Aalto University
School of Science
Master’s Programme in Industrial Engineering & Management

Antti Ihalainen

The role of data ownership in a strategy process: case study from cybersecurity industry

Master’s Thesis

Helsinki, February 28, 2022

Supervisor:	Timo Vuori, Professor
Thesis advisor:	Ilpo Ruohonen, Licentiate of Science (Technology)

AALTO UNIVERSITY
School of Science
Industrial Engineering and Management

ABSTRACT OF THE
MASTER´S THESIS

Author: Antti Ihalainen	
Title of the thesis: The role of data ownership in a strategy process: case study from cybersecurity industry	
Number of pages: 91+4	Date: 28.2.2022
Major or Minor: Strategy and Venturing	
Supervisor: Prof. Timo Vuori	
Thesis advisor: Lic.Sc. (Tech.) Ilpo Ruohonen	

Cloud-based software has taken market share from on-premises software in recent years. With centralised data storage in cloud-based software, the software user doesn't physically hold their own data in their premises as the data is stored by the cloud service provider instead. In parallel with digitalisation and technological development, the strategy literature has generally focused on utilizing data as part of the strategy processes, leaving a research gap for data ownership aspect.

This thesis examines the role of data ownership from the perspective of corporate strategy processes. The research has been carried out in collaboration with a Finnish B2B cybersecurity software company who is interested in the cloud market to support its business in addition to its strong position on-premises software market. A total of 16 people were interviewed for this thesis, both internally within the case company and from its surrounding stakeholders. The interviewees represented different backgrounds, covering both strategy professionals and security experts.

Based on the results of the work, organizations can stand out in their competitive field by owning rare and difficult-to-imitate data, thus helping to create a competitive advantage as part of their strategy processes. In addition, given today's networked value chains and cross-organizational collaboration models, organizations can expand their own and their stakeholders' data assets by sharing data within their own strategic and operational ecosystems.

Keywords: Strategy process, cybersecurity, data ownership	Publishing language: English

AALTO UNIVERSITY
Perustieteiden korkeakoulu
Tuotantotalous

DIPLOMITYÖN
TIIVISTELMÄ

Tekijä:
Antti Ihalainen

Työn nimi:
Datan omistajuuden rooli strategiaprosessia: tapaustutkimus tietoturva-alalta

Sivumäärä: 91+4	Päiväys: 28.2.2022

Pääaine:
Strategy and Venturing

Työn valvoja:
Prof. Timo Vuori

Työn ohjaaja:
Lic.Sc. (Tech.) Ilpo Ruohonen

Pilvipohjaiset ohjelmistot ovat ottaneet sijaa paikallisesti asennetuilta ohjelmistoilta viime vuosien aikana. Pilvipohjaisten ohjelmistojen keskitetyn datan säilytyksen myötä ohjelmiston käyttäjä ei fyysisesti omista omaa dataansa omissa tiloissaan, sillä data säilytetään pilvipalvelun tarjoajan toimesta. Digitalisaation ja teknologisen kehityksen kanssa samanaikaisesti strategiakirjallisuus on keskittynyt yleisesti datan hyödyntämistä osana strategiaprosesseja jättäen aukon tutkimukselle datan omistajuuden osalta.

Tässä diplomityössä tutkitaan datan omistajuuden roolia yritysten strategiaprosessien näkökulmasta. Tutkimus on tehty yhteistyössä suomalaisen tietoturva-alan B2B-ohjelmistoyrityksen kanssa, joka on kiinnostunut pilvimarkkinasta tukeakseen liiketoimintaansa vahvan paikallisesti asennettavien ohjelmistojen asemansa lisäksi. Työtä varten haastateltiin yhteensä 16 henkilöä niin kohdeyrityksen sisäisesti kuin sen ympäröivistä sidosryhmistä. Haastateltavat henkilöt edustivat eri taustoja kattaen sekä strategia-ammattilaiset kuin tietoturva-alan asiantuntijat.

Työn tulosten perusteella omistamalla uniikkia ja vaikeasti jäljitettävää dataa organisaatiot pystyvät erottautumaan kilpailukentässään täten edesauttaen kilpailuedun luomisessa osana strategiaprosessejaan. Lisäksi ottaen huomioon nykypäivän verkottuneet arvoketjut ja organisaatioiden väliset yhteistyömallit, jakamalla dataa yrityksen oman strategisten ja operatiivisten ekosysteemien sisällä organisaatiot voivat laajentaa omaa ja sidosryhmiensä tietovarantoja.

Asiasanat: Strategiaprosessi, tietoturva, datan omistajuus	Julkaisukieli: Englanti

Acknowledgements

The last six years have flown past very rapidly and they have contained a lot of great moments to be remembered thanks to PoRa and Prodeko.

I want to thank Ilpo and Joona for introducing me to the extremely fascinating cybersecurity industry and your valuable support during the whole thesis process. Also, a big thanks to all the interviewees for the interesting discussions related to very topical themes and people who connected me with the interviewees. Moreover, I would also like to thank Timo for supervising and assisting to put this thesis into an academically compelling format.

Most importantly, I would like to thank my friends and family for not only supporting me during the thesis process but also during the whole studies.

Espoo, 28 February 2022

Antti Ihalainen

Table of Contents

List of Figures

List of Tables

Abbreviations

AI	Artificial Intelligence
GDPR	General Data Protection Regulation
ML	Machine Learning
NS	Network Security
PAM	Privileged Access Management
RBV	Resource-Based View
SaaS	Software as a Service
SME	Small and Medium-sized Enterprises
VRIO	Value, Rarity, Imitability and Organization

1. Introduction

Companies are continuously looking for ways to grow their businesses. There exists a large box of tools to achieve the growth for example via enlarging the sales to new customers in the current markets, expanding the business with the current customer base, entering a new business vertical or leveraging mergers and acquisitions. All the tools have typically some strategic reasoning behind that has led to executing these growth initiatives. However, the most widely used strategic frameworks have been created in the last century meaning that they might lack some of the concepts of today's competitive environment, especially in the industries that have developed in the last few decades, such as information technology and cybersecurity.

Considering the growth of the cloud services used in the organizations and the multi-company networks where the companies operate in, data ownership point of view has lacked attention in strategy processes when organizations are conducting their data-driven decision-making within their strategy formulation. Cloud-services, connected industrial devices, and other data generating and storing sources have their own terms and conditions when it comes to ownership and sharing of the data (Birch et al., 2021).

Decision-makers are continuously adapting data-analysis as part of their processes (Miller & Mork, 2013). However, less focus has been paid to the ownership of the data and how the data ownership influences the decision-making process.

1.1 Background

The ICT industry is generally moving from on-premises customized software solutions to on-demand and off-the-shelf SaaS ("Software-as-a-Service") models where the software is delivered over the internet and updated on a regular basis. At the same time the SaaS model provides benefits for vendors, its investors, and customers via predictable cash flows. Therefore, the customers can acquire software without major upfront investments due to monthly billing and scalability because of lack of customer configuration. However, there remain some pockets in the ICT market that are less eager towards moving to cloud environments due to security, privacy, and data ownership aspects.

Cybersecurity industry is one of those segments in the ICT market that has adopted cloud and SaaS models less than the other segments as cloud and SaaS models typically require customers to share their data to some extent with the vendor. Customers typically identify the cybersecurity data strictly confidential which means that it's more challenging to be shared with or stored by a third party and thus the cybersecurity industry has been unable to move completely to the cloud environment.

Although there exists some resistance towards moving the cybersecurity functions to cloud, on the other hand centralized data collection would enable recognizing unexpected behavioural patterns in organizations' IT environment earlier by utilizing the data from the other users' environment. Centralized data could respond to modern cyberattacks with greater extent as the criminals increasingly take advantage of artificial intelligence-based tools. This naturally leads to the question whether cybersecurity vendors should themselves make more use of artificial intelligence-based software, that basically would benefit from centralized data collection, to combat these attacks. Morgan (2019) predicted that the frequency of ransomware attacks would increase from 14 seconds at the end of 2019 to 11 seconds by 2021 while the corresponding frequency in early 2016 was every two minutes. Considering the generality and growth of cyberattacks, centralized cybersecurity databases could enable preventing and responding to attacks to a larger extent than in a scenario where every cybersecurity customer takes care of just their own cybersecurity.

However, some industries, like finance, health, and militaries, are extremely fragile to give any permission to third parties to access their data partly due to a regulative environment and frameworks such as General Data Protection Regulation (GDPR). This strategic choice of customers forms demand to on-premises software solutions providers and thus the software vendors are required to make a strategic choice whether to be present in that market or focus their resources to some other markets.

The rise of the cloud has raised the question over data ownership, who has access to the data and how the data should be shared. This point of view is less discovered in the strategy literature where a typical approach is to explore the market environment and ponder how they can build a sustaining competitive advantage to grow and succeed in the market with company's resources when building corporate strategies (Barney, 1991).

1.2 Objectives and research questions

This thesis has been done in cooperation with a Finnish cybersecurity company that has over a 25-year long history of developing business-to-business cybersecurity software products. The aim of this research is two-edged by contributing both case company's understanding related to the topic and academic research. Firstly, this study accumulates case company's understanding of market entries and how it should position itself in the cloud era as a challenger in the cybersecurity market. The tech giants and other big players are pushing their cloud-based software and cybersecurity solutions to the market which means that the challengers must analyse their competitive edge in the rapidly growing and changing market environment.

Secondly, this thesis contributes to current academic research by combining strategy, data ownership and cybersecurity topics into one study. The intersection of this study is presented in Figure 1. The combination of these three fields have had only a limited amount of research in the past and therefore this provides understanding around the topic for academia, cybersecurity vendors and their customers.

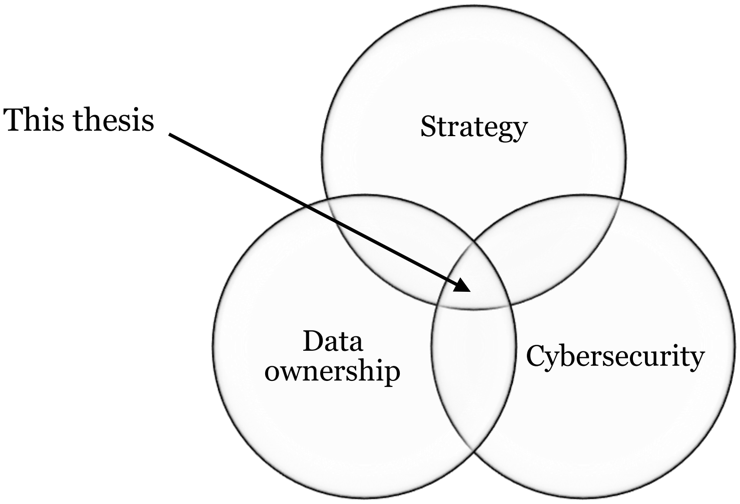


Figure 1: The focus of this thesis

There are three research questions in this thesis to research the topic. The questions are divided into overall strategy level, cybersecurity industry level and case company level to not only research the topic on a theoretical level but also tying the research to practical level at a case company. Firstly, the top-down view of the overall impact of data ownership in a strategy process is taken and hereby the first research question is:

RQ1: What is the role of data ownership when formulating a market entry strategy?

The second research question is formed to tie the research into the cybersecurity context and increase the understanding of what data and what kind of data can create value if it was captured. Therefore, the second research question follows:

RQ2: What kind of data is valuable in a strategy process in the cybersecurity industry?

The aim of the third and last research question is to accumulate the case company's understanding around the topic and how it should align its strategy and resources in the future. The final research question is:

RQ3: How could the case company leverage data ownership as part of its strategy?

The research topic is analysed by conducting a literature review and a qualitative case study. Interviews, observations, and archival materials are used in the research data collection and a more detailed description of data collection and analysis is presented in chapter 3.

1.3 Scope and limitations

Market entry strategy, data ownership and cybersecurity narrow the theoretical boundaries of this study. In addition, presence of a case company steers the research to focus on viewing the topic firstly from a market challenger's point of view but also from business-to-business software provider's perspective. A possible differentiation between the role of data ownership in strategy processes for different types of vendors such as tech giants, other market leaders, business-to-consumer and studies for different industries is left for future research.

1.4 Structure

The structure of this thesis is as follows. This first chapter introduces and motivates the reader to the topic in addition to framing the research objective, the three research questions, scope of the study and the chosen methodology. The second chapter deepens the understanding of existing literature especially in the fields of market entry strategy, data ownership and cybersecurity. In the third chapter the used research methodology is described, and the results of the collected data is analysed in the fourth chapter by using Gioia et al. (2012) methodology. Finally, in the fifth chapter the implications of the thesis are discussed and concluded in addition to raising the limitations of the study and introducing potential topics for future research.

2. Literature Review

Companies are continuously adapting their strategies due to changes in their external operative business environment (Teece et al., 1997). Executives and managers make decisions on how to react to external environmental changes, whether to enlarge, develop or divest parts of their existing product portfolio to gain competitive advantage and how the organization and its processes should be structured to support the corporate strategy. Shatrevich and Gaile-Sarkane (2015) suggest these strategic changes to be bi-directional as the strategies are influenced by the external environment and structural dimensions but also the strategies themselves affect the external environment and structural dimensions (see Figure 2).

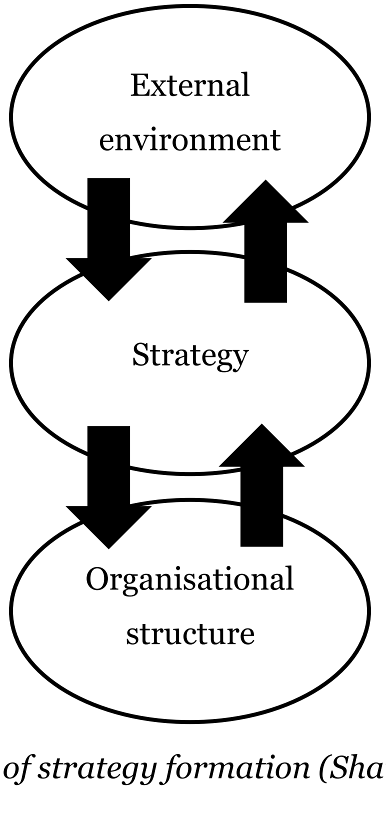


Figure 2: Bi-directionality of strategy formation (Shatrevich & Gaile-Sarkane, 2015)

There exists a large pool of frameworks and theories to be used when making these strategic decisions. However, some of the most used strategy theories introduced by Ansoff (1957), Barney (1991), Porter (1980; 1985), Williamson (1975) among others have been created in the last century when the industrial mix in the economy was different and the value chains were more straightforward compared to the present situation. Thus, those theories will be lacking some of the strategic aspects that businesses, especially in nascent industries such as IT, are currently dealing with.

