potential secondary customers were in many cases actors who were also conducting business with the primary customers of the case company, which was the reason they were interested in data of the primary customer.

All case companies thought that there were secondary customers who were interested in the data they collected, even in the cases where the case company itself had no interest in exploring such opportunities. Four case companies were sure that other organizations were interested in data that they possessed, but had difficulties coming up with more specific examples than generic industries. These examples were mainly based on hunches and thoughts with little to no research behind it. Two case companies had started initial discussions with potential secondary customers, but had yet to lay out all the details necessary to get started. The difference in awareness of potential secondary customers can be exemplified in the following quotes:

> *Well, analysts and the stock market, they would want to know [for example] how much paper is being produced by our customers, and that kind of stuff.*

> *And that's where we see potential, because I think working with municipalities in Sweden or Finland, working together with the government, to show how people are [redacted for anonymity]. I think you have a better potential of getting some PoCs or MVPs off the ground, working in smaller collaborations that way with municipalities.*

> *The possibilities [of monetizing data to secondary customers] are limitless. We've had a lot of interest in many different things. During COVID, we produced a transport index. We also posted it publicly showing how traffic movements changed during COVID. Showing border crossings and helping society understand what's happening to the economy and transport flows throughout the world when the pandemic hit, it's very appreciated. Again, there's a lot of new digital entities trying to optimize the transport system that are very interested in our data. You see a lot of opportunities in other sectors than the transport sector as well.*

Among the case companies that had started initial discussions with potential secondary customers, all had market intelligence teams who had been involved with finding these potential use cases. The potential secondary customers in these cases were organizations with which the primary customers had existing relationships with. One of the few case companies that were in discussions about collaborations with potential secondary customers expressed the following about how secondary customers were found:

> *Right now it's [exploring secondary customers] purely direct collaborations or selling. Something good for you to know: If someone tells you that they have a marketplace [for data], that's not true. No one has a marketplace in the industry, neither in the auto industry, nor in the IoT industry.*

4.4.5 Ability to determine the Value of Data

All case companies in the study struggled with how to think about data valuation. It is an area they've given some thought to previously, as it's relevant and useful when it comes to both internal valuation of a department's data output, as well as when providing primary customers with data as a service. The nascent nature of data valuation in the case companies can be exemplified in the following quote:

> *I can tell you that as of today, neither us or many of the industries knows how to do it [data valuation]. I was actually just in another workshop, a cross industry workshop, last Friday, and that was kind of one of our main discussion points. And we're talking about IoT companies, us, software companies that don't have a framework currently.*

Three case companies had yet to attempt to evaluate its data or put a price on it, mainly because they had yet to charge anything for data that they have given to primary customers as a service. In those cases, the case companies were aware that it was only a matter of time before they would have to figure out how to do it, because they realized the increasing value of data, as well as were planning on charging for their data services. These companies' reasoning can be exemplified in the following quote:

> *We are very early in the monetization stage if I put it very plainly. We recognize the value of data, but we don't have a cost against the data, because some companies say you should represent data in your balance sheet and things like that. I mean, we will be there, eventually.*

Among the case companies that were more mature in providing services based on data, their pricing was often value-based. When the data as a service contributed to customers' industrial operations, it was common to base its value on the cost of downtime and cost of failures that it prevented. In these cases, the companies had a strong domain knowledge and a close relation to their customers, which gave them the necessary knowledge of the customer to understand the value of the data for them. An example from one of the case companies with relatively high maturity is:

> *We have to get paid for the value created from the up-time created. We need to take a share in that, we can't get all of it, but a share that is reasonable to the customer. So there is quite some negotiation in this, but we don't talk about price for data. We talk about the value, how can we share the value created jointly with the customer. And if there is no value created, well then, there is no pay in there.*

In other cases where the value of the data was not as easily measured, for example when it can't be directly linked to measures such as cost of downtime, the case companies still recognized the importance to focus on the value the data provides instead of how much it costs to provide. However, in these cases, the data valuation techniques were in their nascence. The current reasoning of the case companies who were starting to explore this can be seen in the quote below:

> *Our stance today is that data relates to the use case. So one type of data can have two different values for two different use cases. It could be very cheap for companies, let's say again, the marketing example, that data could be evaluated into how much revenue these companies will have in time for what they're looking for. When they're going to put 10 advertising spots, they will have a 10 million SEK per year. Compared to if we have a project with [organization] for five years, that will have a different end value. Will they be able to do it without us? Yeah, most likely, will they be able to do it as good without us? Most likely not. So then we can have an improvement of maybe 10, 20, tops 30%. So that's the premium that they will pay for. So then you have 30% of the premium of the total life. So that's how I evaluate.*

Another challenge that was identified in the interviews with the case companies was how the data literacy of the secondary customer affected the difficulty of determining the value of the data for them. Not only did the potential secondary customers' data literacy affect their understanding of the data's value, but it affected its actual value, as it is the secondary customer who creates the value when using the data. The case companies' perception of this can be seen in the quotes below:

We do not know what companies [potential secondary customers] can do. We are talking with many companies. And they don't know. Like the [company name, extracted for anonymity] example that I gave you, they're traditional [company name], they don't have an idea [of how to use the data] [...] You could have way more information [about primary customer] and you go to the [company name] and they're like, 'sorry, what'?

[How to price data] Depends on the maturity. If it's very immature and we just know that there's value at the end of the line. Then we co-explore in a typical win-win scenario so to say, and then we explore a business model together in the most immature exploration we're doing. Whilst in other cases, it's clear how much more efficiency we're gonna get out of it. And it's clear for them also the value and then it's quite simple negotiations in that sense.

4.5 Factors affecting Primary Customers' Willingness to have Data from them Monetized to Secondary Customers

Monetizing data from primary customers to secondary customers requires that the primary customers are willing to have the data generated from them monetized to others. The interviews showed several factors affecting primary customers' willingness to have the data generated from them monetized to secondary customers. It's important to consider these factors to avoid detrimental effects on the relationship with primary customers and consequently, the core business of the company. The factors to consider found in this study include: impacts on the primary customer, trust, incentives, and legal agreements.

4.5.1 Impact on Primary Customers

The first factor, identified in the interviews, affecting primary customers' willingness to have data generated from them monetized to secondary customers is the impact it has on them. If the impact on the primary customer is negative, the case companies expressed that they would not attempt to monetize that data to secondary customers because it could cause the primary customer to end their relationship. An example of a negative impact on the primary customer is if the data to be monetized contains corporate secrets. This is exemplified in the following quote:

But if it [Data about primary customer] comes to what kind of trees you are using, or how much you are producing, then no. That kind of data is production sensitive, and we don't want to share it because our customers wouldn't come to us. We are a provider of something that is in every industry everywhere, you have to be extremely honest and ethical, *and not share data.*

There can also be positive impacts on the primary customers, for example in cases where the secondary customer aims to further improve the primary customer's value offering. This is the case for one of the case companies, that plan to give third-party developers access to data from their primary customers to improve the value offering of their products.

Lastly, it was found in the interviews that companies' brands can affect their willingness to have data generated from them monetized. One case company that has security as a strong part of their brand, expressed that they would be unlikely to monetize their data to other sources regardless of how anonymized the data is because it's not part of their brand. Alternatively, another case company, whose primary customer have safety as a strong part of their brand, expressed that they are more willing to share data to secondary customers if it can help make their entire industry safer, this can be seen in the following quote:

and they share this [proprietary data] with other companies, because it's part of their whole brand of safety.

4.5.2 Trust

The second factor, identified in the interviews, affecting primary customers' willingness to have data generated from them monetized to secondary customers, is the trust between the company and their primary customers. It was clear from the interviews that primary customers would have to consent to having data generated from them monetized to secondary customers. Whether or not the primary customers would consent is greatly impacted by trust according to the interview data, as can be seen in the following quote:

Anything we do, we very transparently communicate with customers, if customers are not convinced, we will not do it.

The perspectives surrounding the role that trust plays in affecting primary customers' willingness to have data generated from them monetized varied among the case companies. There were two perspectives identified from the interview data. Firstly, there were companies that expressed that monetizing data to secondary customers could ruin the existing trust between the companies. This is because the data from the primary customer was collected based on trust and for a specific purpose, as can be seen from the following two quotes:

The customers are giving the data based on trust. So we don't want to breach trust for a business opportunity.

Because then I will be using data for a different purpose than why they *opened it up to me.*

Secondly, there were companies with the perspective that it's the context of the relationship and the existing trust between the companies that decide whether or not primary customers would allow the data generated from them to be monetized to secondary customers. This can be seen in the following quote:

> *I think it [how data can be monetized] is completely about how your business contract and relation is with them. For example, if you're getting something from your suppliers, you have an existing relationship, then they could just let you take the data.*

Trust was clearly a challenging factor with different perspectives surrounding it. The case companies were aware that to overcome the challenges with trust and get consent from the primary customers, they must be able to anonymize the data. The case companies were also aware that being able to anonymize the data would provide new possible business models for data where others also could make use of the data. One case company said the following in regards to monetizing data to secondary customers, which summarizes the general perspective of all the case companies well:

> *It's [the main challenge to reach their goal of owning data] the secrecy. Building trust and making the customer trust that we don't misuse their data. Because just like for individuals, if you get enough data about someone or something, you can learn things that they really don't want you to know. So it's finding a way to anonymize the data and make the customer trust that we treat the data the right way.*

4.5.3 Incentives

The third factor, identified in the interviews, affecting primary customers' willingness to have data generated from them monetized to secondary customers, is incentives. In this context, the case companies expressed awareness that companies rarely do something without getting something in return. One of the case companies explained it well:

> *I think at the end of the day, everybody asks 'What's in it for them?'*

Creating incentives is essential to get consent from the primary customers in cases where there is no benefit for the primary customer to have the data generated from them monetized to secondary customers. One case company, for example, explained that they would need some sort of incentive for the primary customer to allow their data to be acquired in cases where the data won't be used to provide a service directly to the primary customer. Another case company is exploring if they can lower their prices if their primary customer agrees to give them complete control over the data generated from them, which is also a form of incentive. A third case company, explained that they co-explore with the primary customer in a win-win-win scenario, as seen in the following quote:

> *we will continue exploring, as long as we see it's a win for me, you [the primary customer] and whatever customer it's gonna hit.*

4.5.4 Legal Agreements

The fourth factor, identified in the interviews, affecting primary customers' willingness to have data generated from them monetized to secondary customers is legal agreements. Monetizing data can be very complicated and legal agreements play an important role in monetizing data to secondary customers because it enables primary customers to ensure that the data generated from them is not misused. The complexity can be exemplified by the following quote:

> *So it is definitely not that you get the data, and since you have it, you can just give it away. No, that's not the way it works.*

Legal agreements about how companies can use data generated from primary customers is a grey area for most of the case companies. The case companies were aware that there would be legal challenges when monetizing data from their primary customers to secondary customers, but weren't so aware of exactly what the legal challenges would be. The following two quotes show two typical responses when asked about challenges surrounding monetizing data from primary customers to secondary customers, highlighting that the case companies are very aware that there will be legal challenges:

> *How to prove it in contract writing? I think laws and regulations are lagging way, way, way behind here.*

> *Yeah, there will be legal questions, and they need to be solved. The legal hurdles need to be cleared before you take the next step. I think that's the first thing that should be done.*

Two case companies are currently working with legal agreements connected to monetizing data. One of them has an agreement for every product that they sell that their primary customers must sign. The agreement allows them, through privacy approvals, to utilize data in certain ways. The other case company purchases data from a supplier who has rules and regulations on the data, specifying for example how long they can keep the data and who can access it. They must be compliant with these rules and regulations to utilize the data. The legal agreements in these two examples provide legal security that the data won't be misused, indicating how legal agreements can make primary customers more willing to have data generated from them monetized to secondary customers.