The literature review is approached as follows. The first chapter takes a general view on market entry strategy formulation and process. In chapter 2.2 data ownership literature is introduced. And finally, in chapter 2.3 cybersecurity industry point of view is considered.

2.1 Market entry strategy formulation and process

Companies every now and then enter new markets to grow their business or react to uncertainties and changes in their competitive environment (Lilien & Yoon, 1990). Siggelkow (2002) argues that companies have two ways to enlarge their core: thickening their current cores or patching new cores. Decisions to thicken the existing or patch new cores have some strategic reason behind them.

2.1.1 Building strategy with competitive advantage

Barney (1991) argues that companies can gain sustainable competitive advantage by approaching their strategic analysis with a resource-based view model. This RBV model assumes that companies are heterogeneous and thus differ from each other. These differences are caused by the resources and capabilities of companies which in turn affect whether a company will gain competitive advantage or disadvantage in the industry.

Widely used VRIO framework was introduced in Barney's (1991) paper. The framework suggests that companies can gain sustainable competitive advantage if the resources of a firm are valuable, rare, costly to imitate and exploited by the organization. Peteraf (1993) introduced "the cornerstones of competitive advantage" (see Figure 3) which are partly based on Barney's (1991) RBV paper among others. Peteraf highlights the importance of factors that limit the competition prior and after the strategic choices to enter a market which enables companies to return excess profits by the competitive advantage. By heterogeneity and ex ante competition limitation companies are willing to secure the profitability and favourable cost structure of the market prior entering the field of business. Ex post competition and imperfect mobility are drivers for the strategy professionals to analyse the market environment and competitive positions after entering a market and how the profitability and customer retention will behave.

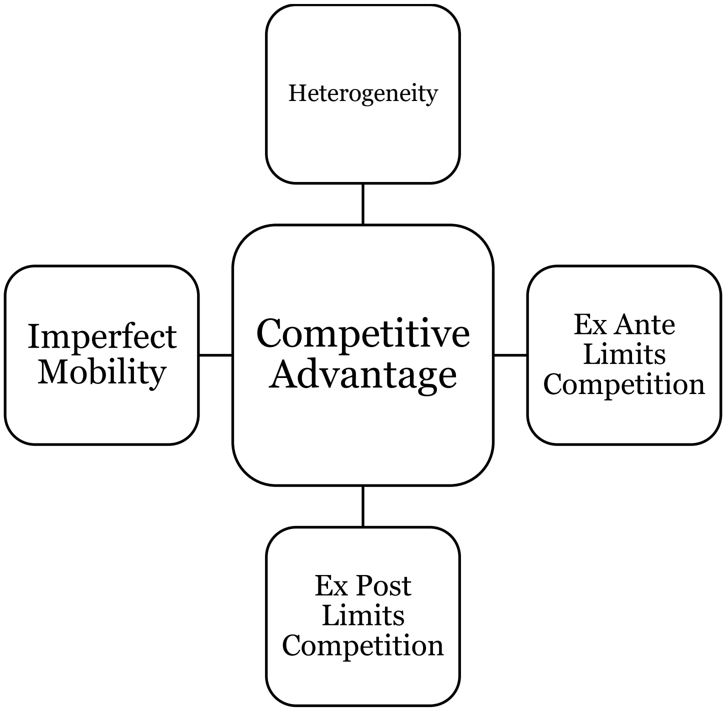


Figure 3: Peteraf's (2003) cornerstones of competitive advantage

However, Teece et al. (1997) have argued the original resource-based view to be static and they rather suggest a dynamic capabilities approach. This dynamic point of view reflects companies' ability to gain competitive advantage in novel and innovative ways. Teece et al. (1997) argue that competitive advantage can be gained not only with their current resources and capabilities but also build, reconfigure, and integrate new internal and external competences to support the business growth with the existing capabilities. Therefore, Teece et al. (1997) argue that companies should develop their capabilities and competences rather than just only products as products can be viewed as a moulded output of the firm's capabilities and competences.

2.1.2 Strategy formulation in a nascent market

Competitive environment typically affects the strategic choices made by the companies (Teece et al., 1997). Strategy formulation is a different process in a nascent market compared to processes in stable or declining industries. According to Ott et al. (2017), strategy formulation can be divided into "strategy-by-doing" and "strategy-by-thinking" approaches while keeping in mind that both approaches are needed in any strategy process.

Nascent markets are typically unpredictable, high-paced, and innovation-driven which requires companies to constantly iterate their routines and market needs (Ott et al., 2017). Therefore, top executives are unaware of all the possible outcomes in the future, and they should rather exploit a strategy-by-doing approach in the strategy process. However, strategy-by-doing is not just doing all the things that seem interesting. Characteristics of the process must be recognized which means that managers must test assumptions of the ideas and cut the losses early enough while learning from mistakes or reflect before scaling (Ott et al., 2017; McDonald & Eisenhardt, 2020).

Chen et al. (2021) state that product portfolio diversity is just one contributor to product success while iterative strategy-by-doing approach after the initial product launch enhances product performance positively. Also, Helfat and Peteraf (2003) argue towards an iterative approach when companies are accumulating capabilities. Helfat and Peteraf use a 6Rs model (Renew, Redeploy, Recombine, Replicate, Retreat or Retire) when companies are making decisions over what to do with their capabilities in an iterative process. The 6Rs model is introduced in Figure 4.

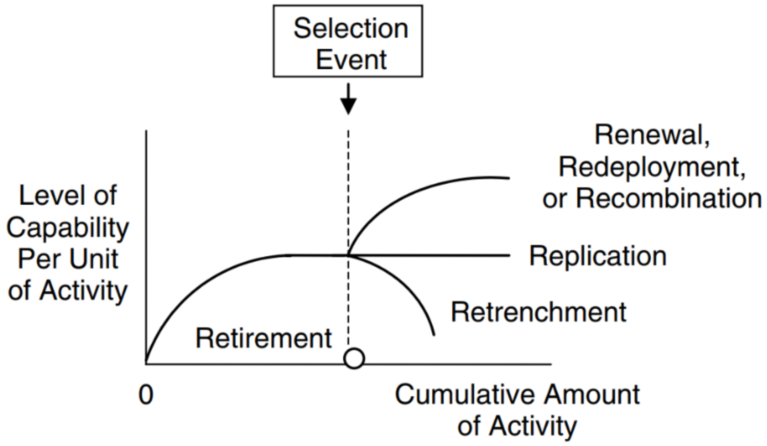

Selection Event
Level of Capability Per Unit of Activity
Renewal, Redeployment, or Recombination
Replication
Retrenchment
Retirement
0
Cumulative Amount of Activity

Figure 4: Helfat and Peteraf's (2003) 6Rs model

In addition to strategy-by-doing, top executives should enhance strategy-by-thinking when formulating strategy. In nascent markets some of the strategic decisions must be made under strict time constraints. This sets the decision-makers to form rapid analysis and decisions where experience from similar types of situations creates value to the strategy process. (Ott et al., 2017) Also, Brittain and Freeman (1980) argued that entering a new market is quicker for companies who possess overlapping capabilities with the market to enter with their existing markets. This enables having an option to develop the required product offering organically in a shorter time horizon compared to competitors. On the other hand, firms with more industry-specialised assets and capabilities have greater possibility to identify and enter a nascent market pocket in its industry (Mitchell, 1989).

The industry standards and operating models are unsettled in novel and nascent industries. Thus, new firms are found to be more flexible to test assumptions, change direction accordingly and reconfigure strategically while established companies have an option to enlarge their offering and capability branching through mergers and acquisitions. (Zuzul & Tripsas, 2020; Helfat & Peteraf, 2003)

2.1.3 Redesigning business model

Due to changes in the competitive environment and customer needs, businesses must renew themselves every now and then to avoid being a victim of creative destruction. Renewing means that companies operate completely in a new different market than before due to multiple rounds of patching new and existing cores (Siggelkow, 2002). However, renewing can also be done via changing the business model to the current customer base with current product offering. Nevertheless, changing the business model completely is challenging for organizations and especially if the differences between the new and old models are considerable (Lassila, 2005).

Amid and Zott (2001) segregate business model and revenue model such that business model refers to value creation whereas revenue model how the value is distributed between the players in the value chain. The definition of these separate models is needed especially when the SaaS concept is next discussed in order to not see the service concept just as a different billing option for a customer to pay for its software.

Software industry is one example of an industry that has been under radical change from on-premises software products to Software as a Service (SaaS) model (Guo & Ma, 2018). Mäkilä et al. (2010) define SaaS as "a software deployment model where the software is provisioned over the Internet as a service." In addition, a typical characteristic of the SaaS model is that the service is to some degree standard to all the customers whereas on-premises products can be modified and configured based on individual customer preferences. Customers benefit from using SaaS instead of on-premises software by easier access to technical expertise, frequent and free software updates, and access to software anytime and anywhere (Lassila, 2005). Sääksjärvi et al. (2005) summarized the benefits and risks associated with the SaaS model from a software vendor's point of view in Table 1.

Benefits	**Risks**
1. Economies of scale in production and distribution due to one-to-many offering	1. Management of complex network of suppliers if SaaS requires integrating third party products and software
2. Predictable cash flows due to recurring revenue	2. Short-term revenue reduction due to loss of license sales
3. Customer base expansion potential due to lower initial investments for users	3. Expected performance and scalability issues depending on the used technical solution
4. Lower sales cycles	4. High initial investment when ramping up the SaaS business due to building and maintaining the needed IT infrastructure and third-party software purchasing costs
5. Lower version management and maintenance costs	5. Lack of customization limits the potential customer base or customization occurs additional costs
6. Barriers of entry for competitors after successful integration	6. Customers expect, and they are promised frequent software updates

Table 1: The benefits and risks of the SaaS model for the vendor (adapted from Sääksjärvi et al., 2005)

Amit and Zott (2001) evaluate business models via four different factors: efficiency, complementary, lock-in and novelty. Efficiency is the most important value driver in the model by Amit and Zott (2001) which supports the usage of SaaS over on-premises solutions as the SaaS vendor can leverage economies of scale and scope in their operations.

Amit and Zott (2001) also found that SaaS offering was seen to be more valuable when a software vendor co-operates both horizontally and vertically with other providers or creates SaaS offering together with other firms. However, Amit and Zott (2001) highlight the importance of remaining in the attractive position in the value chain if additional players are introduced in the network while keeping their incentives in line with other vendors.

The third driver, lock-in, is high in SaaS model as customers are unwilling change their software provider due to laborious IT implementation projects and thus the switching cost to other SaaS or on-premises providers might not be the top priority to the customers especially if the software under consideration is a business-critical tool.

Lastly, novelty was seen to be one of the four drivers. In this case, all the SaaS solutions have already been on the market for a while taking around 25% market share of the total IT spending globally (Mlitz, 2021; Liu, 2021). However, SaaS is still seen as one of the key growth drivers in the information technology industry due to its high growth expectations (Mlitz, 2021). However, the figure varies in certain subsectors so SaaS might be a novel model to some sectors or business functions in the IT sector.

2.2 Data ownership

The amount of data created, captured, and consumed in the world is growing rapidly. Holst (2021) estimated that the amount of data over quadrupled from 15.5 zettabytes in 2015 to 64.2 zettabytes in 2020 and this growth is expected to continue such that in 2025 the amount of data created is 181 zettabytes meaning almost tripling since 2020. The growth is driven by the increase of data capturing devices and easiness to store data in the cloud among other drivers. In this chapter the role of data is discussed and how the data could be leveraged in strategic decision making in addition to how the ownership of data is organized in the cloud era.

Also, general data ownership discussion and worries over personal data usage have increased in the past few years. Especially the tech giants where the cornerstone of the whole business is tied with data analytics and the handlers of sensitive consumer data have raised concerns over who owns the data, who can access the data and how the data is being secured. Different regulators have reacted to the ownership and data sharing topic and for example the European Union created the General Data Protection Regulation in 2018 to increase individuals' control and rights over their own personal data and simplify the regulatory environment across the Union.

2.2.1 The role of data in strategic decision-making

Generally, it's easy to consider data ownership to be a good thing for businesses as they can utilize data to further develop their products and serve their customers better. However, as the IT applications are moving to the cloud and thus the data is stored outside companies' premises, questions arise over the ownership of the data and data's strategic value when making decisions. Obviously, data ownership questions will finally end up comparing benefits and risks both in financial and strategic perspective. For some companies it might be valuable to outsource data ownership as they receive cloud and other services by sharing their data whereas some of the companies are more fragile to share their data thus creating an additional source of vulnerability of data leakage.

Miller and Mork (2013) introduced a data value chain that can be seen in Figure 5. They argue that data-driven decision-making consists of three stages: data discovery, data integration and data exploitation. The value of data accumulates the further you go in the value chain and thus one can see raw data just as a resource to enable data-driven decision-making whereas the more valuable forms of data are more processed into decision-making enablers through analysis and visualisations.

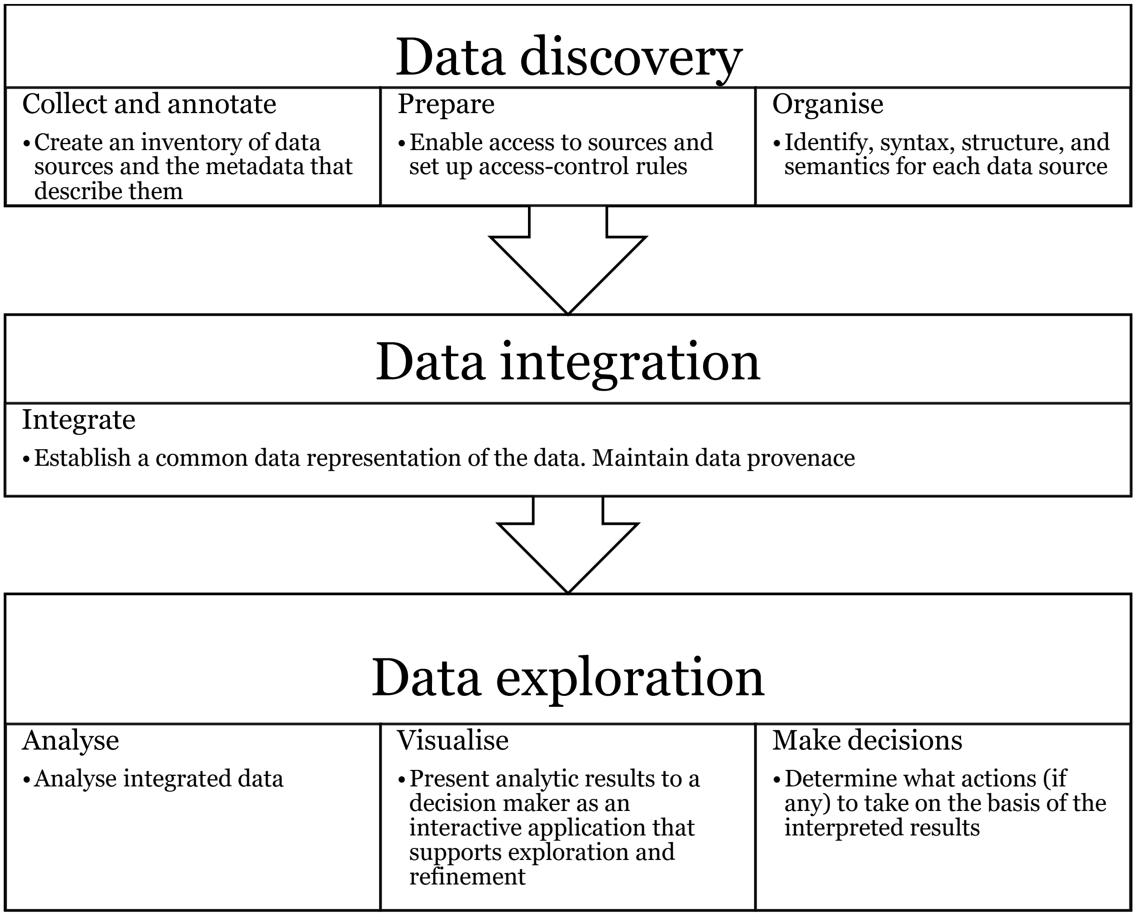

Data discovery
Collect and annotate
• Create an inventory of data sources and the metadata that describe them
Prepare
• Enable access to sources and set up access-control rules
Organise
• Identify, syntax, structure, and semantics for each data source
Data integration
Integrate
• Establish a common data representation of the data. Maintain data provenace
Data exploration
Analyse
• Analyse integrated data
Visualise
• Present analytic results to a decision maker as an interactive application that supports exploration and refinement
Make decisions
• Determine what actions (if any) to take on the basis of the interpreted results

Figure 5: Miller and Mork's (2013) data value chain

Data can enhance companies to gain competitive advantage if the data is rare and challenging for competitors to imitate (Akhtar et al., 2019; Harris & Graig, 2010). If the competitors have access to same or similar data, the product or service will be unable to be a source of competitive advantage and at most being a strength to the company based on Barney's (1991) VRIO-model.

The data itself can be defined to be somewhat close to a commodity such as oil, metal, or any other raw material. The real value of data can be captured only when it's analysed and value adding decisions are made from the analysis. So, in order for companies to be capable of making data-driven decisions, they require data analytics competence in their staff and foster a culture that supports data-driven decision-making (Shah et al., 2012). Also, Harris and Graig (2010) supports the idea of turning raw data into assets by providing a DELTA model where data, enterprise, leadership, talent, and analysts need to be in line with turning data into strategic assets. Akhtar et al. (2019) found that usage of big data-savvy teams' skills and big data-driven actions causes higher business performance that they define to consist of environmental, operational, and financial performance in addition to new business development. Beckwith (2020) also supports the idea of seeing data as one of the input components in data-driven decision-making while he highlights the importance of aligning organizational capabilities, processes, and resource allocation in the creation of business value in Figure 6.

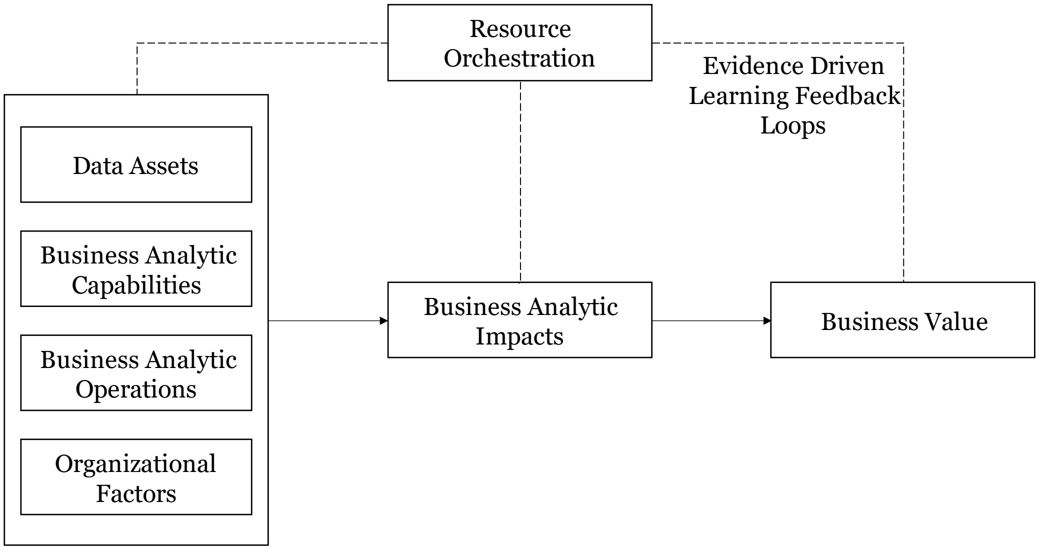

Resource Orchestration
Evidence Driven Learning Feedback Loops
Data Assets
Business Analytic Capabilities
Business Analytic Operations
Organizational Factors
Business Analytic Impacts
Business Value

Figure 6: Beckwith's (2020) business analytics value generation model

So as the amount of data in the organizations grows, it becomes more challenging for organizations to handle the data themselves whereas the attackers are also more interested in attacking the targets that host more data. According to Campos et al. (2016), big data challenges organizations from multiple interfaces that are described as 3V. The volume, velocity and variety of data has grown in the big data era to a level where organizations are unable to analysing their own data in full extent as non-big data-savvy teams will be unable to handle the amounts of data they have, the pace of the data coming in and the diversity of data coming in from various data sources. Thus, data gathering, and analysis can be expected to be done more and more by big data savvy organizations such as cloud providers.

Value through data-driven decisions can also be made without or with less human involvement by utilizing artificial intelligence ("AI"). Ayoub and Payne (2016) argue that AI can support humans in strategy work and risk evaluation especially within the areas that include human biases. However, Ayoub and Payne remind that AI's capability to create valuable and novel insights is dependent on unbiased data as the decisions are made from the given data that the AI has as input.

From a strategic point of view, Kitsios and Kamariotou (2021) argue AI to be a source of competitive advantage. They state that AI can enable companies to react in a shorter time to changes in the competitive environment and thus enhancing organizations' ability to exploit opportunities. However, in order to achieve the competitive advantage with AI and other forms of IT, companies need to align their business strategy with IT strategy if they even have one. Kitsios and Kamariotou (2021) found that this alignment led to higher firm performance, value and sustainability that can be seen from Figure 7.

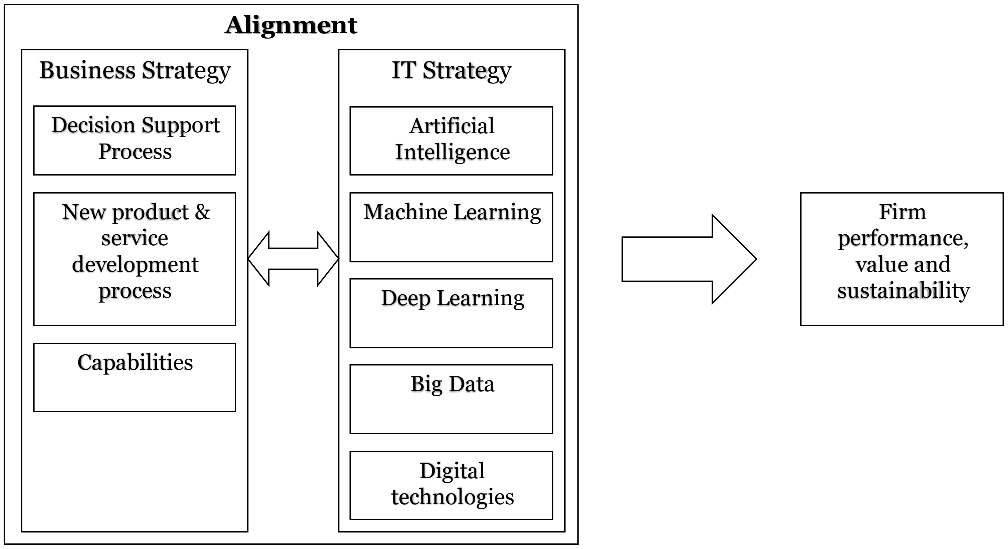


Figure 7: Kitsios and Kamariotou's (2021) IT strategy alignment framework

2.2.2 The value of owning the data

Data ownership is a complex issue. Organizations gather data to emphasize the benefits of the data in three different dimensions which are revenue optimization, cost control and risk mitigation (Fisher, 2009). Although these dimensions create value for firms as analysts and algorithms leverage the data when correctly implemented but at the same time collecting, storing, and using data cause costs and increases risk of being a victim of a cyberattack. Lynch and Hayes (2011) argue that the best way to avoid a data breach is not to store data in the first place. Nevertheless, organizations should see data as a strategic asset rather than a technology asset (Fisher, 2009; Otto, 2015).

Goasduff (2021) argues that data and analytics leaders who share their data to stakeholders will create three times more measurable economic benefit compared to ones who keep their data themselves. The reason for additional economic outcomes is stated to be increased stakeholder engagement as they can immediately enjoy insights and recommendations. However, in order to benefit from successful data sharing a couple of obstacles are required to be tackled. Firstly, the data, the data sources and the data-sharing counterparts must be trusted to achieve the advantages of data sharing. Goasduff (2021) predicts that less than 5% of data sharing programs successfully identify the trusted data and the trusted data sources. Secondly, Goasduff (2021) argues that traditionally organizations have had a mindset of “don’t share data unless” and it should be changed to “must share data unless” to benefit from data sharing. This fundamental change is challenging to achieve and cultural change from siloed data ownership culture to data sharing culture requires fostering by identifying emotional impacts and intrinsic biases that complicates forming the data sharing culture.

Adner et al. (2019) define data to be a scale-free asset that can be replicated from the original owner to other parties. Therefore, sharing intangible assets such data differ from sharing tangible assets where the sharing party abandons assets rather than copies it. This possibility makes data a highly fungible resource meaning that the value decreases mildly when it’s applied in its second-best use-case relative to first-best (Montgomery & Wernerfelt, 1988; Anand & Singh, 1997). The fungibility may even increase if an entity that owns the data sees low possibilities to monetarize the value out of the data. The different level of fungibility of data supports the idea of different organizations sharing data with each other in order to create “data network effects” due to larger sets of data to be used in the decision-making situations. Thus, Adner (2013; 2017) see that companies operating in data aggregation ecosystems form a value-creating position simultaneously being able to attract new entities to join the ecosystem while entering new ecosystems itself.

On the other hand, the value of owning data can be compared to owning other types of intellectual property assets and their characteristics of financial returns. For example, patents represent a widely studied area of IP rights and thus analysing the dynamics of the financial side of the patents will help us to understand the possibilities in data monetarization. Both data with its analysis as presented in chapter 2.2.1 and patents can form a barrier for other companies to enter or scale the operations in the underlying market thus gaining the holder of data or a patent a source of competitive advantage.

Lanjouw et al. (1998) studied patent data and renewing value of patents. They found that the financial returns of patents are typically lognormal as a function of time. This means that a holder benefits the most in the beginning but as the time goes by new innovations and substitutes cut the excess margins that can be gained with the patent.

However, as it’s the case for patents there are obviously also costs associated with for example storing, securing, and updating your data. As Miller and Mork (2013) suggested, the value of the data comes from the analysis which means that the data also needs working on it to proceed in the data value chain to analytics and thus decisions and finally being able to benefit from the possible financial outcomes.

2.2.3 Data ownership in a cloud and SaaS era

As mentioned in chapter 2.1.3, the usage of IT services is moving more and more towards cloud and SaaS models. One typical characteristic of cloud and SaaS is that the end user or the customer don't physically own the data themselves, but the data is physically stored in cloud vendors' data centres outside customers' properties. The ultimate ownership of the data in the cloud service varies as some of the service providers own the data and intellectual property rights to the data put into the service. In turn, some of the cloud providers leave the ownership to the user but might want to have access to some of the data for targeted service providing purposes (Birch et al., 2021). In the dictionaries 'ownership' is defined to be "right to exclusive use of an asset" or "full right to dispose of a thing at will" according to Black et al. (2017) or Kazhdan (1991), respectively.

This cloud model suits many industries and business functions but not for all. For example, some of the industries that handle and store fragile data, such as health and financial industries in addition to armies, have data security as one of their top priorities in their IT infrastructure and strategy. DalleMule and Davenport (2017) defines data fragility into offensive and defensive data based on downside risk related to data ending up in the wrong hands. DalleMule and Davenport (2017) suggest that every company has both offensive and defensive data, but the ratios vary typically based on the industry and its regulatory framework in which the company operates. Therefore, more defensive-driven industries are unwilling to let cloud providers offer their solutions to limit the cybersecurity related risks.

In general, cloud providers should benefit from owning their customers data if they offer to store it. Thus, cloud providers can leverage the insights that their data supports by using artificial intelligence and machine learning algorithms to further develop their products by thickening their products features (Siggelkow, 2002). Partly due to scale benefits, Giustiziero et al. (2022) found that digital era companies tend to be highly specialized in a narrow field of expertise while achieving large scale in that specific market. Giustiziero et al. (2022) argue scalable digital resources to be a large-scale value creator when a digital company “hyperspecifies“ and “hyperscales”.