5
DSCEM

Our findings show that embedded systems companies have a clear interest in monetizing the data they collect to new, secondary customers. However, none of the case companies have yet to launch an initiative, and most of the companies find it challenging to find business opportunities, use cases, and secondary customers. In response to this, we have developed the Data to Secondary Customer Exploration Model (DSCEM). Due to the high demand uncertainty and the unknown nature of exploring new secondary customer groups, the model takes an agile and lean intrapreneurial approach [30]. Not only is there high uncertainty involved, the findings also indicate that embedded systems companies are facing a wide array of possible opportunities to pursue. Therefore, the model applies the concept of a funnel, similar to Bosch et al. [34], to explore multiple ideas in parallel and effectively find which ideas are worth scaling.

The structure of the model is further motivated by the fact that agile and lean techniques are especially useful in contexts where it is possible to have short development cycles [30], which is aligned with the context of creating services based on data that is already being collected. Furthermore, the model emphasizes a fast speed to market, which is vital when working with information offerings, as the market is fast changing [19]. The study by Buff et al. [19] also shows that successful information businesses in other domains work in environments of heavy experimentation and fail-fast processes, which is well aligned with the literature applied to the model.

5.1 Overview of the DSCEM

The purpose of the DSCEM is to create successful business initiatives by leveraging the unique value of data that embedded systems companies collect and providing it to new organizations outside of their core business. It provides guidance and tools to find and test new opportunities in an unfamiliar context by using an experimental and hypothesis-driven approach. The model thus helps determine which ideas are worth developing full-scale solutions for. An overview of the DSCEM can be seen in figure 5.1.

The model's first phase is ideation, a creative process where the team attempts to come up with as many ideas as possible. The following phase evaluates these ideas in a comparable manner, and prioritizes them in a backlog. From the backlog, the ideas are fed into a funnel. The team's capacity determines how many ideas can

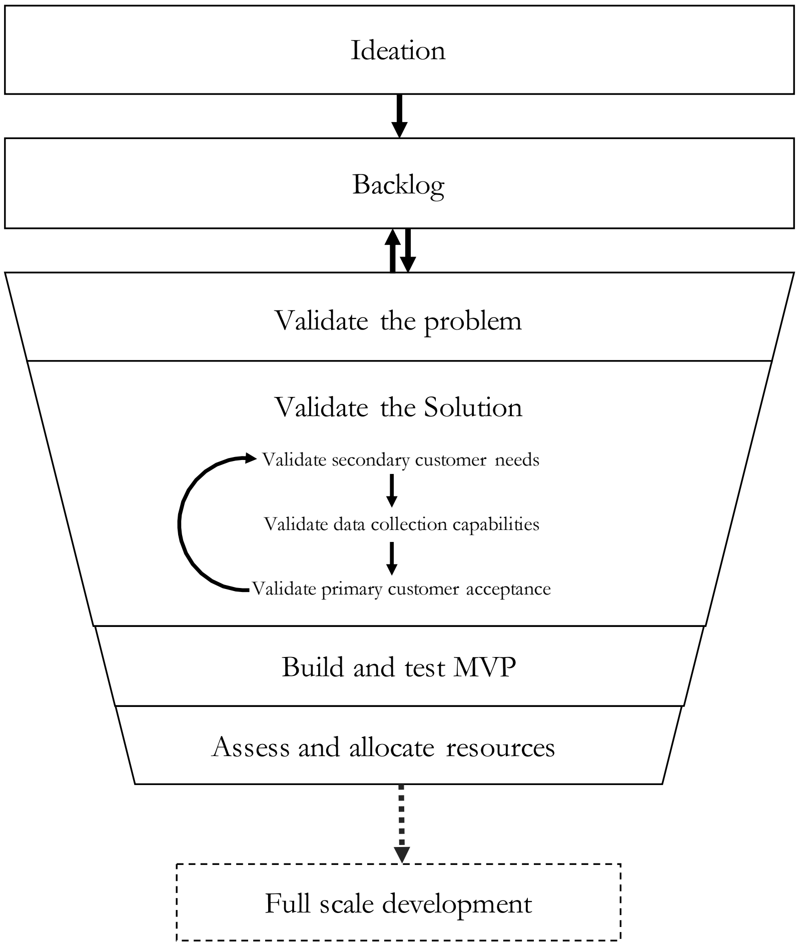

Figure 5.1: Overview of the DSCEM

be explored in the funnel simultaneously. Throughout the funnel, different aspects of the ideas are hypothesized and validated through experimentation. The funnel makes sure that: (1) the idea involves a real problem worth solving, (2) the team has found an appropriate solution to the problem, (3) the organization possesses all the data necessary for the solution, (4) the primary customer, from whom the data is collected, is okay with the data being monetized, (5) the business opportunity can coexist with the core business of the embedded systems company, (6) that an efficient MVP is built that can be scaled, and (7) that the team will have access to the necessary resources to develop the idea full-scale.

5.1.1 Use Cases

To further contextualize and increase the validity of the model, it will be applied in two simulated use cases in the embedded systems domain. They will be described as running examples throughout the explanation of the model. The use cases are derived from real opportunities identified in the case study. Due to limitations of being unable to access data collected by the case companies, a public dataset from embedded systems in automotive vehicles, which have collected data about traffic, driving style, and road condition, is used for the simulation [42]. The dataset is simplistic compared to the real data the case companies collected, which is important to keep in mind for the use cases. However, the dataset is deemed adequate for the purpose of simulating the model. The dataset can be viewed in more detail in Appendix B. The use cases are simulated until the 'Build and test MVP' stage of the model, which was deemed too speculative and beyond the limits of the dataset.

5.2 Prerequisites to the Model

Before it is appropriate to start pursuing business opportunities of developing data offerings to secondary customers, the embedded systems company should consider the following prerequisites. Firstly, the findings show that it is most often appropriate that embedded systems companies should be operating in stage 3 or 4 of the data exploitation dimension presented in section 2.2.2, before attempting to monetize data to secondary customers. The empirical findings show that there is still much progress to be made in regards to data exploitation in their core business. This naturally takes priority as it is the core business that sustains the organization. Only once the embedded systems companies experience enough diminishing returns of using resources to improve their data exploitation in their core business, should they seek new revenue opportunities through secondary customers. However, the findings also show that there exist data exploitation opportunities to secondary customers that require low effort relative to the potential value. These opportunities can be appropriate to exploit before necessarily having reached a high maturity in the core business data exploitation.

Secondly, it was identified that embedded systems companies must be able to anonymize and desensitize the data they collect, if they wish to monetize it to secondary customers. It was clear from the findings that the relationships with the primary customers are too important to risk for a new business opportunity in cases where the data is of a sensitive nature.

Thirdly, data ownership was identified as an important prerequisite before attempting to monetize data to secondary customers. If you're legally unable to monetize data to secondary customers due to ownership agreements, new business opportunities should not be pursued until contracts have been renegotiated to allow this.

Finally, to begin exploring new business opportunities such as this, top management support is required to ensure that the opportunity is in line with the vision of the company and what they wish to do. An example of the importance of top management support from one of the case companies, was when the interviewed CTO was adamant that they would never use their customers' data for other purposes than for that customer. The reason for this was that it wasn't in line with their brand, and that their relationship with primary customers was one of their main competitive advantages. Interestingly, this also applied to data that was not sensitive. This could indicate that it was also a matter of principle and that 'data monetization to secondary customers' can bear negative connotations.

5.3 Ideation

The first stage of the model is ideation, which involves thinking about all the data that the company's embedded systems collect, and generating new ideas for how this data can be leveraged towards secondary customers. Unlike ideation in startup contexts [30], [33], [34], which more freely focuses on finding problems to solve, the ideation in this setting is constrained by the limits of what data the embedded systems company collect. A challenge in the ideation phase in this context, is that it requires embedded systems companies to think about how their data can provide value to actors in domains outside of where their core business operates, which the embedded systems companies are unfamiliar with. A good place to start, based on our empirical findings, is to map and investigate companies that interact with the primary customer base and operate within the same context. From there, the exploration can continue with organizations that interact with the primary customers and operate within a different context. Finally, companies that do not interact with the primary customer but could still leverage the data generated from them can be explored. Figure 5.2 depicts how to map the companies that could benefit from leveraging the data generated from primary customers and in what order the companies are most likely to be able to leverage the data in a valuable way.

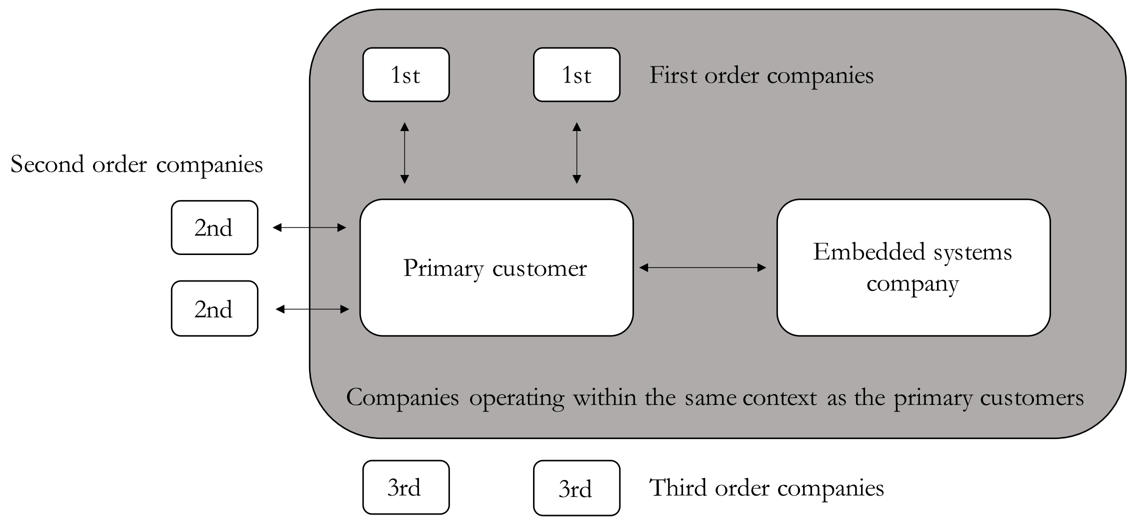

Figure 5.2: Illustration for how to map companies that could leverage data collected by the embedded systems company

Once companies that can leverage the data have been mapped, it's easier to generate ideas with them as guidance. To make the ideas clearly defined and more comparable for later stages, they can be formulated as problem hypotheses in either of the two following forms:
- I believe [type of person/organization] experience [type of problem] while doing [type of task]
- I believe [type of person/organization] experience [type of problem] because of [limit or constraint]

Furthermore, it is recommended to be as specific in the formulations as possible. This will make the ideas more comparable and reduce the risk of confusion amongst the team.

Finally, if the idea seems to involve data that is sensitive to the primary customer, it should be terminated.

Techniques
Brainstorming
The existing literature regarding brainstorming techniques for new business ideas is plentiful [34], [33], [30], and techniques can be picked according to preferences. However, it's important to consider the following questions when applying brainstorming techniques in the context of how data can be leveraged to new, secondary customers:

- What other organizations could be interested in the data?
- Who else interacts with the primary customer and could make use of the data?
- What problems can the datasets solve?

Ideation - Use Case

Brainstorming
Ideas were thought of by mapping interactors of the primary customers (drivers) and brainstorming how the dataset could help them solve potential problems. Figure 5.3 shows the mapping that was done for this use case.