Gregory et al. (2021) suggest that network effects are achieved when companies cooperate with their data thus supporting the idea of forming larger data generating pools in order to achieve enhanced value creation possibilities with multiple stakeholders within the market. Gregory et al. (2021) found that data network effects increase the user value with the AI capability of a platform. Numerous companies in the field of IT have adopted models to enhance data generation through freemium and open-source models. However, Boudreau et al. (2021) argue that the market challengers are unable to capitalize the freemium model into revenue generation despite more populated platforms. Whereas market leaders are sometimes capable of creating value with the freemium business model by penetrating most of the addressable market.

2.3 Cybersecurity industry

Soltys (2020) defines cybersecurity to be “a set of techniques and measures taken to protect digital information against unauthorized access or attack” while Bejtlich (2004, p. 4) defines security in general to be “the process of maintaining an acceptable level of perceived risk.” Moreover, Bejtlich’s security process, which is shown in Figure 8, consists of four steps: assessment, protection, detection, and response.

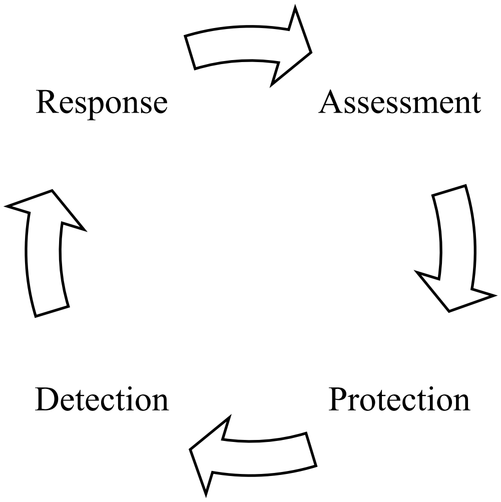


Figure 8: Bejtlich's (2004) security process

Based on Bejtlich’s (2004) definition, assessment is preparation for the other three components in the process. Protection is defined as “the application of countermeasures to reduce the likelihood of compromise” while detection is the process of identifying computer security incidents in cybersecurity cases. Finally, response is the part of the process where the possible detections are validated and required actions are taken. The actions can be divided into two categories. Actions can be either restoring the initial stage and moving on or collecting evidence of illegal actions and continuing the legal way.

Although cybersecurity is featured in the media often when a cyberattack, data breach or the like has occurred, there exists the other side of the coin also that must be considered. The role of cybersecurity is to allow the right people or machines to have access to the right time to the right system thus keeping the unwanted access from happening (Daniel Ani et al., 2017). However, managing of the accesses in the organization can be challenging as Cser et al. (2018) estimate that 80% of data breaches are caused by misuse of administrative privileges such as passwords, tokens, or certificates. The mentioned permission-based framework must be considered in the cybersecurity context as the most secure system cybersecurity-wise is a closed machine or IT system, but it would prevent the user from benefiting from the advantages of the technology.

Risk is closely related to security. Bejtlich (2004) defines risk as a measure of danger to assets that might relate in security context information, intellectual property, or reputation among others. Risk is typically measured as a product of probability and impact of unwanted events whereas Bejtlich (2004) measures risk in a security context as a product of threat, vulnerability, and asset value. In this model, threat is defined to be "a party's capabilities and intentions to exploit a vulnerability" whereas vulnerability is "a weakness in an asset that could lead to exploitation" and asset value is "a measurement of the time and resources needed to replace an asset or resource to its former state."

Given these risk definitions, it's good to note that in cybersecurity or in any other form of security risk cannot be fully deleted but rather minimized and mitigated. Fisher (2009) states that three types of risks exist: calculated, misguided and negative risk. Calculated risk is defined to be taken under fully being aware of the consequences of the possible outcomes which for example might mean in the cybersecurity context that an organization consciously doesn't fully secure its IT environment. Misguided risk refers to a suboptimal state where the risks are believed to be calculated but the assumptions behind the risk assessment are wanting or even incorrect. Misguided risk in cybersecurity could mean that a company imagines leaving some small and irrelevant parts of their IT unsecured but in the case of cyberattack it turns out to be a more critical part of their business operations than they thought in the first place. Finally, Fisher's (2009) third form of risk is the negative risk which is defined to be unethical employees and negligent business practices which in the cybersecurity world might mean for example a human error by accidentally sending confidential information to a wrong person. Therefore, in cybersecurity it's essential to secure the most valuable assets by minimizing vulnerabilities. Herrmann and Pridöhl (2020) summarizes the relationship between vulnerability and risk in Figure 9.

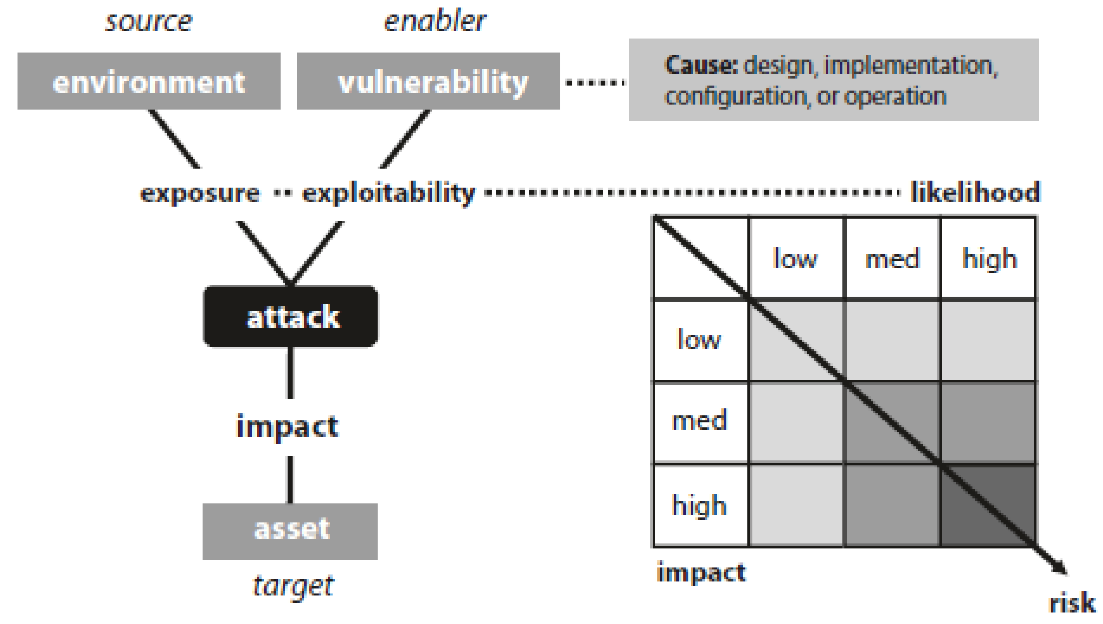


Figure 9: Herrmann and Pridöhl's (2020) presentation on the relationship between vulnerability and risk

Typical framework to determine the objectives of cybersecurity is CIA + AAA (Soltys, 2020; Herrmann & Pridöhl, 2020). This framework emphasizes Confidentiality, Integrity and Availability in the first place and these objectives are supplemented by the three A's: Authentication, Authorization and Accounting.

Confidentiality is one of the key objectives of cybersecurity to guarantee that information is used by the right people or machines. Typical way to ensure the appropriate access is using passwords, keys, and certificates (Xue et al., 2018). In cybersecurity, the cloud environment is seen as challenging from a confidentiality point of view as the one who stores files might want to keep the data confidential against the cloud provider while the cloud users are unable to ensure that their data will be completely prevented from ending up to malicious users. (Soltys, 2020; Herrmann & Pridöhl, 2020)

Another key objective of cybersecurity is integrity. This means that all the modifications in data can be detected to secure safe transmission, procession and storing of the data while avoiding any accidental or malicious changes in the data during its lifecycle. Soltys (2020) gives an example from online money transfer where it's crucial that the amount of money transferred remains unchanged during the whole process. (Soltys, 2020; Herrmann & Pridöhl, 2020)

The final key objective of cybersecurity is availability which means that the users have the information available at the time when the information is needed. This objective is becoming more and more relevant in the cloud era when the data is located outside the organizations' premises. Therefore, the computer systems and data transit channel should act correctly. (Soltys, 2020; Herrmann & Pridöhl, 2020)

The latter objectives, the three As, of the Soltys' (2020) model are authentication, authorization, and accounting. Authentication means in the cybersecurity context that a user that is attempting to access is in possession of credentials such as password, key, certificate, or some biometric authentication form like a fingerprint. Authorization signifies what a user is able and allowed to access once it has authenticated. Finally, the reason for accounting is to keep record logs which enables automation and close to real-time reactions. (Soltys, 2020; Herrmann & Pridöhl, 2020)

Goodwin et al. (2015) define cybersecurity information sharing to be consisting of information types, key actors, mechanisms of exchange and methods of exchange. Each of these pieces must be considered when sharing cybersecurity related data. Their framework can be seen from Figure 10.

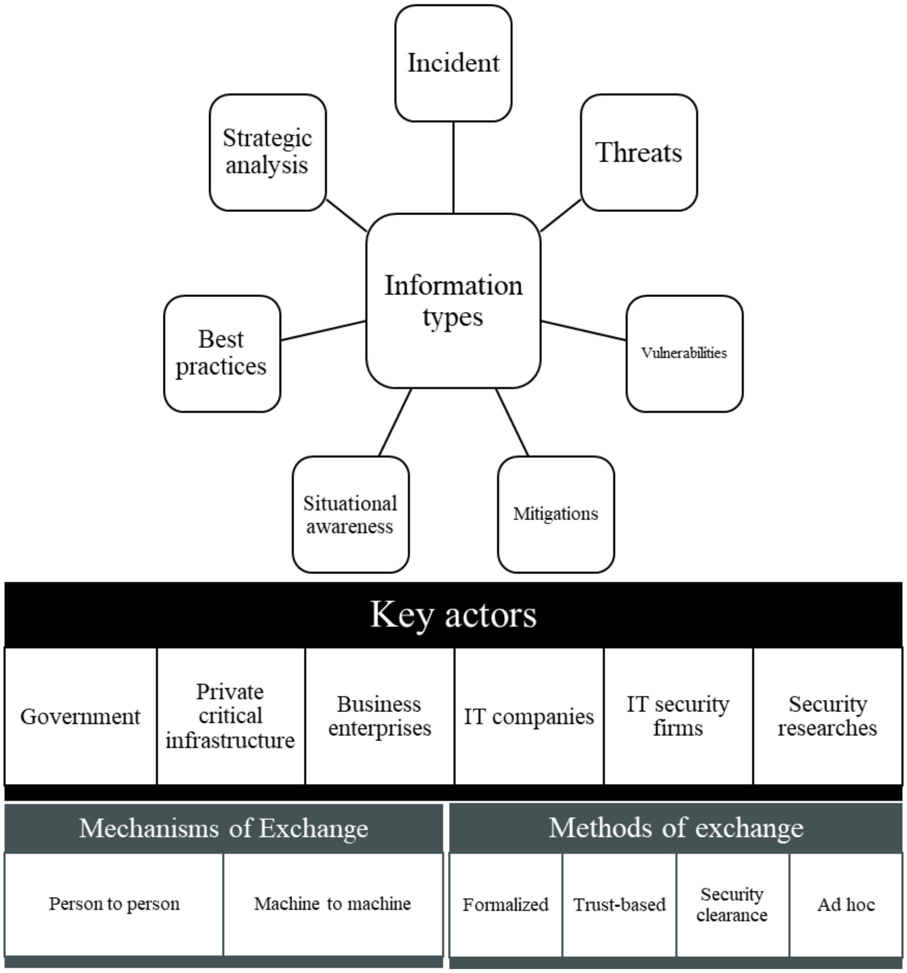


Figure 10: The foundation for cybersecurity risk management (Goodwin et al., 2015)

2.3.1 Cybersecurity in a cloud and SaaS era

As mentioned, the SaaS model has taken market share from the on-premises model as IT functions have moved to cloud in the past decade. However, the cybersecurity challenges change as the data have moved from hard drives to data centres as the data needs to be stored and transmitted securely (Campos et al., 2016). Sherman (2011) states that the cloud environment has increased demand for authentication as earlier it was enough to protect the corporate local area network (“LAN”) while currently using similar data and applications require a secure way to store and transit the data.

Cloud applications, remote access usage and growing number of employees’ personal connected devices in workplaces increase the network traffic from the organization’s premises. This growth in connectivity naturally raises security related risks but sophisticated IT architectural solutions exist such as zero trust. According to Rose et al. (2020), zero trust approaches the cybersecurity from authenticating and authorizing the users and subject that they are applying to use rather than statically exploring user’s network-based perimeters. In the other words, physical or network location or ownership of some device doesn’t guarantee a granted trust under zero trust environments.

Cybercriminals have also noticed the importance of cloud infrastructure. Atapour-Abarghouei et al. (2020) find that the recent cyber crimes are targeted more and more to cloud services as the clouds contain a lot of data from multiple customers in addition to having a strategically significant role in how companies operate their businesses and thus the attackers can demand larger ransoms from broader audience than attacking single organizations or consumers. Therefore, Campos et al. (2016) highlight the importance of confidentiality, integrity and availability when operating in cloud environments.

There exist different cloud infrastructures for different use cases of cloud (Gupta et al., 2013). Public cloud is the most known cloud infra as public cloud players like Amazon, Google and Microsoft serve both consumers and organizations. Goyal (2014) states that public clouds are typically the most vulnerable type of clouds and therefore attackers are interested in those in addition to the fact that these clouds contain a lot of data and have broad user bases which further increases the interest to abuses.

Another major deployment model is private cloud. Mell and Grance (2011) defines private cloud to be a cloud infrastructure that is created exclusively for a single organization that may comprise multiple customers or business units. An organization might want a private cloud if they operate multiple stores and want every store to have the same customer data simultaneously while having an increased cybersecurity features than in a public cloud. Private clouds can be owned and operated by the customer, a third party or a combination of them while the cloud may exist on or off premises. (Gupta et al., 2013)

The remaining deployment models are community and hybrid clouds where the former is a cloud for a specific group of consumers of organizations with similar concerns such as mission or security requirement whereas the latter is simply a combination of public and private clouds. (Gupta et al., 2013)

Security and privacy are among the three most concerning factors when adopting SaaS models generally, not just cybersecurity specific SaaS (Kumar et al., 2015; Gupta et al., 2013). Therefore, concerns and reluctance can be assumed to be even stricter when it comes to cybersecurity and other business critical SaaS. Safari et al. (2015) argue that currently at least SMEs are satisfied with their SaaS products although McLilly and Qu (2020) found that 65% of SMEs recognize being under a cyberattack while just 4% of SMEs have adopted the national best practices to defend from the attacks. Reasons for this phenomenon might be that SMEs handle less critical data than large corporations and thus the data security aspects are less relevant for them or as McLilly and Qu (2020) argue that SMEs simply don't have resources, familiarity, or proficiency to implement cybersecurity projects.

Trust is an important factor in the security industry. In order to build trust with customers, Safari et al. (2015) and Atapour-Abarghouei et al. (2020) argue that SaaS vendors should sign a service level agreement or a formal contract when it comes to security. However, Xin and Levina (2008) remind that it's challenging to form a contract that covers all the future aspects.

Cybersecurity customers are eager to mitigate the risk of cyberattacks and thus diversifying their cybersecurity supplier base (Naicker & Mafaiti, 2019). Multisourcing enables the customer of the contract to mitigate the supplier related risks, benefit from vendor competition and collaboration which leads to the possibility to acquire a broader set of skills and technologies with cost savings. This cooperative dimension for cybersecurity vendors will be underlined in the SaaS era as there typically exists a system integrator that integrates multiple different technologies and services under its offering. This arrangement benefits the integrator, the technology and service providers of the system and the customers. The technology and service providers get distribution channels to their offering that they wouldn't be able to form without the network. On the other side of the table, the customers benefit from this scheme by being able to operate via one provider as for many organizations critical cybersecurity procurements are bureaucratic processes with the requirement of many signings from managers and executives (Ruohonen et al., 2016). By forming this kind of one stop shop for the end customers, the counterparts of the network will benefit from the possible network effects and scale benefits that a larger network is able to produce.

2.3.2 Data sharing and data in the cybersecurity industry

Data in cybersecurity refers to for example IDS logs, firewall logs, network traffic data, packet data, and honeypot data etc. (Sarker et al., 2020). Other data that customers of cybersecurity vendors produce and use is unnecessary for cybersecurity vendors from a product development point of view. Bhatia et al. (2016) studied users' willingness to share their cybersecurity data with the vendor and how usable the data is for the vendors to develop their products. They found that the most beneficial data to be shared for both counterparts is for example operating system type and version, IP address and device information. On the other extreme, users' personal information such as emails, image files, passwords and chat histories were the least value creating pieces of information to be shared.

Cybersecurity data is typically sensitive and thus sharing this data is something that customers are unwilling to do. As data sharing is lacking (Atapour-Abarghouei et al., 2020) or data is only shared across organizations after a cyberattack has occurred (Goodwin et al., 2015), cybercriminals can gain advantage as organizations don't have a proper way to learn from each other to protect themselves from the attacks. Therefore, there could be a rationale for an existence of SaaS and cloud vendors that collect the data centrally thus being able to analyse the exceptional behaviour patterns and prevent cyberattacks. Hosting the data enables also applying new technologies, such as AI and ML, to detect the exceptionalities. Atapour-Abarghouei et al. (2020) suggest that cybersecurity vendors should leverage these new technologies in their offering as the cybercriminals are using them also. Sarker et al. (2020) describe the data value chain in the cybersecurity context when discovering the possibilities for the usage of ML in the value chain in Figure 11.

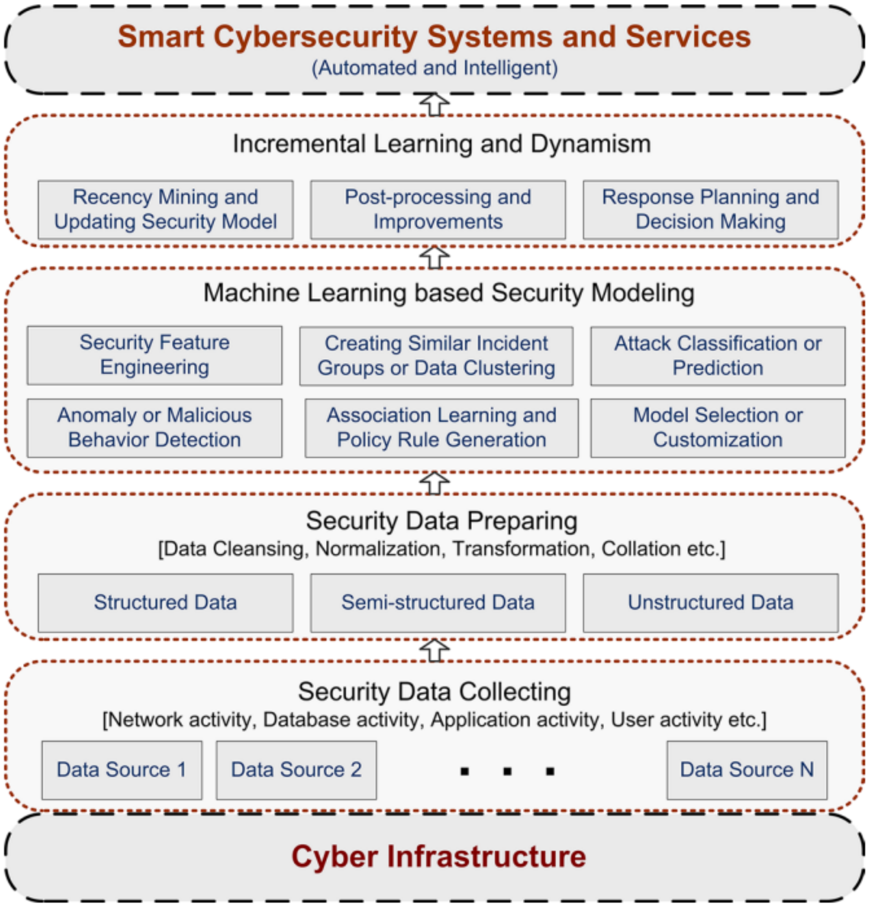

Smart Cybersecurity Systems and Services
(Automated and Intelligent)
Incremental Learning and Dynamism
Recency Mining and Updating Security Model
Post-processing and Improvements
Response Planning and Decision Making
Machine Learning based Security Modeling
Security Feature Engineering
Creating Similar Incident Groups or Data Clustering
Attack Classification or Prediction
Anomaly or Malicious Behavior Detection
Association Learning and Policy Rule Generation
Model Selection or Customization
Security Data Preparing
[Data Cleansing, Normalization, Transformation, Collation etc.]
Structured Data
Semi-structured Data
Unstructured Data
Security Data Collecting
[Network activity, Database activity, Application activity, User activity etc.]
Data Source 1
Data Source 2
. . .
Data Source N
Cyber Infrastructure

Figure 11: Data value chain in the cybersecurity industry in ML era (Sarker et al., 2020)

One fundamental challenge of cybersecurity data sharing is that the customers should be continuously sharing the data, but the direct benefits for the customer remain unclear or the upsides from the sharing extend far into the future (Goodwin et al., 2015). Goodwin et al. (2015) argue that company-to-company data sharing is the richest and most valuable form of data sharing as the parties benefit right away from the collaboration. Goodwin et al. (2015) also found company-to-vendor data-sharing valuable if it’s voluntary, but they state that mandatory data-sharing doesn’t improve operational security. This is especially the case if the other party of the sharing is authority as in that scenario the data sharing is typically one-directional.

It’s not only the data sharing as a privacy that is challenging but also sharing includes an additional layer that requires cybersecurity. Also, the cybersecurity vendors should evaluate if they can create enough value from the data that it surpasses the risk of carrying more data and being a more attractive target to criminals. Summing the different factors in data sharing, Having the technical, legal, and psychological barriers of sharing data, Koepke (2017) argues that the lack of incentives hinders the data-sharing in the cybersecurity field.

Forming a functioning data-sharing ecosystem requires certain frames to be conducted in order to make all the participants enrich the ecosystem. As Ring (2014) mentions, data-sharing is a challenging task just within a large organization. Therefore, detailed data definitions (Beynon-Davies & Wang, 2019), trust over the quality of the shared data (Cilluffo, 2017), and positive feedback on sharing is required in order to keep the data-sharing running (Brilingaitė et al., 2022). Thus, Atkins and Lawson (2021) found that cybersecurity data is shared more likely in two scenarios. Firstly, if the companies are small and thus, they don’t have visibility and resources to study all the vulnerabilities. Or secondly, the companies operate within the same industry that share a congruent and impactful threat that would endanger the existence of a firm or even the whole industry if a key supplier to the industry is attacked.

2.3.3 Moving cybersecurity functions into the cloud

According to Lassila (2005), changing business models is a challenging task for any company to execute. Although, cybersecurity industry might remain on-premises operating model, there still can be changes in the business models with the on-premises model or some of the cybersecurity applications will move into cloud and thus SaaS model. The most fragile customers may stay in the on-premises model for long while SMEs and some other customer groups have adapted cybersecurity SaaS solutions with a good satisfaction. Therefore, cybersecurity vendors are facing a strategic choice whether to enter the SaaS market, remain at on-premises model, adopt a hybrid of these two or something else. As it's often the case, the customer demand determines vendor's choices. Campos et al. (2016) support the importance of different users letting a possibility to choose the type of security they want.

Liao and Chen (2014) claim that the survival probability of an online business consists of three stages. In the first stage the market entrants invest heavily in network security ("NS") as security leadership is seen as a first mover advantage and as a necessity for an online business to operate in a nascent market. In the second phase the market has matured from the cybersecurity point of view as companies' willingness to invest in network security decrease.

In the second phase the market faces free riders that are unwilling to invest in network security as the users take security for granted or lack incentives to invest in cybersecurity due to aiming at increased cash flows or lack of cyberattacks. (Liao & Chen, 2014)

In the final and third phase, the cybersecurity technology has also developed to a level that protects its users to survive from greater cyberthreats and thus the remaining companies in the market are again eager to invest in network security (Liao & Chen, 2014). This model is visualized in Figure 12.

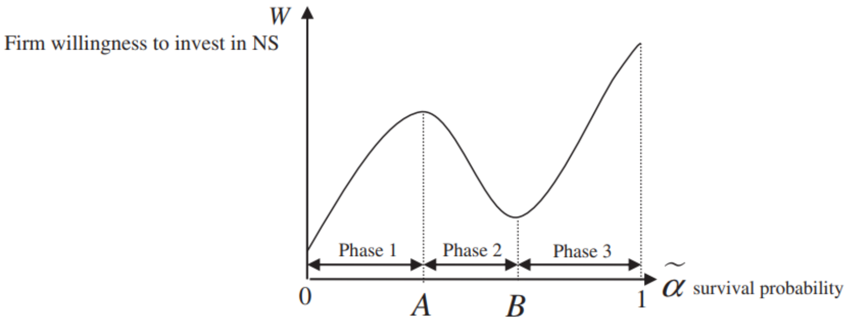


Figure 12: Firm willingness to invest in NS and survival probability according to Liao and Chen (2014)

McDonald and Eisenhardt (2020) argues that companies should aim at differentiating from their substitutes rather than their peers in rapidly growing markets to support the growth of the nascent market over the incumbent one. Therefore, players in the nascent market should be prepared for the strategic response of incumbent players. Guo and Ma (2018) state that SaaS entrants are typically lagging behind in product quality compared to incumbent perpetual software vendors, but SaaS vendors typically improve their product quality faster and on-go and thus they pose a threat to incumbent software vendors.

According to Guo and Ma (2018) incumbent software vendors react to SaaS vendors market entry and alter the competitive environment by decreasing prices when the SaaS's product quality becomes competitive. With this strategy incumbent vendors respond to SaaS model's lower upfront investments for the customers. Typically, this strategy leads incumbents to focus on market segmentation where they seek for segments where they could avoid competition and thus charge higher prices.

Finally, SaaS's quality bypasses the on-premises software quality through the continuous product development and thus on-premises vendors are left to compete mainly with lower pricing. Guo and Ma (2018) argue that incumbent software vendors should avoid SaaS's market entrance by offering periodic software quality improvements by free software incremental quality improvements and releasing new versions of the software with major quality jumps.

The cloud environment enables cybersecurity software vendors to have an additional way of delivering and operating their products. In the other words, the vendors can decide to offer only on-premises, only cloud or having both options in their offering. Campos et al. (2016) remind that it’s important for the users and customers to have an option to choose from based on their preferences.

2.4 Synthesis of the literature review

Strategy processes and data-driven decision-making are broadly studied in the literature. Both fields are seeing an increasing number of papers on the effects that AI and ML among other new technologies can enhance data analysis (Gregory et al., 2021; Kitsios & Kamariotou, 2021; Sarker, et al., 2020). However, the existing literature on strategy processes lacks the understanding of the data ownership aspect while the data ownership literature has focused generally on data-driven decision-making rather than specialising on some specific verticals (Miller & Mork, 2013).

Strategy formulation is an iterative process especially in nascent markets (Ott et al., 2017) while the strategy itself should be based on organization’s current and upcoming capabilities and resources (Teece et al., 1997). When an organization forms a strategy iteratively in a nascent market, it gathers data continually enabling rapid reactions to change in a market environment which are characteristic in nascent markets. However, having this raw data in place is not valuable itself (Beckwith, 2020; Miller & Mork, 2013) but rather the value is created when the data is analysed, and the analysis leads to enhanced decision-making.

Although raw data isn’t valuable itself as Beckwith and Miller & Mork (2020; 2013) describe but it can create competitive advantage for its owners if the data is rare and challenging for competitors to imitate thus enabling the data owning entity to utilize the data in its strategy process in a way than its rivals are uncapable of (Akhtar et al., 2019; Harris & Graig, 2010). Owning rare data creates value for its holder if the decision-making process is well-managed but the benefits don’t limit to the primal data-owning entity itself. Data-sharing enables the data-sharer to let other stakeholders benefit from having access to data, possibly creating network effects to the sharing organization as well (Adner et al., 2019; Anand & Singh, 1997; Gregory et al., 2021; Montgomery & Wernerfelt, 1988).

However, from a cybersecurity point of view data-sharing creates a challenge to allow access to an organization’s data as there always exists a risk of data or access ending up in the wrong hands. However, the literature has divided data sources into more defensive and more offensive ones depending on the level of security required (DalleMule & Davenport, 2017). Thus, sharing defensive data might limit the impact of risk associated with sharing although the likelihood could increase (Herrmann & Pridöhl, 2020). Nevertheless, owning and sharing data within the cybersecurity domain increases the security of the products by allowing the product development from having wider visibility over the vulnerabilities and threats associated in the cybersecurity space (Atkins & Lawson, 2021).