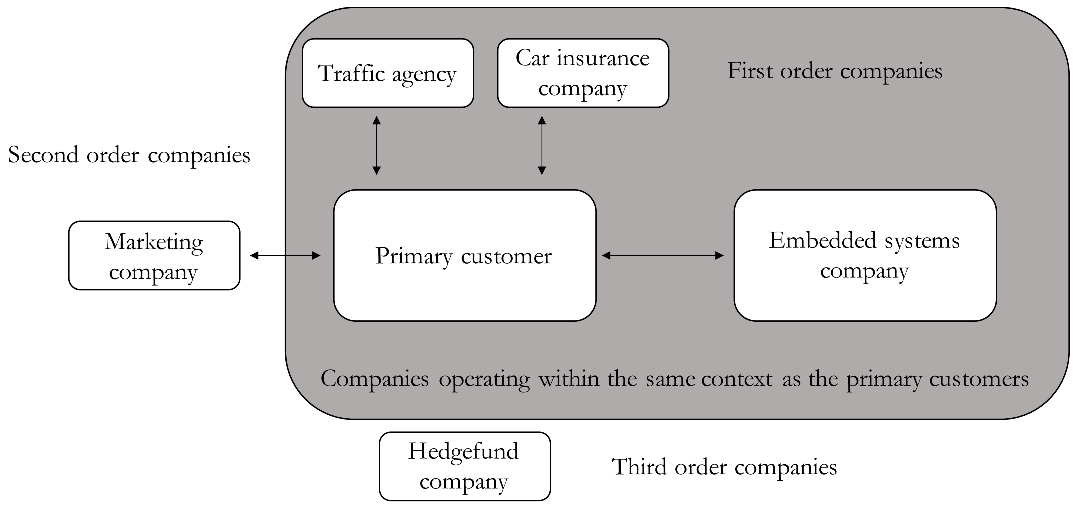

Figure 5.3: Mapping of companies that could leverage the data collected by the automotive company.

For this use case, the two 'first order companies' were chosen because they are assumed to most likely be successful in leveraging the data.

To drive a car you must have car insurance, therefore it's reasonable to assume that car insurance companies could potentially be interested in data about drivers' behaviour to price their insurance more accurately. Data about road conditions can similarly be of interest to the organization that is responsible for the roads, which we assume to be a government traffic agency. The following two problem hypotheses were created from this brainstorming session:

Idea 1: Government traffic agency. I believe municipalities/governmental organizations experience a lack of information on road conditions in Sweden when planning for road maintenance.
Idea 2: Car insurance. I believe car insurance companies experience suboptimal pricing because of a lack of information about the drivers.

5.4 The Backlog

After the ideation stage, all ideas are placed in an idea backlog. The backlog is based on agile principles. In the backlog, the ideas will be placed in a prioritized order based on both value and risk. Each idea should also be assigned an effort, to determine how many ideas can be in the funnel simultaneously. To do this, the ideas must first be made comparable, and then evaluated in relation to each other.

Techniques
Making the ideas comparable
To prioritize the ideas in the backlog, they have to be comparable. As the ideas are in their nascence and not yet thoroughly researched and detailed, this should be considered more of an estimation of their importance. The team should not spend too many resources on attempting to prioritize them in a perfect order. As recommended by Maurya [33], it is useful to document the ideas in the Lean Canvas to make them more comparable. Maurya [33] further recommends updating the Lean Canvas throughout the process to track the ongoing learning.

Prioritizing the ideas
After the ideas are defined in a comparable way, they are evaluated along several parameters to get an understanding of their potential value. Answering the following questions provides a good initial understanding of the value and risks of each idea, which acts as a basis for prioritizing the ideas.

Questions to help estimate the value and risks of the ideas:

- Value
  - How significantly do we think this data would help the secondary customer?
  - How large is the market size?
  - How exclusive is the data?
  - How accurate is the data?
  - How complete is the data?

  - How is the timeliness of the data?
- Risks
  - How sensitive do we think the data is?
  - How familiar are we with the domain?
  - How easy is it to contact potential secondary customers?
  - Are there other actors already offering this service? How competitive and saturated is the market?
  - How technologically advanced do the potential secondary customers seem? How is their data literacy?
  - How difficult does the idea seem technically?

After answering the questions and getting a good understanding of the ideas, the matrix in figure 5.4 is a useful tool for prioritizing ideas. Figure 5.4 is based on traditional impact-effort matrices, a well known tool for prioritizing ideas. It is important to note that the team should not attempt to plot the ideas in an exact way, but more in relation to one another. The initial ideas can be difficult to plot, but the more ideas that are added to the matrix, the better it acts as a tool for prioritization.

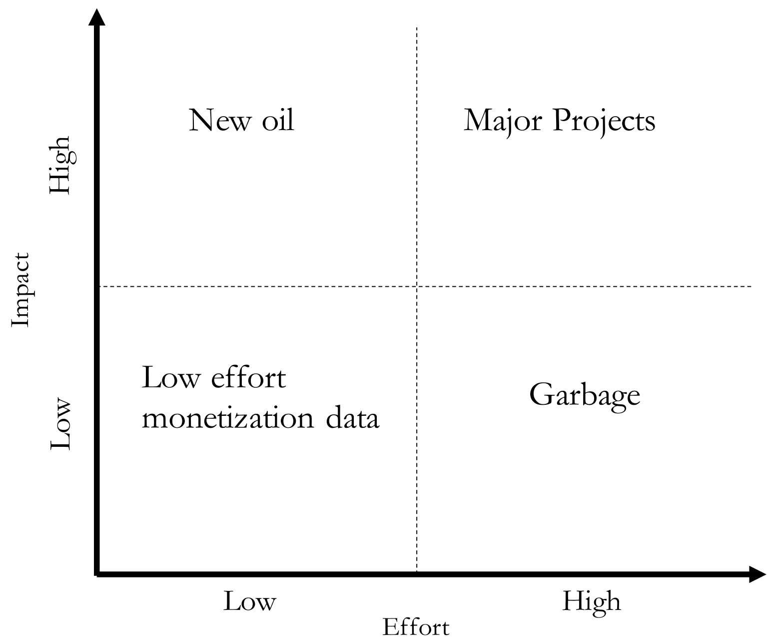

Figure 5.4: Matrix for prioritizing ideas based on the estimated impact and effort.

The Backlog - Use Case

Prioritizing the ideas

The two ideas from ideation are prioritised by answering questions regarding the value and risks and comparing the answers to each other. The answers to the questions can be seen in Appendix C.

Based on the answers provided in Appendix C, the rough estimation indicates that the importance for the government traffic agency idea is relatively low, and the car insurance idea is high. In terms of effort, both the road conditions idea and the car insurance idea is low because it wouldn't require any new data. Relative to each other, the road conditions idea is estimated to require less effort. From this, they are plotted into the matrix as seen in Figure 5.5.

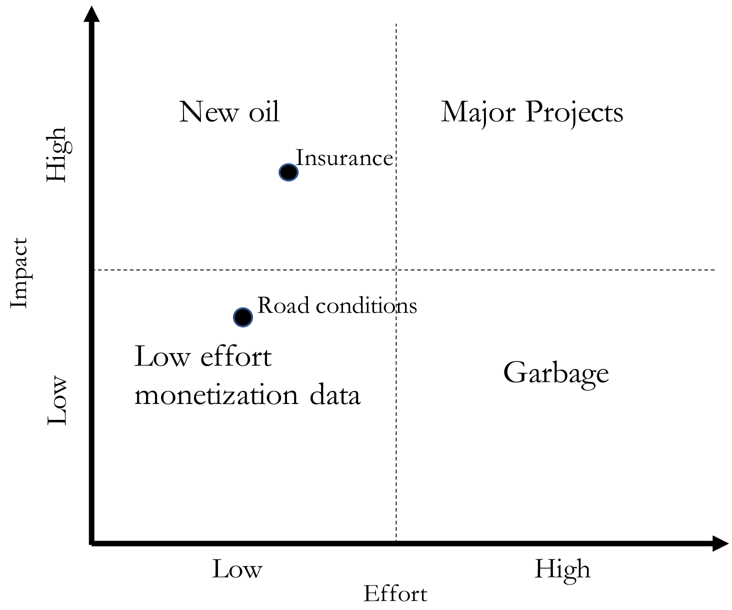

Figure 5.5: Example of using the prioritization matrix.

5.5 The Funnel

Ideas from the backlog are fed into the funnel in accordance with the prioritized order. The capacity of the team determines how many ideas that can be in the funnel simultaneously. As the early stages in the funnel are less resource intensive, it generally contains more ideas in parallel, compared to later stages where ideas are increasingly validated and focused on turning into real, full-scale offerings.

Each stage of the funnel contains a set of exit criteria, techniques, and questions that need to be answered before moving on to the next stage. These are answered through validated learning from BML-cycles [29]. The decision process after each BML-cycle must be disciplined and objective. The DSCEM uses the decision process described in section 2.3.4.3 by Constable [31].

5.5.1 Validate the Problem

Validating the problem is the first stage in the funnel and aims to learn about the problem to determine if it is worth spending resources to investigate solutions for it. This stage answers (1) if the need for a solution to this problem is real? (2) who the secondary customers are? and (3) if the business value of solving the problem is worthy of a new initiative within the organization?

Exit criteria

- The majority of contacted secondary customers express interest to the extent that they are willing to pay for a solution to the problem.
- The majority of contacted secondary customers are willing to participate in the solution validation phase.
- Can list the secondary customers' needs, operations and touch-points to start testing solutions.
- The data necessary to solve the problem can't be obtained as easily through other means.

Techniques

Target potential secondary customers

This technique is useful in cases where there are multiple potential secondary customers with different characteristics, and you seek to identify which is most appropriate to target. The ideation phase will give a general idea of who the secondary customers are. To specifically identify what secondary customer to prioritize, the organization needs to identify innovators and early adopters willing to try incomplete solutions [43]. Alvarez [43] lists the following criteria to help target what secondary customers would be most likely to pursue a new business opportunity:

- Low-tech vs. tech-savvy.
- Low autonomy vs. high autonomy.
- Conservative corporate culture vs. progressive corporate culture.
- Risk-averse vs. risks are rewarded.

The findings of this study have helped identified the following criteria:

- Followers vs. leaders.
- Customers who would benefit the most from a solution to the problem.
- Low vs. high data literacy.
- Connections, if the company or individuals at the company has a connection to a potential secondary customer.

Research targeted potential secondary customers

Research their reason for existence, business models, the relevant decision-making units, and competition. The research should be lightweight and aim to get a better understanding of the secondary customers and appropriate departments to contact for interviews.

Problem interviews

The interviews are useful to learn about the needs and operations of the secondary customer, to truly understand what is valuable to them [25]. During the interview, the interviewer should explore if the hypothesis is actually a problem that the secondary customer needs solved, if there are other problems that the data can solve, and to learn how the secondary customer goes about solving the problems currently.

Validate the Problem - Use Case

For this simulation, both of the ideas from the previous stages go through the funnel in parallel.

Target potential secondary customers

Government traffic agency: The municipalities of Sweden are evaluated across the dimensions of 'Target potential secondary customers'. The purpose is to identify which municipalities have a progressive digital agenda with sufficient digital infrastructure.

Car insurance: By analysing the car insurance industry along the dimensions of 'Target potential secondary customers', it is seen that there exists three main types of companies. The incumbents who currently have the majority of the market are relatively traditional and seemingly unchanging. Therefore, these organizations are not the primary targets, as they are not likely to be as receptible. The mid-tier insurance companies seem to be more open to new ideas and change in order to get an edge over the market leaders. Therefore, they are identified to be somewhat likely to be open to the idea. Finally, there are new car insurance companies that are seeking to disrupt the industry and embrace new technologies and business models. These are the most appropriate targets for an initiative that aims to change how car insurance prices are set.

Research targeted potential secondary customers

The targeted potential secondary customer is researched with the help of information available on the internet.

Government traffic agency: Their reason for existence is to ensure that means of transport work as they should. They do this by being responsible for the long-term infrastructure planning for road, rail, sea, and aviation travel as well as for the construction and operation of state roads and railways. They value safety, the environment, and efficiency. They are funded by the state.

Car insurance: Their reason for existence is to provide insurance to car owners, should anything happen to their cars. Customers pay a monthly cost to be insured. This cost is currently decided by factors such as age, gender, driving history, and area of residence.