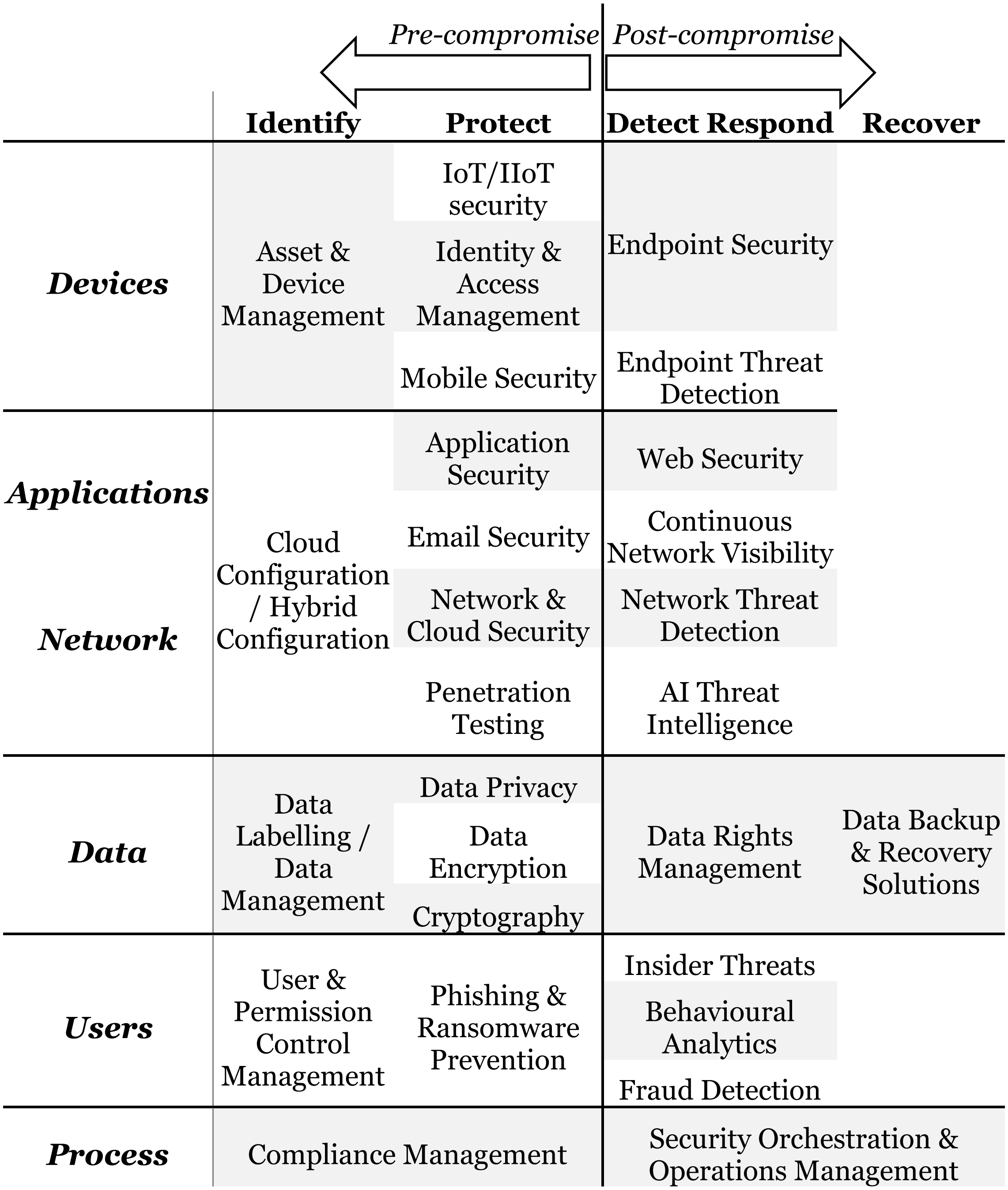
3. Research Methodology

This thesis is done for a Finnish cybersecurity vendor to investigate the research questions between September 2021 and February 2022. This section describes the research process. Firstly, the research setting is provided and the background of the situation of the case company and its industry is characterized. Secondly, the data collection process is discussed and after that how the data was analysed. Finally, validity of the gathered and analysed data is described.

3.1 Research setting

This thesis is conducted for a Finnish cybersecurity company that was founded in the mid-90s. The company grew rapidly in its early years supported by its highly competitive technological innovations in the secure communications field. However, the financial performance in the last couple of years as the revenue has remained in the same range since 2015 while the company has been unable to pay dividends for its owners driven by negative earnings. As a result of lack of growth and unprofitability, the ownership and thus the management and the strategy of the company has changed to a more growth-orientated one.

Roughly simplifying, a company can grow organically by price and volume increases with current customer portfolio, serving new customers within its existing markets or entering a new market. In this thesis the latest of these options is under the research as market entries should interest a lot of companies in the cybersecurity industry. Yu (2016) divided the cybersecurity market into dozens of segments and sub-segments, presented in Figure 13. Currently, the cybersecurity market is fragmented, and it consists of more than 1200 vendors globally (Miller, 2018). Typically, companies are focused on a few segments but to enlarge their offering they might be actively looking for new markets to enter.



	Identify	Protect	Detect Respond	Recover
Devices	Asset & Device Management	IoT/IIoT security	Endpoint Security	
		Identity & Access Management		
		Mobile Security	Endpoint Threat Detection	
Applications	Cloud Configuration / Hybrid Configuration	Application Security	Web Security	
		Email Security	Continuous Network Visibility	
Network		Network & Cloud Security	Network Threat Detection	
		Penetration Testing	AI Threat Intelligence	
Data	Data Labelling / Data Management	Data Privacy	Data Rights Management	Data Backup & Recovery Solutions
		Data Encryption		
		Cryptography		
Users	User & Permission Control Management	Phishing & Ransomware Prevention	Insider Threats	
			Behavioural Analytics	
			Fraud Detection	
Process	Compliance Management		Security Orchestration & Operations Management	

Figure 13: Yu's (2016) cybersecurity vendor landscape

3.2 Research design

Inductive studies develop theories and concepts based on data, concentrates and observations that are attached to the studied phenomenon and its meanings and thus inductive studies are often called as a grounded theory approach (Saunders et al., 2009, p. 118; Dubois & Gadde, 2002). One of the characteristics of inductive studies is its flexibility which allows agile changes during the process as qualitative data is used instead of quantitative data (Saunders et al., 2009, p. 127). Qualitative data, in this study especially from interviews, is compared with the literature review gathered. Therefore, this research can be defined to be an inductive study.

This thesis is approached by the "Gioia methodology" and "Building theories from case study research" introduced by Gioia et al. (2012) and Eisenhardt (1989), respectively. The Gioia methodology aims at developing concepts by systematic inductive approach (Gioia et al., 2012) while the case study approach introduced by Eisenhardt (1989) is widely used when novel theories are built iteratively and based on empirical evidence. The chosen approach enables creating generalizable theories.

In case study research "how" and "why" questions can be studied according to Yin (2017) as the answers to these questions describe the root causes for the phenomena. Yin (2017) defines a case study to be an empirical method where the underlying contemporary phenomenon (the 'case') is studied thoroughly and in a real-world context in addition to dealing with technically distinctive environments with a lot of more variables of interest compared to data points. Therefore, a case study approach has been selected to examine the research questions besides having a deep understanding regarding the topic.

3.3 Data collection

Data collection began with conducting a literature review. The literature review introduced the author to the cybersecurity industry while also forming the theoretical foundation for the study. Eisenhardt (1989) states that literature review is an essential part of theory building as comparison of the gathered data to existing literature improves the validity of the results in addition to enabling the research to approach the topic potentially from additional points of view and thus diversifying the theoretical framework of the study. The materials utilized in the literature review included the most relevant strategy, data and cybersecurity articles and e-books from diverse sources. The identification of the most relevant sources was done by searching with pertinent keywords in search engines. In addition, the reference lists at the end of each read article was leveraged to further deepen the theoretical knowledge of each topic.

The empirical data of the research includes interviews, observations from day-to-day strategy work in a case company and archival data. These data collecting methodologies are suggested by Marshall and Rossman (1989). The interview data was collected with a semi-structured method, where the interviewer asked questions around strategy, data, and cybersecurity themes. The preliminary list of questions was sent to interviewees typically a couple of days ahead or at least 24 hours before the interview in order for them to familiarize and prepare themselves more deeply into the topic. As the interviews were semi-structured and the question list were sent beforehand, the interviewees had a possibility to answer and describe the themes flexibly and agilely during the interviews which lasted typically around 60 minutes.

In accordance with remote work recommendations, the interviews were held with Microsoft Teams. As the remote work recommendation has taken place for over 18 months, the interviewees were used to meeting over the video connection and thus acting normally during the video calls. Utilization of meeting recording enabled the later analysis of the interviews and the responses while keeping the interviewer's attention in discussion and asking further follow-up questions in semi-structured interviews rather than focusing on taking notes.

3.3.1 Sample selection and interviews

The interviewees represented both case company's internal employees and external professionals from diverse backgrounds to gain a broad understanding from multiple points of view of the topic. The cybersecurity related work experience distribution was diverse as the most novel had been in the industry for a couple of years while the most experienced interviewees had worked over three decades among cybersecurity. In addition, the case company acquired six months before the beginning of the thesis another cybersecurity vendor that utilizes a SaaS model. Interviewing employees and customers from the acquired company and from the existing business of the case company broadens the understanding of the topic. The interviewees represented different backgrounds which can be divided into four different groups based on their background in Table 2.

Interviewee group	**Description of the group**
Technical focus	Internal, non-customer-interface employees with responsibility of technical development of the existing and new products
Commercial focus	Internal, customer-interface employees with responsibilities of selling product,

	account management and strategic development of the product portfolio
Customer or partner	Customers or partners of the case company
Other industry expert	External cybersecurity professionals with experience from broader set of cybersecurity market segments that the case company is represent

Table 2: Interviewee backgrounds

The number of interviews collected for the research summed up to 16 and the detailed summarization is presented in Table 3. The more detailed information, for example the exact job titles, will remain undisclosed to guarantee anonymized responses from the interviews. As the case company's origins are in Finland and thus a big part of the employees is Finnish, most of the interviews were held in Finnish which might be some source of bias. Since most of the interviews were held in Finnish, the quotes in chapter 4 are translated into English. Although the origins of the interviewees are mostly Finnish and thus forming a possible bias in the data, the characteristics of the researched topic are global, and the operations of the case company are worldwide as the products are sold to Americas, Asia, and Europe while the interviewees' focus is on the products and global operations rather than in some specific geographical area.

Informant no.	Role	Date	Length (min)
1	Internal 1	November 2021	55
2	Internal 2	November 2021	57

3	Internal 3	November 2021	59
4	Internal 4	November 2021	56
5	Internal 5	November 2021	56
6	Internal 6	November 2021	53
7	Professor	December 2021	70
8	Advisor 1	January 2022	46
9	Customer 1	January 2022	31
10	Customer 2	January 2022	50
11	Partner 1	January 2022	56
12	Partner 2	January 2022	50
13	Advisor 2	January 2022	57
14	Customer 3	January 2022	30
15	Advisor 3	January 2022	52
16	Customer 4	January 2022	28

Table 3: List of interviews

The interview meetings began by noting to the interviewee that the interview and all the responses are anonymized and a permission to record the virtual meeting was asked for later analysis of the interview. The interviews themselves began with personal introductions as the interviewer and interviewees were mainly unfamiliar with each other before the research process. After the introductions and small-talks, a more content-focused part of the interview officially took place. The preliminary list of interview questions was divided into three main sections: strategy, data ownership and cybersecurity. Two versions of a preliminary set of questions were used in the interviews based on whether the background of the interviewee was the case company's internal or external. The internal version was a bit more extended as a little more emphasis was put on the internal strategy and product development processes that external participants have limited access to. The preliminary lists of interview questions can be seen in the Appendices 1 and 2 for internal and external interviews, respectively.

3.4 Data analysis

In the first phase of the research process, the underlying dynamics of the cybersecurity market was orientated as the researcher was moderately unfamiliar with the cybersecurity industry. The case company has a product portfolio consisting of four main products in addition to the recent acquisition. Therefore, the background and strategic reasoning for the market entries to present markets of the case company was discovered in parallel to industry orientation.

After the onboarding and orientation, the more detailed research of the topic was kicked off. According to Eisenhardt (1989), analysing data is "the most difficult and the least codified" section of the inductive case study process although analysing data is in the centre of theory formation. However, Gioia et al. (2012) offer a useful methodology in creating a generalizable emergent theory. The key principles of the Gioia methodology are divided into three stages: 1st -order codes, 2nd -order themes and finally aggregate dimensions.

Recording the virtual interview meetings enabled the interviewer to both return to interviews and eased the transcribing process to have access to the interview data also in written format. The written interview data was once again read, and the first order codes were being identified from the responses. Gioia et al. (2012) present first order codes to be "informant-centric terms and codes" and due to informant centrality rather than attempting to distil categories. The amount of 1st order codes flooded, and thus alike codes were merged leading to the number of codes that were lighter to handle. At this point the used codes are informant centric as Gioia et al. (2012) propose.

Having the 1st -order codes in place, the codes were further analysed under more abstract 2nd-order themes that are generated researcher-centric. This breakdown enables rationalizing the interview data at a more theoretical level (Gioia et al., 2012). Finally, aggregated dimensions were generated from these themes leading to data structure that can be seen from Figure 14.

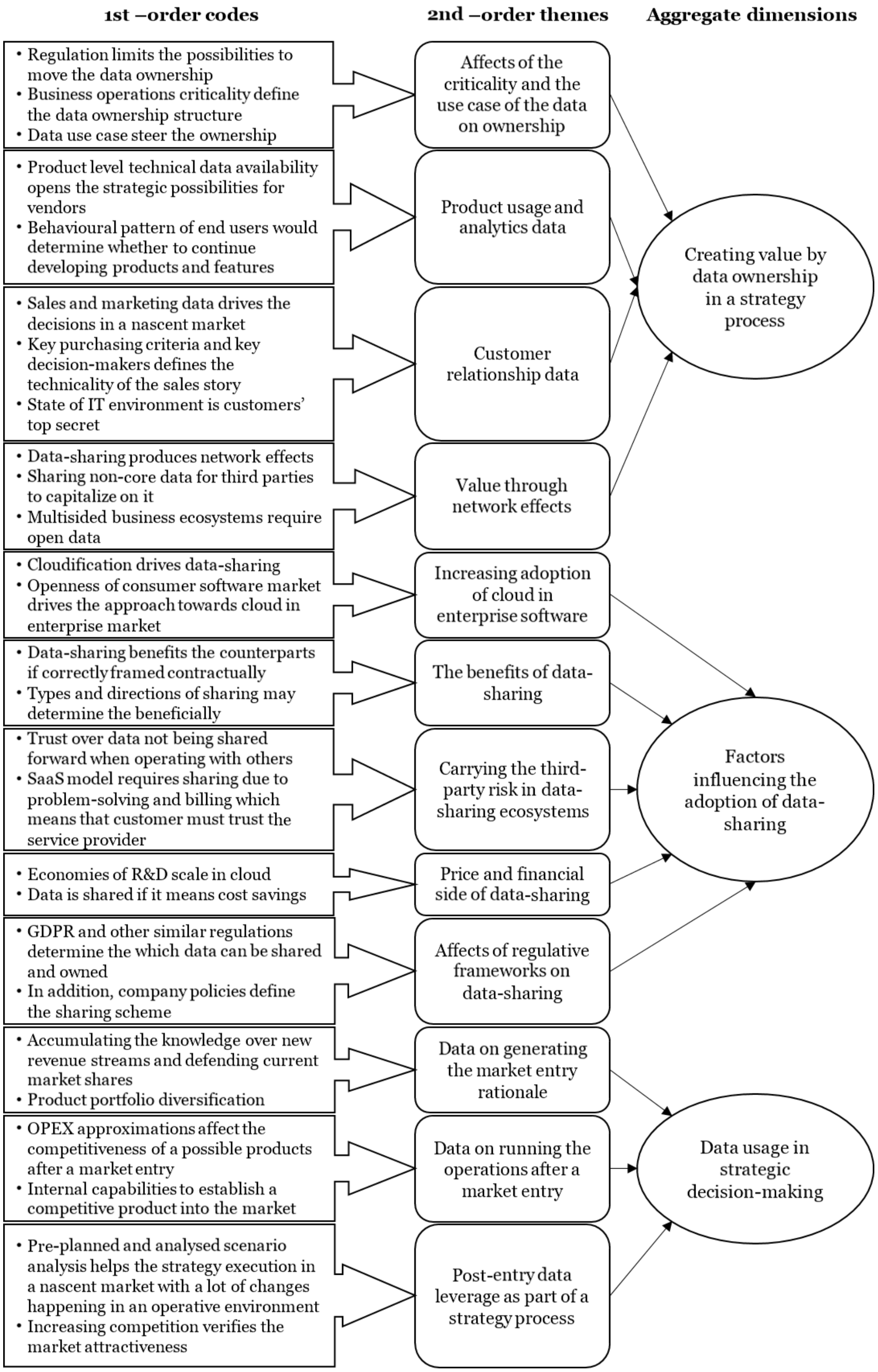

1st –order codes
2nd –order themes
Aggregate dimensions
• Regulation limits the possibilities to move the data ownership
• Business operations criticality define the data ownership structure
• Data use case steer the ownership
Affects of the criticality and the use case of the data on ownership
• Product level technical data availability opens the strategic possibilities for vendors
• Behavioural pattern of end users would determine whether to continue developing products and features
Product usage and analytics data
• Sales and marketing data drives the decisions in a nascent market
• Key purchasing criteria and key decision-makers defines the technicality of the sales story
• State of IT environment is customers' top secret
Customer relationship data
• Data-sharing produces network effects
• Sharing non-core data for third parties to capitalize on it
• Multisided business ecosystems require open data
Value through network effects
Creating value by data ownership in a strategy process
• Cloudification drives data-sharing
• Openness of consumer software market drives the approach towards cloud in enterprise market
Increasing adoption of cloud in enterprise software
• Data-sharing benefits the counterparts if correctly framed contractually
• Types and directions of sharing may determine the beneficially
The benefits of data-sharing
• Trust over data not being shared forward when operating with others
• SaaS model requires sharing due to problem-solving and billing which means that customer must trust the service provider
Carrying the third-party risk in data-sharing ecosystems
• Economies of R&D scale in cloud
• Data is shared if it means cost savings
Price and financial side of data-sharing
• GDPR and other similar regulations determine the which data can be shared and owned
• In addition, company policies define the sharing scheme
Affects of regulative frameworks on data-sharing
Factors influencing the adoption of data-sharing
• Accumulating the knowledge over new revenue streams and defending current market shares
• Product portfolio diversification
Data on generating the market entry rationale
• OPEX approximations affect the competitiveness of a possible products after a market entry
• Internal capabilities to establish a competitive product into the market
Data on running the operations after a market entry
• Pre-planned and analysed scenario analysis helps the strategy execution in a nascent market with a lot of changes happening in an operative environment
• Increasing competition verifies the market attractiveness
Post-entry data leverage as part of a strategy process
Data usage in strategic decision-making

Figure 14: Data structure

3.5 Ensuring reliability and validity of data

The reliability and validity of both data and the results defines the quality and credibility of the research (Yin, 2017). Saunders et al. (2009, p. 156) state reliability of the research referring to producing consistent findings about the underlying phenomenon from selected data collection techniques and analysis. In the other words, the same results should be reached on different occasions and by different researchers. As Saunders et al. (2009, pp. 156-157) argue there are four main threats to reliability that are especially noteworthy in qualitative and interview-based research: participant error, participant bias, observer error and observer bias.

Interview participant errors and biases were mitigated by ensuring anonymity of responses, selecting comfortable interview conditions when it comes to time and preparation and carefully selecting the participants to have diverse backgrounds. Observer errors and bias were mitigated firstly by having weekly meetings with the advisor of the thesis and a person from the case company that had conducted a master's thesis a couple of months before the beginning of this research. However, further research may take place in this field as some detailed fields within the cybersecurity remain subjective such as if it's preferred to use password managers to have strong and different passwords to every system but containing a risk of losing all your passwords and if your passwords manager application is attacked. Secondly, having the virtual interviews recorded enabled the observer to return to interviews and the responses in them. However, as the growth and change in the cybersecurity is rapid with its double-digit growth rates similar results may be challenging to achieve in some time as the underlying operative environment changes and evolves all the time.

Saunders et al. (2009, p. 157) defines validity to signify whether the research explains the studied phenomenon and the study considers the correctness of the results. Therefore, multiple people from multiple organizations from multiple backgrounds were interviewed to enhance the validity of the thesis and thus enabling the researcher to form an objective vision of the topic. Moreover, the current literature is for the most part in line with the results of this study.

4. Results

The results of this thesis are presented and analysed in this chapter with the methods explained in chapter 3. This section is structured into four subchapters corresponding to the aggregated dimensions in Figure 14. Each of these subchapters will be further divided into 2nd-order theme related chapters where 1st-order code quotes from the informants are linked into the respective theme of the chapter (Gioia et al., 2012).

This results section will be divided into three main chapters based on phase in a strategy process that the findings handle. In the first chapter, the extent of the data available for a strategy process is under consideration. Next in the second chapter, the focus is on organizations operations in networks and how the data is handled and shared with other stakeholders. Finally in chapter 4.3, the perspective lies on what angle data represents more generally in a strategy process especially when having an iterative approach due to dynamics of a nascent market (Ott et al., 2017). The visualised model of this section can be seen from Figure 15.

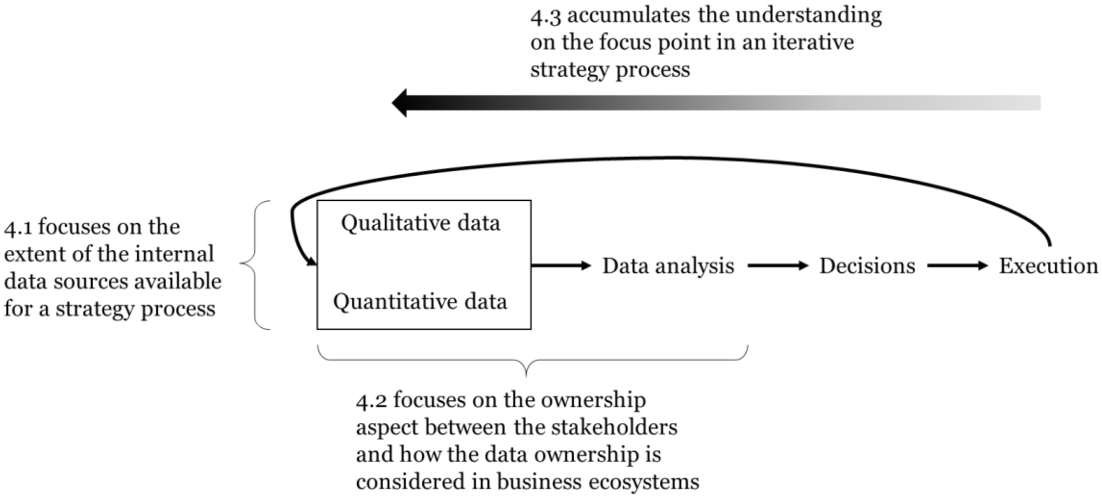

4.3 accumulates the understanding on the focus point in an iterative strategy process
4.1 focuses on the extent of the internal data sources available for a strategy process
Qualitative data
Quantitative data
Data analysis
Decisions
Execution
4.2 focuses on the ownership aspect between the stakeholders and how the data ownership is considered in business ecosystems

Figure 15: Theoretical model on data ownership in a strategy process

4.1 Creating value by data ownership in a strategy process

Having the right data, right time and the right analysis of data can form a better and more fact-based understanding of the business situation than a pure gut feeling. As the wrong type of data or data that wrongly describes the underlying phenomenon might be harmful for businesses, organizations must identify the key data sources that they require in their day-to-day operations and how that data should be secured and shared with other entities.

4.1.1 Effects of the criticality and the use case of data on ownership

Empirical data in this chapter highlights the importance for organizations of owning the data that they are handling. The respondents supported software users to own their data regarding their operations while the other types of data that are irrelevant or cause network effects can be shared to other parties.

Organizations' willingness to own data rather than share it depends on multiple factors. First, the criticality of the data determines organizations' willingness and even ability to share the data. For example, in the EU, GDPR has set the standards on how health organizations can handle and share their data over patients due to the sensitivity and criticality of the information. Elsewhere, some companies might want to keep their data over the next strategic actions as secure as possible to hinder the actions from the competitors.

> *"My strong view is that especially operationally and strategically critical data should be owned by the underlying organization. Then whether to share or give other parties an access to that data depends on the benefits and the possible network effects." – Informant 3*

> *"If we think about business-critical data then the ownership should be there at the company as it's exclusively the most*

> *important data for running the business and thus unwilled to be shared. But then the diagnostics and the like data can be shared outside." – Informant 1*

As both previous informants highlight, there are differences between what data should be owned and what kind of data could be shared. This shareable data varies between the companies and industries which data can be classified to be owned and which could be enabled to be shared as individual organizations benefit from different data in separate ways.

> *"There can be so much data to be used that focusing is needed. By focusing you can differentiate from the competitors such that the competitors find it irrelevant to gather the data that you find relevant. Then it's a question how you gather that data, is it available within your field of focus, or if you need to enrich the data such that your competitors won't have it available." – Informant 13*

Applying the criticality and focus of the data into the cybersecurity context, the businesses and organizations should familiarize themselves regarding their data and identify the criticality of it and how the criticality should be arranged when it comes to accesses and data ownership.

> *"Firms should define what the data is, but more and more companies should think about whether they understand what their critical data is, who can access that data and moreover where that data is located." – Informant 4*

Cybersecurity product companies evaluate previously mentioned aspects when designing their products for customers while balancing how they can gather data themselves securely from their installed product base.

> *"We have three types of data: analytics data that describes how our products are used. [...] Debug data that describes the potential technical problems of products. [...] And then there is the production data of the customers related to their own operations." – Informant 4*

From these three data types, cybersecurity vendors are interested in analytics and debugging data to develop their own products. Customers' production data is something that is irrelevant and unnecessary in cybersecurity operations from vendors' perspective. However, as some of the analytics data and debug data contain or might be linked to customers'' critical information the procedure to avoid sharing confidential data is to aggregate and anonymize the primal data source.

4.1.2 Product usage and analytics data

As described in the previous chapter, analytics and debug data would be the sort of data in cloud services that are among the interests of a cybersecurity vendor. These types of data would unveil how the products are used and how the future R&D resources should be allocated in order to support the user experience. Again on the other hand, the end users' production data is not necessary within the interest of a cybersecurity vendor.

Product usage and analytics data would create value for a cybersecurity product vendor on how to further develop their current products and what new features and completely new product families should be introduced to serve the current customer base better and potentially enlarge it.

> *"Understanding on industry level what features are used in the products, how much, what types of connections are formed, how much data goes through them and what types of end*

> *devices are used would be valuable data. It would help develop the products to better match the end user demand." – Informant 4*

In addition to product and feature usage data, the traffic data would be an interesting option to have access to as the malicious usage would be more possible to be recognized.

> *"If we talk about machine learning and AI, I believe that it would be beneficial if we got some behavioural patterns of the end users. Then we could develop some intelligent automatic solutions. Gathering that kind of data would be valuable to both a vendor and a customer." – Informant 1*

These mentioned data sources would be value-creating to have access to, but the products typically collect the information regarding some other technical parameters.

> *"In principle we see the validity of a license and the timestamp when the license was previously checked in addition to IP address, the version of an operating system of a device and the version of the product that is running." – Informant 1*

Cloud services and SaaS products might require more data to be shared to the vendor in the sense of a billing basis as the services can be priced based on the amount of software used, the number of users or a time spent in the service.

> *"We have data related to how many messages have been sent via our solution and how many users as its a typical billing basis. In addition, some of the products might have transaction-based billing. We get data over how many signatures have been made, how many forms have been filled,*

> *how many individual users and email addresses there are. The billing is based on these things." – Informant 6*

4.1.3 Customer relationship data

When talking about strategy processes, not only technical product related data create value alone. The organization that is conducting a strategy also requires intelligence related to its current and potential customers and how they are making their decisions. This type of data is typically available to an organization's current customer base which might provide competitive advantage. Knowledge over customer relations is highlighted especially in the cybersecurity industry as some of the sales cycles last due to the strategic importance of defensive data related investments among the customers.