Problem interviews

Government traffic agency: There are many possible outcomes from interviewing the traffic agency. The main objectives of the interviews is to get a better understanding of the problem surrounding road condition maintenance, how important a better solution is to them, how they are dealing with the problem today, how they would use the data, and to initiate a relationship with them. The answers from the interview will provide great insights for validating the problem hypothesis and to answer the exit criteria. For the simulation, the outcome of the interviews is that we have learned the following. They are cur-

rently experiencing a lack of information about their roads, and to be able to get more information at a higher frequency would be of great value to them. It was confirmed that they would use this information to better plan their road maintenance work. However, it also came up that they would like to use information on traffic patterns in planning their infrastructure work. The government traffic agency expressed this as a more important problem. Therefore, the idea pivoted to primarily focusing on road infrastructure over maintenance of road conditions. While we are still very familiar with the data that is likely to be involved, we aren't as familiar with how its usage in infrastructure would work. Thus, more interviews were conducted to better understand this. It was found that it would for instance be used identifying what roads are most congested when planning where to add additional car lanes.
Car insurance: The outcome of the problem interviews confirmed that the more modern insurance companies were interested in developing more tailored pricing models towards their customers. It was also confirmed that they did not currently possess the necessary data to do it well. Finally, it was confirmed that they saw a great potential value in a solution that would enable them to do this. They explained what information they require of their customers' driving behaviour, for example how often they were speeding.

Research and problem interviews were conducted until the exit criteria were deemed to be fulfilled.

5.5.2 Validate the Solution

The main purpose of this stage is to learn what data monetization solution fits the problem and the customer, and what features are necessary for an MVP, given the company's available data and primary customers. In order to effectively and efficiently go through the process of validating data monetization solutions to the secondary customers' problems, a fundamental theme throughout the stage is to use demos just good enough to see and understand the customer's reaction and identify what specifications are necessary for an MVP [33].

The range of possible solutions to the problem are limited by the embedded systems company's data collection capabilities and by the primary customers' acceptance of monetizing the data they generate. Therefore, each element of the solution needs to be validated to make sure that it is aligned with what data the company can collect, and what data can be used after consideration of the primary customer. This is done in the Need-Data-Primary (NDP) loop, as illustrated in figure 5.6. The loop consists of three processes: validating the customer needs, validating the data collection capabilities, and validating the primary customer acceptance. The loop is executed iteratively until the exit criteria in the 'Validate the customer needs' stage are fulfilled, and the current iteration of the loop is completed. This ensures that each element of the solution is aligned with the customer needs, the embedded system's company's data collection capabilities, and the primary customers.

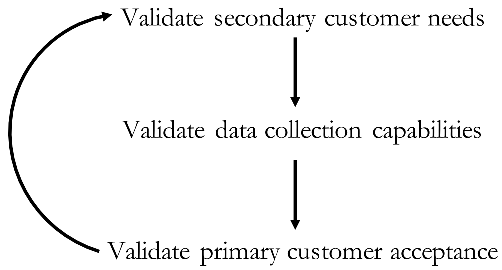

Figure 5.6: The NDP-loop

5.5.2.1 Validate the Customer Needs

The first step of validating the solution, and the NDP-loop, is to hone in on the customer's needs. The main purpose of this stage is to find what solution that best can help the customer with their problem validated in the previous stage. A part of this is to learn about what features are required for a good MVP. As a starting point, it is important to learn which of the three general data as a service offerings, raw and processed, analytics and reporting, or action offerings [19], that best fits the problem and the customer. From there, further details and specifications to the solution can be explored. Each aspect of the hypothesized solution needs to run through the NDP-loop. For example, if conducting a mockup exercise, each element of the mockup should be tested through the NDP-loop. Finally, a rough estimate of the ROI should be established to get a general idea of the financial benefits of pursuing the idea.

Exit criteria

- The majority of contacted potential secondary customer companies are willing to commit time, money, or reputation for the MVP.
- Can explicitly list the customers' requirements on a solution to the problem.
- Have a clearly defined specification list for the MVP.
- The business case meets the required return on investment (ROI).

Techniques
<u>Solution interview</u>
The solution interviews are important and useful to establish a closer relationship with the potential customers. It's an opportunity to test hypotheses about the solution and get a deep understanding of what solution would fit their problem. Important things to learn from the interviews are:

- What do they want to know from the data to help with their problem?
- What type of data offering?
- What protocols and formats would the solution have to be aligned with?
- How much will the customer themselves process and analyze the data after delivery?
- What are the minimum features required before launch?

It is often a good idea to combine the solution interview with a demo of sorts [33], such as mockups or prototypes.

Mockups
Mockups are an appropriate method to test products or features in early stages, especially in information products such as data and analysis [31], [33]. The mockups can help understand the whole user journey of the secondary customer and identify how a user would interact with the data monetization offering [31]. The mockup is a sketch of a hypothesized solution, which can be done on paper, through photoshop, HTML/CSS, etc [33].

Proof of concept
The main purpose of proof of concepts is to test a product or a feature internally to verify that it can be achieved in development, i.e., it's feasibility. For developing data driven services, it is beneficial to develop a real use case of how the data can be used to provide value to the customer. The proof of concept can therefore be done in a short time-frame and must not provide a complete solution to the problem.

Prototypes
When mockups are too simple to test the hypotheses to the potential secondary customer, then prototypes can be a useful technique for demos. The main purpose of the prototype is to as quickly and resource-efficiently as possible learn about what data offering fits the customer. It can for example be partially working, hacked together software, or just a single feature [31]. Prototypes should only focus on learning, and not consider reusability or scalability [31].

5.5.2.2 Validate Data Collection Capabilities

This stage focuses on the data necessary to provide the solution. After validating an element to the solution in the 'Validate the Customer Needs' stage , the company must reflect on their data collection capabilities. The stage aims to answer (1) do we currently collect all the data necessary for the element? (2) if there is more data required for this than what is currently being collected, can we obtain this data? and (3) can we collect the data in a more appropriate way, e.g., more frequently or in another format?

Exit criteria

- Can collect and use all the data necessary to develop the data as a service offering.

Techniques
Validate data collection
First, it must be evaluated if the embedded systems company is currently collecting all the data required for the element of the solution. This can be done by mapping out all of the data required for the element of the solution and comparing it to the

data the the systems collect. If it is not being collected, the company must explore if it can be collected from their embedded systems, and if it is worth the effort. Our findings show that for collecting new data points to be a realistic option for embedded systems companies, they must have an advanced data infrastructure, with versatile and dynamic storage. This is due to the large volumes of data embedded systems companies collect, which means that each additional data point increases data volume significantly. Furthermore, the hardware dependencies of embedded systems companies' products means that costly changes could have to be made to each individual product to make the new data possible to collect. Finally, it must be assessed whether the timeliness of the data can be achieved, for example if the data solution is only valuable in real time, it must be assessed if the embedded systems can provide the solution in real time with the help of technologies such as cloud computing.

5.5.2.3 Validate Primary Customer Acceptance

Primary customer validation is unique to this type of business opportunity where an organization aims to monetize the data they have collected from their primary customers, to new, secondary customers. The core business is what sustains the embedded systems company, and the primary customer needs to come first. Therefore, effects on the primary customers need to be taken into consideration when exploring data-driven business opportunities. The purpose of this stage is to determine if the solution can coexist with the organization's primary customers, from whom the data is generated. It aims to answer questions such as (1) can the necessary data contractually and legally be used for monetization to secondary customers? (2) what are the risks to the core business of using the data?

Exit criteria

- Primary customers are okay with having data generated from them monetized to secondary customers.

Techniques
<u>The LIIT Evaluation Chart</u>
The findings from the interviews indicate that the main factors affecting how willing the primary customer is to have data generated from them monetized are: (1) impact on the primary customer, (2) trust between the embedded systems company and the primary customer, (3) incentives to share the data, and (4) the legal contracts between the embedded systems company and the primary customer. Therefore, the embedded systems company must balance incentives, trust, and the impact the business opportunity will have on their primary customer and core business, to find a solution that can coexist well with the core business. That balance should then be reflected in the legal contracts between the embedded systems company and the primary customer. The impact on the primary customers sets the foundation for how the balance will be achieved. It can be divided into how sensitive the data is, how anonymized it is, and how it affects the embedded systems companies' brand. Sensitivity of the data can be affected by many factors. Data can be sensitive in

the way that it affects the primary customers' competitiveness. However, sensitivity can also stem from the data simply being uncomfortable for the primary customers to have collected or used, even if there is no real risk associated with it. The degree of anonymity determines if the data can be associated with the primary customer, making it fundamental in determining the potential impact. Most of the time, if the data is completely anonymous, the impact will be negligible. Finally, exploiting primary customers' data to secondary customers can affect the potential secondary customers' perception embedded systems company's brand.

If there is potential negative impact on the primary customer, incentives are necessary for the primary customer to be willing to have the data generated from them monetized to secondary customers. In this case, incentives can be indirectly given by reducing prices of existing products and services to the primary customer, or directly given by sharing the revenue received from the secondary customer. Another dimension to incentives is the degree to which the primary customer can monetize the data they generate themselves. The more potential they see in their data, the less likely they are to be willing to have embedded systems companies monetize the data from them.

No matter the degree of impact on the primary customers, it's always important to consider how the business opportunity can affect the relationship between the embedded systems company and the primary customers. This is where the role of the trust between the organization and their primary customers comes into play. Trust is achieved through transparency by ensuring that the primary customers understand what data will be monetized and how it will be used. Trust is also built by setting up limitations and controls on the data that the primary customer is comfortable with.

Figure 5.7 shows a radar diagram for evaluating each aspect of the LIIT chart. This can be used by embedded systems companies to assess how well the idea can coexist with the core business. The model is not only useful for assessing whether to persevere, pivot or perish the idea, but also a tool for comparing options or different ideas.

Renegotiate/rewrite contracts for data ownership

The agreements and contracts in place with the primary customer have been made with the initial relationship in mind and might not cover essential details like who owns the data and how it can be utilised. These will need to be renegotiated and rewritten to start monetizing the data to secondary customers.

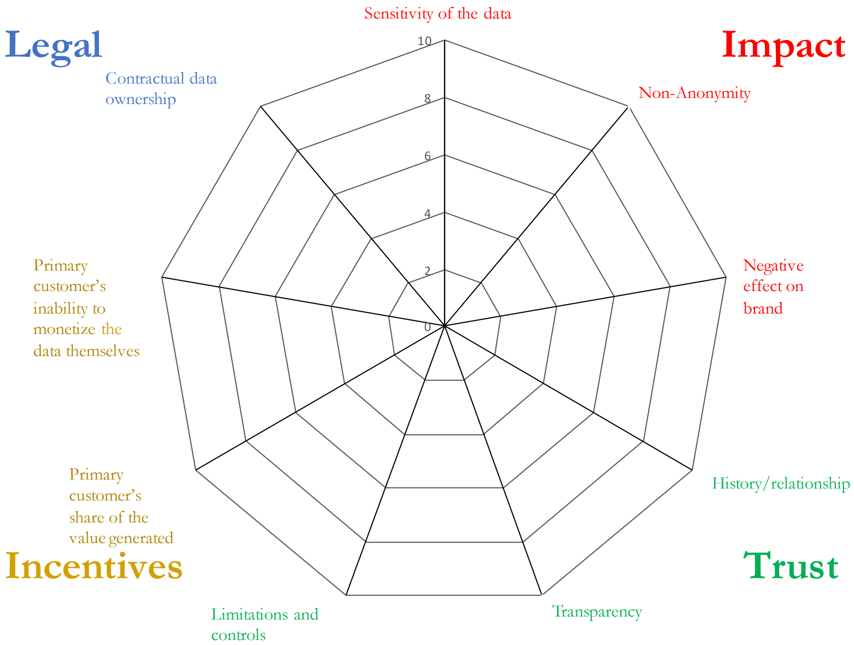

Figure 5.7: The LIIT evaluation chart

Validate the Solution - Use Case

Government traffic agency

To help improve the planning and decision-making of infrastructure works, the potential secondary customer wants to be informed of traffic patterns and road conditions in the area that they are responsible for. Furthermore, a governmental traffic agency is likely not specialized in processing and analyzing raw data. Therefore, we hypothesize that an appropriate type of data offering for this problem is a reporting and analytics offering. To test this hypothesis, a solution interview followed by a mockup is appropriate. This first gives a deeper understanding of what they would want in a solution, followed by seeing how they would interact with an analytics and reporting data offering. Each element of the solution will be validated through the NDP-loop.

Validate data collection capabilities

Required data: GPS, timestamp, and operational asset data. In this simulation, which is based on publicly available data [42], the required operational asset data is seen in Appendix B. In a real context, there are multiple ways to do this and a much larger set of available data. The dataset is purposely kept simple to be used as a proof of concept [42], and the same can be said for this use case simulation.

Available data: In this simulation we have all the operational asset data required. We are missing an appended GPS-location and timestamp. We know from the interview findings that this information can be collected as well.

Validate primary customer acceptance

To assess the impact on the primary customer, the dimensions of the LIIT-evaluation chart is assessed. The data necessary for these ideas is currently already being collected by the cars as part of the fundamental operations and is contractually owned by us. The sensitivity of the data is identified as low but not negligible, as most of the data is asset data, but GPS-location and timestamp could be used to learn personal information about the driver. However, it is possible to anonymize the data by only offering a status of road congestion with a timestamp and not sharing any information about the individual cars. As the purpose of this offering is to improve infrastructure, and therefore improve society, the effect on brand is seen as positive. The primary customer has incentives to share this data, as they would benefit from better infrastructure. Additionally, they have no means to monetize this data themselves.

As the impact is estimated to be low, and there are incentives for the primary customers, the need for trust is marginal, which means that degree of transparency and limitations and controls must not be high. To minimize cost, we continue with minimum viable transparency and limitations and controls through the early phases of the idea. We currently also have a low degree of history and relationship with the drivers, but as the need for trust is low, that is not deemed as a risk. To summarize, the filled out LIIT evaluation chart can be found in figure 5.8.

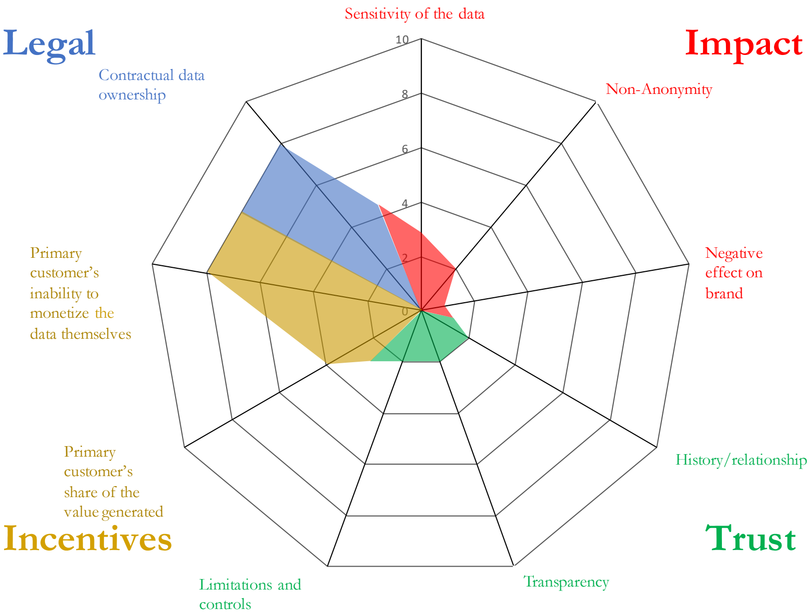

Figure 5.8: The LIIT evaluation chart - Traffic patterns and road conditions

Overall, this initiative is assessed to be low risk for the core business and relationship with primary customers. It can even be seen as a positive for the relationship with the primary customer, as it aims to improve the driver experience.

Car insurance

As this type of insurance pricing has not been done before, the insurance companies are not entirely sure of what they want. From the initial interviews, we learned that they would not know what to do with the raw data. However, they would like to know about the driving style of the driver they insure. From this information, the hypothesis is that a service, which provides the insurance companies with drivers' driving style, would enable them to innovate their pricing models.

Validate data collection capabilities
Required data: Asset data, on an individual and non-anonymized level. Data on the speed limits of the roads that the drivers use.
Available data: Asset data found in [42].
Need to collect what specific car the data comes from to price accordingly and the speed limit on the road the car is on.

Validate primary customer acceptance
Although the data is of a sensitive and private nature, it's only meant for customers willing to be priced based on their driving behaviour. This can be seen to have a positive effect on the embedded systems company's brand because it incentivises drivers to drive safely. Furthermore, there are potentially great incentives for drivers to agree to this, because it could give them a better and fairer price of insurance. Transparency and limitations and controls are very important for drivers to understand exactly how the data will be used to price them, and to ensure that it won't be used for any other purpose. In this case, each driver affected would only be so by signing up on an insurance plan that would have this type of individual-based pricing. They would also be able to quit just as any other insurance plan. To summarize, the filled out LIIT evaluation chart can be found in figure 5.9.

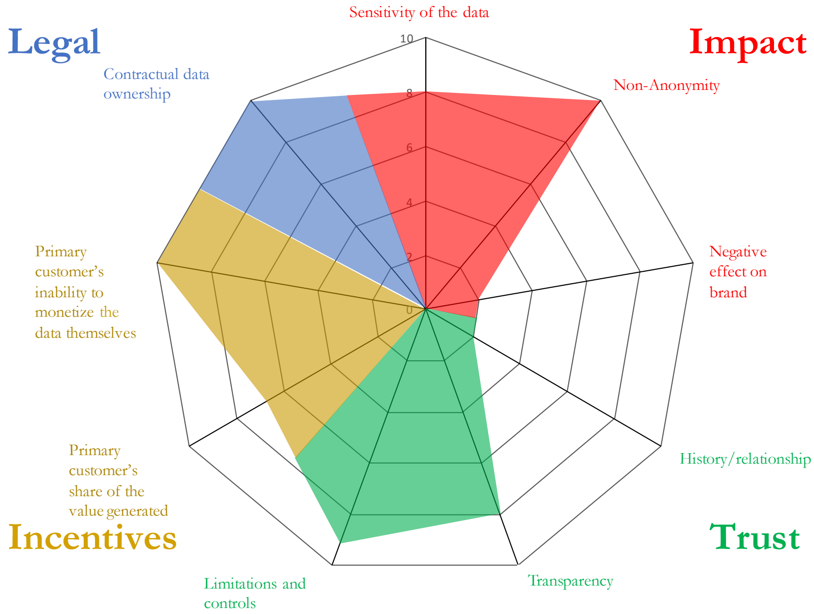

Figure 5.9: The LIIT evaluation chart - Driver Behavior

5.5.3 Build and Test the MVP

Now that the embedded systems company is aware of a solution to a problem worth solving, is aware of how the solution should look, can collect all the data required, and has concluded that it won't negatively affect the relationship with their primary customers, the next step of the funnel is to build and test an MVP. The purpose of this stage is to iteratively build an MVP and learn as much as possible about the secondary customer from their feedback. It aims to answer if the functions of the MVP satisfy the secondary customers' requirements from the previous stages in the funnel.

Exit criteria

- Secondary customer approves of the MVP and all of its functions
- Secondary customer is willing to pay for a full scale solution

Techniques

Lean Software Development Tools

It's easy to get carried away and develop more than necessary for the MVP [33] and thus, waste resources. Therefore, Lean Software Development is an optimal approach to developing the MVP. Lean Software Development focuses on only delivering what the product needs fast, releasing it to the customer for feedback, to then iterate based on this feedback to help guide the decision of whether to persevere, pivot, or perish. There are several tools to apply the principles of Lean Software Development, the following ones are of relevance in the context of delivering data as a service to secondary customers:

- Seeing waste. To mitigate the issue of developing more than you need for the

MVP it's beneficial to be aware of the seven wastes of software development so that you can avoid them: partially done work, extra processes, extra features, task switching, waiting, motion, and defects. This is especially important for data driven services as they can involve extremely large volumes of data.

- Value stream mapping. The second technique to mitigate the risk of developing more than you need for the MVP is to map your value stream. Taking time to ask how activities add value to the customer is an important reflection to determine what's really necessary for the MVP. This is especially important for activities conducted on the data, ensuring that they won't be wasteful in cases where lots of data is involved.
- Options thinking. Due to the highly uncertain nature of developing an MVP in this context it's important to develop in a way that delays decision making as much as possible. Options thinking fits perfectly in this context due to iterations that provide frequent feedback allowing decisions to be made based on facts.
- Pull systems. Pull systems ensure that the features are developed based on what the customer needs and therefore ensures that the MVP is developed as fast as possible.
- Testing. Testing from the very start of developing the MVP ensures that resources are being put on developing the correct features and ensures longevity of the software.

Wizard of Oz-testing

Wizard of Oz-testing refers to when the customer thinks they are interacting with a real product or feature but it's actually being done manually behind the scenes [31]. Wizard of Oz testing can be useful when the real data offering involves advanced algorithms.

5.5.4 Assess and Allocate Resources

Finally, before developing a full scale solution, resources must be assessed and allocated. It should be clear what types of competencies and resources are required to realize the initiative. As seen on the spectrum of the information offerings consumption path [19], the solution can be categorized as either a raw and processed data solution, a reporting and analytics solution, or an action solution. This categorization acts as a starting-point to help the company evaluate what resources and competencies are required for the solution. The further along the spectrum, the more resource- and skill-intensive the solution [19]. Depending on the outcome of the evaluation, the solution can either be aligned with the company's resources and competencies, require additional resources and competencies, require pivoting to a less advanced information offering that is not as resource intensive, or perish if deemed unfeasible with the company's current resources.

As seen in both the findings and related work [24], [26], developing these kinds of services in an embedded systems company can face scepticism and resistance. In this stage, there should now be a clear business case. Therefore, it is appropriate in this stage to anchor the initiative in the organization and convince the organization of the viability of monetizing data to secondary customers.

Exit criteria

- Have the budget estimated to be necessary
- Have the right people for it, such as the following roles:
  - Data management professionals
  - Developers/designers
  - Technical engineers
  - Data scientists
  - Salespeople
- Have a data infrastructure that enables the solution
- The resources are best spent on the solution

Techniques

Calculate ROI

In previous stages of the funnel, only ballpark figures of the ROI can be given. However, in the final stage and before developing a full-scale solution to the customer, it's important to calculate a more detailed ROI to ensure that the solution is priced correctly and that it is supported by a strong business case. This is possible in this stage of the funnel because the company is aware of costs for the data collection, primary customer incentives, and the resources required to develop the full-scale solution.

6

Validation

The feedback from the validation of the model was positive, and the case representatives found the DSCEM to be useful. The validation was conducted in the form of validation interviews, workshops in which the model was applied, and finally, use case simulations. The main subject of discussion and change throughout the validation was the order of the steps in the process. Most of this discussion stemmed from 'catch-22'-like dilemmas, where a piece of information was wanted before a particular stage, but that information could not be obtained before having done that stage. Through iteratively improving the model, and discussing and testing new structures, the authors found a structure that appeared to reach a consensus. The remainder of this chapter presents the key insights made from the validation efforts of the study.

Key insights from the validation interviews

- If there is a good business case behind an idea, it should not be abandoned because there aren't resources available to pursue it. Instead, focus should be on determining how the resources should be gathered to pursue the idea. Therefore, assessing and allocating resources is the final step of the model because the business case is as clear as possible in this stage.
- The validation interviews emphasized the importance of involving the primary customer as early as possible in the process for the case companies to be comfortable with the initiatives, but not too early before there is a clear enough business case to present. Therefore, the primary customers are involved in the second step of the funnel.
- In the validation interviews, it was discovered that companies found it highly difficult to adjust their data collection based on data monetization to secondary customers ideas. They expressed that it would only be possible for those with well-developed and versatile data infrastructure. Otherwise, it would add a load on the organization that was not intended, which they would not be able to handle. This still remains in the model, as it was possible to some extent to collect necessary additional data, and seen as more possible in the future. However, for many organizations, it would currently not be possible.
- The model was deemed as useful by the practitioners for pursuing these types of opportunities.

Key insights from the workshop

- Only a rough estimate of ROI is required for the initial phases of exploration. A more detailed and thorough ROI is needed before continuing with full-scale solutions.

- Discussions surrounding the topic of exit criteria in early exploratory phases. On the one hand, it was argued that it is useful to determine which ideas are worth pursuing and when to abandon ideas and not waste resources. On the other hand, it was argued that they could act as resistance for novel ideas and be used by the more traditional individuals of the organization to hinder change.
- Prior to the workshop, assessing and allocating resources was done before building the MVP. After discussions in the workshop and insights from the validation interviews, it was deemed more appropriate to do after building and testing the MVP, when more is known and when resources are needed for full-scale development.
- The workshop confirmed the usefulness of the NDP-loop. It was seen that data collection capabilities and primary customer acceptance needed to be continuously evaluated and that separating them into distinct sequential phases would be counterproductive.

Key insights from the authors use case simulation

- The use case simulation further contextualized the model, confirmed that the order of the different steps seem feasible, and that the different techniques are appropriate to determine if the exit criteria can be fulfilled.

7
Discussion

This section discusses the findings and contributions from the perspective of the research questions. Furthermore, the section addresses ethical implications and threats to validity.

7.1 How are embedded systems companies exploiting data today? (RQ1)

The results indicate that embedded systems companies are currently exploiting data to enhance their core business, operating in steps one, two, and three in Bosch and Olsson's [4] Data Exploitation Dimension. Although there exists research that have investigated RQ1 already [4], [19], it was deemed necessary to understand the case companies' current state to identify correlations between their data exploitation maturity and their ability to monetize data to secondary customers. Furthermore, understanding each case company's current state allows for richer insights in the validation interviews. Additionally, as data exploitation in the embedded systems domain is a relatively new research area, it is valuable to provide more information about the current state.

The findings of importance to this study, from the perspective of RQ1, are that companies are still exploring new ways to exploit the data they collect and that monetizing data to secondary customers is of interest to companies in the embedded systems domain. It has been explored by some of the companies in this study, but none have yet to do it. This is an important finding because it confirms that monetizing data to secondary customers is an area worth researching and that answering RQ2 and RQ3 are of value to embedded systems companies.

7.2 What factors affect an embedded systems company's ability to monetize data to secondary customers? (RQ2)

To summarize, figure 7.1 shows the three key areas identified to affect how likely data monetization to secondary customers is to be successful.

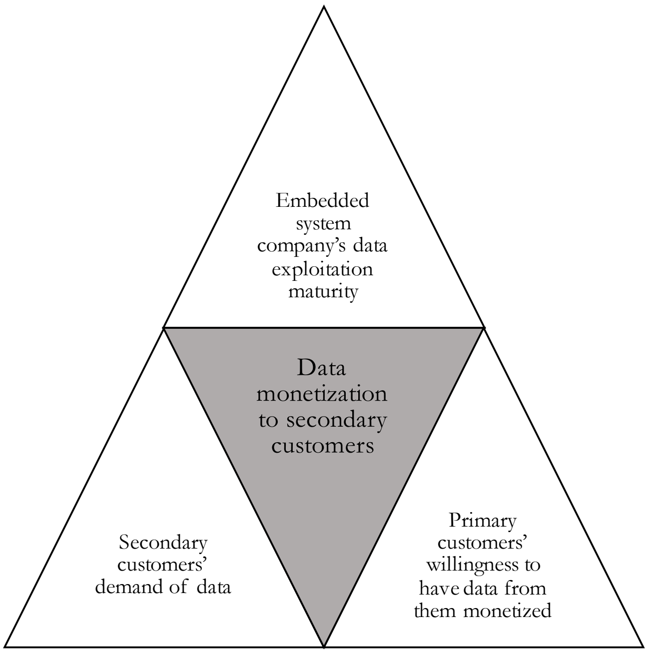

Figure 7.1: The Trinity of Data Monetization to Secondary Customers.

The first area is the embedded systems company's data exploitation maturity. Within this area, there are several factors that have been identified as having an effect on embedded systems company's maturity to monetize data to secondary customers. Bringing awareness to these factors is important because it acts as guidelines for the domain to understand what capabilities need to be focused on as an organization, to successfully monetize data to secondary customers. The results identify both organizational and technological factors important for monetizing data to secondary customers. Due to the factors being primarily organizational, the study focuses more on what embedded systems companies must do on an organizational level, to successfully monetize data to secondary customers. Furthermore, it is useful for academia in identifying new research areas worth exploring to further illuminate the area of monetizing data to secondary customers.

The second area is the primary customers' willingness to have the data generated from them monetized to secondary customers. The factors that have been identified in this area are of key importance because the impact on primary customers was a main concern for monetizing data to secondary customers among all case companies. Proving to the primary customer how monetizing the data generated from them can be beneficial is a difficult task because 'monetizing data' appears to often have a negative connotation. Therefore, this study recommends being transparent to the primary customer about exactly how the data they generate will be used, to whom, and what is in it for them. Approaching the primary customer too early about a data monetization opportunity can cause doubt if these questions can not be answered and have detrimental effects to the monetization opportunity, or worse, to the relationship with the primary customer. However, there is a balance here that needs to be found, because the validation interviews indicate that primary customers will need to be involved as early as possible to mitigate the risk of spending resources on a business opportunity that the primary customers would never agree to.

The third and final area is the secondary customers' demand of the data. For the data to be of real value to the secondary customer, it needs to be difficult to obtain through other means than the embedded systems company. As described in section 2, it is proprietary data that is of most value for monetization [10], which embedded systems companies possess but currently do not leverage to its full potential. The DSCEM provides the embedded systems company with a structured process and tools for rapid validated learning about the secondary customers needs [29], and a better understanding of the secondary customers' value-in-use of the data, which is key for delivering services [25].

7.3 How can embedded systems companies develop services based on data from primary customers, to secondary customer bases? (RQ3)

This research question is answered by the DSCEM, which is one of the main contributions of the study. The model provides guidelines for embedded systems companies to create successful data monetization initiatives to secondary customers worth pursuing with full-scale solutions. It uses the factors identified in RQ2 as a basis to help determine which initiatives to continue pursuing, and which ones to abandon.

DSCEM is inspired by agile, lean, and startup literature [29], [30], [33], [34], due to the high uncertainty related to pursuing initiatives outside of embedded systems companies' core business, as indicated by the findings. The study contributes to existing literature by identifying a type of development methodology appropriate for developing data monetization solutions in the embedded systems domain to secondary customers, and extending and modifying it to this purpose. Novel parts of the model include:

1. *Prerequisites to developing data monetization solutions to secondary customers.* The findings indicate that developing data monetization solutions to secondary customers isn't appropriate for all embedded systems companies, therefore, prerequisites to the model are important for companies who are uncertain if they should explore these opportunities.
2. *Ideating when constrained by the data the embedded systems company collects.* The findings indicate that embedded systems companies lack a structured way of knowing who potential secondary customers could be. This supports the need for a more structured way of mapping potential secondary customers based on the data that the embedded systems company collects, which is provided in figure 5.2.
3. *Prioritizing data monetization ideas.* The findings indicate that embedded systems companies lack a structured way of how to prioritize data monetization ideas. This is provided in section 5.4

by the questions to estimate value and risks.

4. *NDP-Loop.*
   The validation interviews identified the importance of directly considering the data collection capabilities and primary customer acceptance for every aspect of the solution being validated. Therefore, this study proposes the NDP-Loop, which iterates through elements of the solution and validates that the necessary data can be collected for each element, and that the primary customer accepts the element of the solution.
5. *Validating data collection capabilities.*
   The need to validate data collection capabilities to ensure that the data can be collected in the right way for the solution is necessary because embedded systems companies are limited by the data that can be collected. It is also necessary because DSCEM allows for ideas to be pivoted, which can require data outside the limits of what can be collected. The study extends current development process models by adding this as a stage and identifying when it is appropriately executed.
6. *Validating primary customer acceptance.*
   The findings indicate that the effect on embedded systems companies' relationship to their primary customers is the main reason that companies are uncertain about developing data monetization solutions to secondary customers. A contribution of this study is the LIIT chart to validate primary customer acceptance. Additionally, the study contributes by identifying when in the development process this stage is appropriate to conduct.

7.4 Ethics

Embedded systems company's data clearly has value, and this value can be leveraged more than it is currently, by monetizing it to secondary customers. This begs the question: Should embedded systems companies do this? It is a difficult question because data can be seen as an asset just like any other asset, and if you own an asset you are free to do whatever you would want with it. However, the data can also be of a private nature and involve risks in how it is used. As a result of this, there are lots of ethical discussions about the monetization of data. In a study by Someh et al. [44], they rank the concepts that underlie ethical issues in data analytics from organizations' perspective, and the highest ranked concept is data trading. Data trading is defined as "*The extent to which organizations collect, buy, aggregate, share, and sell data from multiple sources in a manner that respects individuals' rights.*" [44, p. 726].

There are three key aspects that can create ethical challenges and need to be considered in an ethical discussion about monetizing data to secondary customers: explicit informed consent, transparency, and anonymity [44]. Firstly, data can often be collected from customers without their explicit informed consent [44], [45]. This was also discussed in a validation interview, where the interviewee explained that it would be possible to collect data about their customers without clearly communicating to them what data they collect and for what purpose. This can be because terms and conditions are often vague about these areas [46]. Private individuals are especially vulnerable, as they don't have entire legal teams to analyze user agreements. Secondly, although it might be disclosed that organizations will use their customers' data in new markets, it can be done with no transparency [47]. Transparency involves providing information about what data will be shared, with whom, and for what purpose [44]. Thirdly, sharing data can risk the anonymity of the data [48]. Although the data being shared can be anonymous, it can be combined with other datasets and thus, breach otherwise private information [49]. In a study by Lewis et al. [50], they studied how students' Facebook profiles changed over time. Privacy was ensured by converting all names to numerical identifiers and removing or encoding all other information that could be traced back to individual students [50]. However, shortly after, researchers discovered it was possible to de-anonymize the data [51]. The same risks exist when monetizing data to secondary customers, further demonstrating the importance of an ethical discussion.

The DSCEM includes steps and processes for embedded systems companies to deal with the ethical implications of monetizing data to secondary customers. In the prerequisites, the model specifies that sensitive data should be avoided and to investigate the legality before attempting to create new business initiatives. Furthermore, the model recommends to validate the primary customer acceptance, identifying the importance of transparency, limitations, and controls on the data. However, for challenges such as explicit informed consent, it's up to the embedded systems companies to act in an ethical manner, because our findings show that there can be either detrimental effects, or no effect, to an embedded systems company's core business should they ignore explicit informed consent. Finally, it's important to understand that GDPR already exists with regulations about how personal data may be utilised and thus, protecting people from their data being misused. In cases where the primary customers are organizations, they are likely to have legal teams able to ensure proper use of the data they generate and won't conduct business with organizations who aim to misuse their data. Therefore, overall, data monetization to secondary customers is a fully viable opportunity for embedded systems companies if it is done with care.

7.5 Threats to validity

This section discusses the validity of the study, to assess the extent of which the research findings accurately represents what was to be studied [37].

7.5.1 Internal Validity

Internal validity refers to the trueness of the results, and the correctness of the conclusions [37]. A threat to the internal validity stems from the nascent nature of monetizing data to secondary customers. As most case companies either haven't started or have barely begun exploring data monetization to secondary customers, there is limited data to collect for assessing factors affecting maturity and success, and for an empirical foundation for developing the DSCEM. To mitigate this risk, the study uses multiple sources of evidence to triangulate the validation, including validation interviews, simulating use cases, and applying the model to real use cases together with one of the case companies.

The potential secrecy of organizations surrounding the topic of monetizing data to secondary customers is another threat to internal validity. It is arguable that more mature organizations, that have an understanding of how insights from data can potentially change their business models, are unable to talk openly about it at this point in time. Therefore, it is relevant to consider how openly they can discuss monetizing data to secondary customers, since it can firstly be a very strategic topic and potentially affect their future business models, and secondly affect their relationships with their primary customers. To mitigate this risk, the first phase of the interviews investigated the company's ability to exploit data in more fundamental ways. This allowed the authors to evaluate the feasibility of the company's ability to monetize data to secondary customers. As the study found that most organizations are still facing challenges such as providing analytics for predictive maintenance, the likelihood of them currently monetizing data to secondary customers but being unable to disclose it, was deemed as very low. In the cases where the embedded systems company had a high maturity in terms of data exploitation with strong data analytics competencies, the authors asked questions to dig deeper into possible implicit data exploitation. The authors saw no signs of undisclosed data monetization to secondary customers in five out of the six cases. The low likelihood of these five companies conducting data monetization to secondary customers but being unable to disclose it, was reinforced by the fact that they had no NDAs involved. However, the case company that was considered most mature in terms of being ready to monetize data to secondary customers, was unable to disclose certain aspects due to NDAs. This means that the possibility of them already monetizing data to secondary customers can't be dismissed.

Finally, it is arguable that the authors signing NDAs with all case companies would allow for potentially deeper and richer insights. However, as this would severely limit what would be able to be included in the study's findings, it was avoided.

7.5.2 External Validity

External validity refers to the generalizability of the results beyond the context of the case study [37]. In this study, the results and contributions should be viewed as explanations and tendencies rather than predictions [52]. In terms of generalizability, the study aims to contribute with drawing on specific implications and rich insights [52].

The case sample can affect external validity [37]. As the case companies were selected through convenience sampling, and all companies interested to participate were accepted to the study, there is a low risk of selection bias on behalf of the authors. Despite this, the selection of cases can be affected by the authors' awareness of case companies and network of contacts, as that had some effect on who was contacted. Furthermore, as most companies who participated in the study did so through a software collaboration organization, there can be bias in what type of company participates in such a collaboration.

There are demographic factors to be considered that can threaten the external validity of the results because the case companies all had offices in Scandinavia. However, the case companies are international and the roles of the participants were diverse. Another threat to the external validity of the study is if there is a difference in data exploitation maturity between the companies willing to participate in the study, and those who were not. The setting of the case also affects the external validity [37], meaning that insights from our case settings may be difficult to generalize to other industries or contexts. Finally, a threat to external validity can come from the small sample size of case companies, due to the time constraints and case company availability. Despite this, the researchers believe that the set of case companies are diverse, which contributes to the degree of generalizability.

7.5.3 Construct Validity

A threat to the construct validity of the study is that the interviewees misunderstand the questions, especially considering the novel nature of the subject. To mitigate this, the authors established a common vocabulary throughout the interview and explained terminology as the terms appeared.

Another threat to construct validity comes from researcher expectancies and bias. There is a risk that researchers can interpret results in their favor and introduce bias to the results. This risk was mitigated by the authors coding the qualitative data separately during analysis. To further mitigate this risk, the codes and themes could have been validated with the interview participants. However, this was deemed unfeasible due to the time constraints of the thesis.

A threat to construct validity of the validation interviews for the model, is that each validation interview started with a short presentation of the model, that possibly could have impacted the interviewees answers and opinions. To mitigate this risk, the model was presented in a neutral way to not influence the interviewees' thoughts. Furthermore, the authors did not express any opinions of the model throughout the validation interview. Another threat to the construct validity of the validation interviews is that the participants answer questions how they think the authors would like them to answer, because it can be uncomfortable to give critique. To mitigate this risk the authors were clear that the purpose of these interviews were to improve the model and asked questions that allowed the interviewees to express their thoughts.

Finally, the construct validity, in terms of choice of data collection methods, could have been increased if the interviews and workshops were complemented with looking into the case companies beyond the interviewees and investigating what is actually happening at the case companies. This would improve construct validity as the perceptions of the interviewees, or the image they want to portray of the company, can differ from reality. However, this was deemed unfeasible due to the degree of company access it required.

8

Conclusion

There is currently unexploited potential in the data that embedded systems companies collect. More value can be generated by sharing it to other stakeholders outside of their core business. There is a clear interest from embedded systems companies to do this, however no organization has yet done it, and only few have started exploring these opportunities. The embedded systems domain faces unique challenges to monetize data to secondary customers due to hardware dependencies, long development cycles, and an established traditional core business sustained by an essential primary customer group. As previous work in other domains, such as the digital domain, do not experience these challenges, there is a gap in knowledge of how to monetize data in the embedded systems domain.

The study found that:

- The embedded systems companies of the study had a low maturity in terms of being able to monetize data to secondary customers. The primary factors affecting the maturity identified in the study were: available resources, data ownership, ability to anonymize data, knowledge of secondary customers, and ability to determine the value of data.
- The main concern of monetizing data to secondary customers was from consideration of the primary customers. The study contributes by identifying four key factors for considering the primary customer: legal agreements, impact on the primary customer, incentives to share the data, and trust. The study further contributes in this regard by providing guidance for assessing these factors and how well the data monetization opportunity can coexist with the core business.
- There lacks a clear way of how to explore opportunities to further leverage the value of the data embedded systems companies possess. The DSCEM offers guidance to embedded systems companies for how to effectively explore data monetization opportunities to secondary customers.

To conclude, the study provides guidance to practitioners and academia with the initial steps of exploring data monetization to secondary customers opportunities and initial conceptualizations of monetizing data to secondary customers in the embedded systems domain.

8.1 Future Work

Due to the unexplored nature of this research area, validation of the DSCEM was limited, especially beyond the 'validate the solution'-stage. Future work could apply the model in a real context with real datasets from an embedded systems company to improve and further validate the model. There is also room for further exploration of how to appropriately build MVPs, assess profitability, and allocate resources for full-scale development related to data monetization to secondary customers . Such research can act as a proof of concept for monetizing data to secondary customers and abolish the negative connotations surrounding the term.

The thesis explores data monetization to secondary customers from a closed innovation perspective. Future work could investigate the possibilities of data monetization to secondary customers through more open innovation.

Bibliography

[1] M. E. Porter and J. E. Heppelmann, “How Smart, Connected Products are Transforming Companies,” *Harvard Business Review*, no. October, pp. 96–114, 2015. [Online]. Available: https://hbr.org/2015/10/how-smart-connected-products-are-transforming-companies?autocomplete=true

[2] H. H. Olsson and J. Bosch, “Data Driven Development: Challenges in Online, Embedded and On-Premise Software,” in *International Conference on Product-Focused Software Process Improvement*, vol. 11915. Springer, 2019, pp. 515–527.

[3] J. Bosch, “Speed, data, and ecosystems: The Future of Software Engineering,” *IEEE Software*, vol. 33, no. 1, pp. 82–88, 2016.

[4] J. Bosch and H. H. Olsson, “Digital for real: A multicase study on the digital transformation of companies in the embedded systems domain,” *Journal of Software: Evolution and Process*, vol. 33, no. 5, p. e2333, 2021.

[5] A. Bharadwaj, O. A. El Sawy, P. A. Pavlou, and N. Venkatraman, “Visions and voices on emerging challenges in digital business strategy,” *MIS Quarterly*, vol. 37, no. 2, pp. 633–661, 2013.

[6] H. H. Olsson and J. Bosch, “Going digital: Disruption and transformation in software-intensive embedded systems ecosystems,” *Journal of Software: Evolution and Process*, vol. 32, no. 6, 2020.

[7] N. Rea and A. Sutton, “Putting a value on data,” PwC, Tech. Rep., 2019. [Online]. Available: https://www.pwc.co.uk/data-analytics/documents/putting-value-on-data.pdf

[8] A. Lewis and D. McKone, “The Edge of Big Data: Enabling New Ways to Create Value,” in *Edge Strategy: A New Mindset for Profitable Growth.* Harvard Business Review Press, 2016, ch. 8, pp. 121–135.

[9] D. Opresnik and M. Taisch, “The value of Big Data in servitization,” *International Journal of Production Economics*, vol. 165, pp. 174–184, 2015.

[10] T. H. Davenport and T. C. Redman, “Your Organization Needs a Proprietary Data Strategy,” *Harvard Business Review*, no. May, pp. 2–6, 2020. [Online]. Available: https://hbr.org/2020/05/your-organization-needs-a-proprietary-data-strategy

[11] A. R. Munappy, D. I. Mattos, J. Bosch, H. H. Olsson, and A. Dakkak, “From Ad-Hoc Data Analytics to DataOps,” in *Proceedings of the International Conference on Software and System Processes.* Association for Computing Machinery, 2020, pp. 165–174.

[12] C. H. Liu and C. L. Chen, "A review of data monetization: Strategic use of big data," in *Proceedings of the Fifteenth International Conference on Electronic Business (ICEB).* Chinese University of Hong Kong, 2015, pp. 40–46.

[13] S. Heath, "What is an embedded system?" in *Embedded Systems Design*, 2nd ed. Newnes, 2002, ch. 1, pp. 1–14.

[14] T. Saarikko, U. H. Westergren, and T. Blomquist, "Digital transformation: Five recommendations for the digitally conscious firm," *Business Horizons*, vol. 63, pp. 825–839, 2020.

[15] E. Fielt, P. Westerveld, K. Desouza, and G. Gable, "Business model innovation and strategic transformation when confronting digital disruption: The case of data-driven business models for professional services," in *Proceedings of the 29th Australasian Conference on Information Systems (ACIS2018).* University of Technology Sydney ePress, 2018, pp. 1–7.

[16] M. Zaki, "Digital transformation: harnessing digital technologies for the next generation of services," *Journal of Services Marketing*, vol. 33, no. 4, pp. 429–435, 2019.

[17] G. Liebel, N. Marko, M. Tichy, A. Leitner, and J. Hansson, "Model-based engineering in the embedded systems domain: an industrial survey on the state-of-practice," *Software and Systems Modeling*, vol. 17, no. 1, pp. 91–113, 2018.

[18] M. Humayun, "Role of emerging IoT big data and cloud computing for real time application," *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 4, pp. 1–13, 2020.

[19] A. Buff, B. H. Wixom, and P. Tallon, "Foundation for Data Monetization," *MIT Center for Information Systems Research*, pp. 1–16, 2015. [Online]. Available: https://www.sas.com/content/dam/SAS/en_au/doc/whitepaper1/data-monetisation-2.pdf

[20] D. Coyle, S. Diepeveen, J. Wdowin, L. Kay, and J. Tennison, "The Value of Data: Summary report 2020," Bennet Institute for Public Policy, Tech. Rep., 2020. [Online]. Available: https://www.bennettinstitute.cam.ac.uk/media/uploads/files/Value_of_data_summary_report_26_Feb.pdf

[21] L. D. Thomas and A. Leiponen, "Big data commercialization," *IEEE Engineering Management Review*, vol. 44, no. 2, pp. 74–90, 2016.

[22] A. S. Bataineh, R. Mizouni, M. El Barachi, and J. Bentahar, "Monetizing Personal Data: A Two-Sided Market Approach," *Procedia Computer Science*, vol. 83, pp. 472–479, 2016.

[23] B. H. Wixom, "Cashing In on Your Data," *Center for Information Systems Research. Research Briefing*, vol. XIV, no. 8, 2014.

[24] D. Kindström and C. Kowalkowski, "Development of industrial service offerings: A process framework," *Journal of Service Management*, vol. 20, no. 2, pp. 156–172, 2009.

[25] C. Grönroos, "Service logic revisited: Who creates value? And who co-creates?" *European Business Review*, vol. 20, no. 4, pp. 298–314, 2008.

[26] I. Gremyr, L. Witell, N. Löfberg, B. Edvardsson, and A. Fundin, "Understanding new service development and service innovation through innovation modes," *Journal of Business and Industrial Marketing*, vol. 29, no. 2, pp. 123–131, 2014.

[27] R. Sherman, "Project Management," in *Business Intelligence Guidebook*. Elsevier Inc., 2015, ch. 18, pp. 449–492.

[28] M. Poppendieck and T. Poppendieck, *Lean Software Development: An Agile Toolkit*. Addison Wesley, 2003.

[29] E. Ries, *The Lean Startup: How Today's Entrepreneurs Use Continuous Innovation to Create Radically Successful Businesses*, 1st ed. Crown Business, 2011.

[30] T. Eisenmann, E. Ries, and S. Dillard, "Hypothesis-Driven Entrepreneurship: The Lean Startup," *Harvard Business School Background Note 812-095*, 2013.

[31] G. Constable and F. Rimalovski, *Testing with Humans: How to Use Experiments to Make Faster, More Informed Decision Making*, 1st ed. Giff Constable, 2018.

[32] A. Croll and B. Yoskovitz, *Lean Analytics: Use Data to Build a Better Startup Faster*, 1st ed. O'Reilly Media, Inc., 2013.

[33] A. Maurya, *Running Lean: Iterate from Plan A to a Plan That Works*, 2nd ed. O'Reilly Media, Inc., 2012.

[34] J. Bosch, H. H. Olsson, J. Björk, and J. Ljungblad, "The early stage software startup development model: A framework for operationalizing lean principles in software startups," in *International Conference on Lean Enterprise Software and Systems*, vol. 167. Springer, 2013, pp. 1–15.

[35] A. C. Edmondson and S. E. Mcmanus, "Methodological fit in management field research," *Academy of Management Review*, vol. 32, no. 4, pp. 1246–1264, 2007.

[36] S. Easterbrook, J. Singer, M. A. Storey, and D. Damian, "Selecting empirical methods for software engineering research," in *Guide to Advanced Empirical Software Engineering*. Springer, 2008, ch. 11, pp. 285–311.

[37] M. Easterby-Smith, R. Thorpe, and P. R. Jackson, *Management and Business Research*, 6th ed., K. Smy, Ed. Sage Publications, 2018.

[38] R. Yin, "Case study research: Design and methods, applied social research," *Methods series*, vol. 5, 1994.

[39] P. Runeson, M. Höst, A. Rainer, and B. Regnell, *Case Study Research in Software Engineering: Guidelines and Examples*. John Wiley and Sons Ltd, 2012.

[40] V. Braun and V. Clarke, "Using thematic analysis in psychology," *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77–101, 2006.

[41] J. A. Maxwell, *Qualitative research design: an interactive approach*, 2nd ed. Sage Publications, 2005.

[42] M. Ruta, F. Scioscia, G. Loseto, A. Pinto, and E. DI Sciascio, "Machine learning in the Internet of Things: A semantic-enhanced approach," pp. 183–204, 2018.

[43] C. Alvarez, *Lean Customer Development: Building Products Your Customers Will Buy*. O'Reilly Media, Inc., 2014.

[44] I. Someh, M. Davern, C. F. Breidbach, and G. Shanks, "Ethical issues in big data analytics: A stakeholder perspective," *Communications of the Association for Information Systems*, vol. 44, pp. 718–747, 2019.

[45] A. Abbasi, S. Sarker, and R. H. Chiang, "Big data research in information systems: Toward an inclusive research agenda," *Journal of the Association for Information Systems*, vol. 17, no. 2, pp. i–xxxii, 2016.

[46] M. R. Wigan and R. Clarke, "Big data's big unintended consequences," *Computer*, vol. 46, no. 6, pp. 46–53, 2013.

[47] K. E. Martin, "Ethical Issues in the Big Data Industry," *MIS Quarterly Executive*, vol. 14, no. 2, pp. 67–85, 2015.

[48] S. Zuboff, "Big other: Surveillance Capitalism and the Prospects of an Information Civilization," *Journal of Information Technology*, vol. 30, no. 1, pp. 75–89, 2015.

[49] S. Barocas and H. Nissenbaum, "Big Data's End Run around Anonymity and Consent," in *Privacy, Big Data, and the Public Good: Frameworks for Engagement.* Cambridge University Press, 2014, ch. 2, p. 44–75.

[50] K. Lewis, J. Kaufman, M. Gonzalez, A. Wimmer, and N. Christakis, "Tastes, ties, and time: A new social network dataset using Facebook.com," *Social Networks*, vol. 30, no. 4, pp. 330–342, 2008.

[51] D. Boyd and K. Crawford, "Critical questions for big data: Provocations for a cultural, technological, and scholarly phenomenon," *Information, Communication & Society*, vol. 15, no. 5, pp. 662–679, 2012.

[52] G. Walsham, "Interpretive case studies in IS research: nature and method," *European Journal of Information Systems*, vol. 4, no. 2, pp. 74–81, 1995.

A

Topic Guide

Interview Preparation

- Research the case company, the practitioner's LinkedIn, and projects they currently are involved in.

Interview Questions

Opening questions (5 min)

- Purpose of this interview
- Can we record this interview in order to use it for analysis?
- Could you tell us a little about your background and your role at [Company]?

Questions around key topics (50 min)

Collecting and processing data

- What type of data is currently being collected by your company?
  - From when you started working at [Company], how has the type of data you're collecting changed? And why has it changed, do you think?
  - What are the plans/goals for future data collection?
- Who typically owns the data that you collect?
- How do you decide what data to collect?
  - How do you know if it's valuable?
  - Challenges?
- How is the data processed from the point of collection to when it gets used?
- Do you receive/get external data from other stakeholders?
  - What kind of data?
  - How do you use it?
  - How did this exchange of data start?
  - What do you give in exchange for the data?
  - What were/are the challenges?
  - What could be better?
    - If No: Do you think there is data from other stakeholders that could be of value to [Company]?
    - If yes: What is stopping you from getting it?

Data exploitation

- How is your data currently being used (Challenges and Opportunities), how are you creating value from it?
  - How has the usage at [Company] of data changed since you joined?
  - How are you planning on using the data in the future?
  - What are your goals?
- Is [Company] currently sharing data with other stakeholders? Who, why, and how?
  - Yes: What do you get in exchange for the shared data?
- Who could be interested in your data?
- Are you currently exploring any opportunities for sharing data with other stakeholders?
  - How?
  - Yes: Could you give an example?
- What do you think would need to be done in order to start monetizing data to new secondary customers who don't actually use your products but are interested in the data?
  - What is currently preventing you?
  - What resources or competencies do you need to do this?
  - What/whose support within the company would be needed?
- Do you think the customers from whom you collect data, would mind that you generate value from it to other types of customers? In which situations? Why?

Closing questions (5 min)

- So to conclude, if you could, with the wave of a magic wand, solve any problem related to monetizing data to secondary customers, what problem would you choose?
- Thanks
- Next step is to build a model with all the interview data, which we will validate through interviews. Would you be interested in helping us validate the model?

B

Automotive Vehicle Dataset

General information

The dataset can be studied in the following URL: `https://www.kaggle.com/gloseto/traffic-driving-style-road-surface-condition`

- Low-level parameters acquired by the car via OBD-II and through the micro-devices embedded in the user smartphone, with the goal of accurately characterizing the overall system composed by driver,vehicle and environment
- Predicted attribute: road surface, traffic and driving style

Data related to the following cars:
- Peugeot 207 1.4 HDi (70 CV)
- Opel Corsa 1.3 HDi (95 CV)

Attributes

- altitude change, calculated over 10 seconds;
- current speed value; average speed in the last 60 seconds;
- speed variance in the last 60 seconds;
- speed variation for every second of detection;
- longitudinal acceleration, measured by the smartphone accelerometer and pre-processed with a low-pass filter;
- engine load, expressed as a percentage;
- engine coolant temperatures in celsius degree;
- Manifold Air Pressure (MAP), a parameter the internal combustion engine uses to compute the optimal air/fuel ratio;
- Revolutions Per Minute (RPM) of the engine;
- Mass Air Flow (MAF) Rate measured in g/s, used by the engine to set fuel delivery and spark timing;
- Intake Air Temperature (IAT) at the engine entrance;
- vertical acceleration, measured by the smartphone accelerometer and pre-processed with a low-pass filter;
- average fuel consumption, calculated as needed liters per 100 km.

C

Use Case - The Backlog Prioritization Questions

Value:

- How significantly do we think this data would help the secondary customer?
  Government traffic agency: The data would likely help them significantly in increasing their response time and accuracy in their road maintenance work. However, it is not business critical but rather a nice-to-have.
  Car insurance: It can provide them with a more accurate way of pricing their insurance plans. As this could significantly change their way of pricing insurances, there is high potential value in this idea.
- How large is the market size?
  Government traffic agency: A large institution with strong buying power. Therefore, while there is only one actor, the market is still large.
  Car insurance: The insurance industry is a large market with a strong buying power.
- How exclusive is the data?
  Government traffic agency: There are multiple automotive companies that are able to provide data on road conditions.
  Car insurance: There are multiple automotive companies that are or will soon be able to provide data on driver behaviour.
- How accurate is the data?
  Government traffic agency: This would need to be assessed within the organization by asking someone knowledgeable in the embedded systems company. Accuracy of the data would in this case be a source of competitive advantage, if other automotive companies were to develop similar data as a service in the future.
  Car insurance: Same as above.
- How complete is the data?
  Government traffic agency: The size of the fleet collecting the data, as well as the sensor equipment on the cars would determine the completeness of the data. For this idea, a high completeness is of great value, as the governmental traffic agency would likely want a complete overview of road conditions over a wide geographical coverage. As the government traffic agency needs to consider all car users within their municipality, the fact that we only can provide data from cars with our software and not represent everyone in the municipality is a risk that needs to be assessed in the problem solution.
  Car insurance: Same as above. However, as the data potentially could be of a

more sensitive or personal nature, the completeness of the data could also be affected by the extent to which the primary customer agrees to have the data monetized.

- How is the timeliness of the data?
  Government traffic agency: The value of data about road conditions is greatly affected by timeliness. It is likely that a real-time data stream or a high frequency batch of data would be most appropriate to allow the traffic agency to quickly react to road conditions.
  Car insurance: The timeliness of data for pricing insurance rates is likely not as important, as long as it is sent in a reasonable time frame.

Risk:

- How sensitive do we think the data is?
  Government traffic agency: As long as the data is only about road conditions, and won't be connected to any personal data of the driver, or exact timestamp of when the car was on that specific road, it will probably not be sensitive at all.
  Car insurance: The data could be quite sensitive. Especially if they would want data on their individual customers to determine the insurance rates. However, anonymized and aggregated driver behaviour data linked to age or certain cars could perhaps be of interest to them as well. That data wouldn't be as sensitive. If instead it is found that individual information is what is required, then the data will be very sensitive. This can be mitigated if the primary customer completely understands and actively accepts its use.
- How familiar are we with the domain?
  Government traffic agency: We are quite familiar with the domain. As we are experts at automotive vehicles and how they interact with roads, we know and understand road conditions well.
  Car insurance: We are very familiar with data of driving behaviour and car usage as our software is already heavily focused on it for our core business. We are also quite familiar with the insurance companies' usage of the data.
- How easy is it to contact potential secondary customers?
  Government traffic agency: Easy to contact.
  Car insurance: It is easy to initiate discussions with the insurance companies, as they seem interested in learning more about their customers.
- Are there other actors already offering this service? How competitive and saturated is the market?
  Government traffic agency: Not that we are aware of.
  Car insurance: Not that we are aware of.
- How technologically advanced do the potential secondary customers seem? How is their data literacy?
  Government traffic agency: They have medium data literacy, but are currently undergoing a digitalization journey and are looking for ways to become more digital. However, the data literacy varies significantly amongst different municipalities.
  Car insurance: They have a high data literacy and are used to working with large amounts of information and statistics. However, their digital competen-

cies can in some cases hinder their receptiveness of data-as-a-service.

- How difficult does the idea seem technically?
  Government traffic agency: The idea seems feasible as the cars already have to analyze road conditions when in motion. However, datasets that are currently used separately would have to be combined to analyze road conditions.
  Car insurance: We are currently collecting and using data related to driver behavior. However, the idea would require datasets to be combined to provide insights about driver behavior, which is not currently being done.