Value creating data for a cybersecurity product vendor can also be something else than technical data gathered from the products. Having the market data in place would describe the changes in the market environment from more top level rather than at the grass root level of one customer.

> *"As a growth company within a growth industry all the data that is connected to sales is central and important. Customer behaviour and satisfaction are two things that pop into my mind in the first place." – Informant 12*

As sales processes are found important in the growing field of cybersecurity, the sales related data would create value and improve targeting the right kind of marketing efforts and approaches to the right people. As the cybersecurity field is a relatively technical subject, the data over the target audience and the level of technicality in sales and marketing activities would create value to ensure that the decision-makers have the right information in their hands. Moreover, some of the cybersecurity purchases are bureaucratic processes with multiple agreements required in the customer organization which causes inertia in the sales process and decision making.

> *"It would be important to understand the reason why a customer has bought a solution from us, who are the decision-makers in the customer organization. This information can be leveraged in new customer acquisition. Also, the data on the IT environments and their development trends would be valuable as only few or none are willing to publicly describe their IT environments." – Informant 3*

Having the regulative and highly technical complexity involved in the cybersecurity field, some of the more business-oriented decision-making executives might be unaware of the cybersecurity situation in their business operations. Thus, identifying a right target audience when it comes to data ownership and sharing points of view in cybersecurity operations.

> *"It [cybersecurity] is an intangible thing and possibly challenging to translate into 'executive language' as we are talking about thieves in the WLAN environment, not in the physical world." – Informant 8*

As the decision-makers might be unaware of the whole cybersecurity theme, it's easy to delay the cybersecurity investments. The typical business decisions are made for example by discovering the payback times, rate of returns and discounted cash flows of the investment. In that framework some of the cybersecurity investments are easy to postpone as the possible returns are more abstract to model in the sense of not having the costs by a possible cyberattack or data breach etc.

> *"Although the customers know that they have vulnerabilities in their cybersecurity processes, they have so many other*

> *things to be done. Therefore, cybersecurity might not be the number 1 priority but number 2 or 3 or these things might be even remaining undone." – Informant 1*

4.1.4 Value through network effects

Data ownership can benefit an owner in other ways than analysing the raw data by the owning organization itself. The respondent highlighted that in today's connected world sharing your data is sometimes even necessary to running the operations within a multicompany ecosystem.

One of the benefits of data-sharing is that third parties can develop their products that interact with the data sharing entity or that the third party can analyse the data in a way that the original data owner has no resources to execute. Sharing data can be seen to develop the ecosystem that the company operates.

> *"We have had internally a dialogue which data can be classified to be IPR, and which is general. Then we have shared quite openly the data with our competitors. Thus, we believe that the overall cybersecurity will become better and in the long run everyone will benefit from it." – Informant 11*

Although the network effects might take place by sharing some data with direct competitors, companies can also have a bit gentler approach to data sharing.

> *"More and more companies see that if owning some data is not their core and there is something valuable in that data, why not grant third parties access to it. If I'm not going to use the data, then it would be better to give it to someone else who can capitalize on it by creating additional value from it." – Informant 12*

Sharing data can help organizations to develop their own products and services but also other parties can analyse the data in a different way or with different capabilities that can benefit all the participants in the data sharing ecosystem.

> *"Organizations have a possibility to make decisions whether they want to be a platform in some area to provide edge benefits to counterparts in that ecosystem. [...] If the data to be shared has matured to a kind of bulk and thus challenging to create additional value with it then one can openly share the data. Thus, other organizations can process the data further that can create new business opportunities within the ecosystem." – Informant 13*

Informant 16 gave a practical example of what it means to operate in an ecosystem of multiple stakeholders and how the data is being organized among the counterparties.

> *"I think it's modern and our operating model is strongly based on networks. For example, one of our projects contained 13 different companies to develop an optimal solution. This type of project wouldn't have succeeded with our internal capabilities, but the network had the capabilities. However, one must make sure that in these types of networks you can provide value." – Informant 16*

4.2 Factors influencing the adoption of data-sharing

The cloud is here to stay. As organizations operate to an increasing degree in cloud environments data sharing becomes unrecognizably more common as the data in the cloud environment becomes technically shared with the cloud provider whether the provider has any intention to use the data or not. This cloudification has enabled users to share data more efficiently within an organization and across organizations.

4.2.1 Increasing adoption of cloud in enterprise software

Firstly, technological development has simplified the data-sharing scheme as data can be shared across devices, people and organizations automatically and continuously with cloud services and applications. As the technological barriers are being diminished due the rise of cloud, the mental models of siloed organizations need to be tackled.

Both supply side growth and consumption habits have driven the increased popularity of cloud. Firstly, the possible obvious bottleneck of cloud supply has shrunk during the last decade. Nowadays, cloud software options are available for almost every field in the software market, if not all of them.

> *"Cloud has changed the whole data-sharing game radically. The change started somewhere in 2010 and then a larger cloudification has happened." – Informant 6*

As cloud has been available for both enterprise and consumer market, the consumption habits of individuals and thus also decision-makers have got used to using cloud services in their day-to-day life.

> *"Everyone is used to the fact that Google or Facebook or others have permission to your personal data when using the systems. In the enterprise world the familiarity of giving permission is not yet at the same level but I guess that it will ease as you are used to the data-sharing with vendor nature in the consumer application side." – Informant 1*

4.2.2 The benefits of data-sharing

Two things have to be accomplished in order for data to be shared. Firstly, the sharing and receiving party have to trust each other. And secondly, sharing data must benefit both parties instantly or in a longer horizon. Otherwise, companies have a natural opportunity to not share their data if they feel that the other party or parties are unreliable, and the benefits are unclear.

Data-sharing is irrelevant if it doesn't benefit any counterpart. The starting point for the whole data-sharing discussion is how the data-sharing benefits the sharer whether the benefit is a financial one or result in enhanced products and services.

> *"If a vendor is trusted, terms & conditions, contracts are done well and the supplying organization can take care of their cybersecurity, it would benefit all [the counterparts if data was shared]." – Informant 12*

The respondents had a wide range of approaches to data-sharing where the most cautious opinions suggested not to share any data with other entities while some other respondent was willing to share data even with the direct competitors. Also, in the middle of these extremes the attitudes varied among the interviewees.

> *"Cybersecurity is a big part in successfully delivering and developing digital services. Success is tied up with how we cooperate with others and who we cooperate with. And if we cooperate, we must share data and give it forward. All kinds of sensitive data can be anonymised. [...] We have had internally a dialogue which data can be classified to be IPR, and which is general. Then we have shared quite openly the data with our competitors. Thus, we believe that the overall*

> *cybersecurity will become better and in the long run everyone will benefit from it." – Informant 11*

> *"I'm not extremely excited about sharing data. The reason is the security problems such as the data should be anonymized, all the identification information etc should be removed that could be traced back. This type of clean data could be beneficial but often the shared data contains all kinds of things and it's shared across the internet." – Informant 5*

The rest of the attitudes stood between these points of view by weighting the direction of sharing and the outcomes of the sharing data.

> *"We are ready to use data towards customer interface, but we are a bit jealous outside because we don't want a third party to operate in our field of maintenance operations. So, all the data received is welcomed but outside we are a black hole." – Informant 9*

Despite the possibilities of having machine-to-machine communication in the cybersecurity data-sharing scheme, person-to-person communications still have their place in informing the personal networks of the possible vulnerabilities in cybersecurity. On the other hand, the regulation also sets the standards on how we can share and benefit from sharing data.

> *"Of course, we all have our own networks where we share information more liberally. But as we are a listed company, we can't talk as openly as possible within our networks about the cybersecurity setbacks that we face but rather we have to publish a stock market release." – Informant 10*

4.2.3 Carrying the third-party risk in data-sharing ecosystems

Leaving the data ownership into other entities hands in a strategy process, organizations carry a risk that the needed data is unavailable when required. Therefore, there is a clear incentive to carry business critical information by yourself or at least have other ways of accessing the not so critical data.

Operating in the cloud might save some costs for customers but at the same time it increases the uncertainty over the safety of the cloud service. As the users outsource hosting the service, they outsource part of the security of their operations as well as now they don't have all the wires in their hands but rather, they are dependent on cloud service providers' confidentiality, integrity and availability of data and service.

> *"Firms are using more cloud but cybersecurity wise but of course it would be better to host your own critical information by yourself." – Informant 5*

Having the data on someone else's servers, one can never be completely sure who can access the data and how the data centres etc are secured.

> *"It's always challenging if [case company's] customers' systems send information outside because customers are distrustful over if any data is being sent out and used wrongly." – Informant 1*

Typically cloud service providers have service level agreements with their customers where they guarantee certain aspects in their cloud service such as availability etc. On the other hand, these agreements enable the providers to have access to the log data to discover possible bugs and vulnerabilities in their product.

> *"A SaaS product would require access to log data on demand because if there was some problem, we would be the one who has the responsibility over running the service." – Informant 4*

On the downside, if the organization uses cloud services it has to trust that the cloud service provider takes care of the security in their applications and updates the security for the users within the cloud.

> *"The problem of customers in these cloud services is that if the cloud or a SaaS vendor doesn't have the cybersecurity updates. Then we will have these kinds of things that have popped up to the public news. A couple of times, all the information has gone." – Informant 5*

4.2.4 Price and financial side of data-sharing

Data is shared more easily if it can be verified to provide cost savings or potentially new revenue streams. Only the most critical software applications and data can underrate the cost advantage of the cloud applications which enables easier data-sharing basis.

Obviously, price and lifecycle costs interest the users of an IT solution. That's also the case in acquiring cybersecurity software although the short-term cost optimization would mean postponing the cybersecurity investment thus taking a risk of being a victim of a cyberattack. Cloud environments have brought clear cost advantages to the whole IT industry which can resonate to end-user interfaces as lower prices.

> *"Cloud is an area where economies of scale are strong. [...] R&D costs play a huge part in total costs for cloud companies and having a large customer base enables depreciating the*

> *costs over the larger audience which challenges the market entries of new players." – Informant 7*

As the development costs are covered by several organizations and thus the need for overlapping costs is lesser.

> *"Less money is needed to produce more when more investors and customers are involved in a project." – Informant 12*

Pricing and lifecycle costs can also be the key purchasing criteria in corporations as informant 8 describes based on his Central European experiences regardless of the data ownership and sharing side considerations involved.

> *"It's clear that there is not even a topic for a conversation [whether to choose cloud or on-premises] in German speaking countries. If the cloud has a price advantage it's all clear and the cloud is chosen." – Informant 8*

Also, the financial benefits of data sharing stand in the future and thus it's more challenging to argue for the executives the logic behind for the customer. In principle, customers feel like the benefits of data-sharing wouldn't exceed the disadvantages quick enough.

> *"Data collection in our case wouldn't mean instant price discount which might be the case for example in grocery stores, but the data collection would mean better service quality in the future." – Informant 3*

4.2.5 Effects of regulatory frameworks on data-sharing

Although businesses might want to gather and share data with others in order to develop their products, offering and corporate strategies, all the companies operate under some regulative forces. Data privacy and security has lately been within the regulators' interest in how companies can leverage the data they are operating with. However, analytics and debugging data that is within vendors' interest doesn't belong to regulators main concerns as they are more cautious over citizens personal data.

The regulative environment also frames the cybersecurity industry at the top level and the discussion is active over tech giants' roles and their data collection. One of the largest regulative forces in the EU area has been the adoption of GDPR that all companies must align with. This affects data ownership relationships and data sharing. However, GDPR takes place only in the EU area and within the industries that handle consumer level data.

> *"Some industries, such as healthcare and finance, handle more personal level data and thus they are more affected by the GDPR as the data handling requirements have changed to more critical. However, for example in manufacturing there is less legislation related to data ownership." – Informant 3*

Many of the data ownership regulations are local or at some union level. This makes the following and data sharing possibilities challenging for companies that operate globally. Therefore, an interviewed company has adopted an external tool to track the changes in data sharing and handling related legislation.

> *"We use a product that compares legislation in different areas. As we operate globally, it's almost impossible to keep track of every change in all of the countries." – Informant 10*

Although governments and other legislators can issue different kinds of laws and settings, platform companies themselves have raised concerns over how the data is handled on their platforms.

> *"GDPR has changed this industry a lot. In addition, for example Apple itself has changed its tracking protection in iOS 15 to much stricter, so one is not able track the users so easily."*
> *– Informant 5*

Partly due to regulation and companies' own governance factors also their procurement processes might be time consuming, bureaucratic, and regulation driven or even lacking to ability to proof of cybersecure operations might lead to not being considered when agreeing on contracts and partnerships.

> *"Many of our customers operate in highly regulated environments. We received about a 500-item request for a proposal list from one of our customers to which all the questions must be answered and checkbox features to be covered. Therefore, when you go selling a PAM solution you have to have these features in your product." – Informant 4*

> *"Today it would be challenging to think about taking some SaaS product to our [system integrator] offering, that is not GDPR compliant in cybersecurity sense or doesn't fulfil the standards" – Informant 12*

4.3 Data usage in strategic decision-making

Cybersecurity market is growing and evolving rapidly which leads to a situation where vendors' existing portfolio might be substituted if not developed. On the other hand, the providers can enter new markets and introduce new products to remain competitive in the market. Although the underlying cybersecurity market is growing double-digit percentage figures annually, informants highlight that the development of a completely new, best-of-breed, and future-proof product is likely to take time while sometimes selling and marketing these next-generation products can be surprisingly time-consuming as customers' ICT architecture might not be aligned with the most novel technology.

4.3.1 Data on generating the market entry rationale

The previous chapters have described data ownership, collection and sharing while this chapter focuses on the usage of that data especially with the strategy process point of view. When companies are expanding their businesses to reach wider than their current markets, strategy professionals should analyse the rationale to enter a market based on the data and other capabilities that they possess.

As a challenger in the market, one of the strategic initiatives of the case company is business growth. Entering a new market is one way of expanding the business and total addressable market if a vendor sees that its portfolio on the current market provides unfavourable organic growth possibilities or the existing product portfolio is at a risk of being replaced by other substitutes.

> *"I think there are two types of market entries. Firstly, one can grow your revenue by searching for new growth possibilities as the old markets are 'repressed' or there are not enough growth possibilities. Secondly, one can enter a new pocket in the market to defend or support vendor's current businesses."*
> *– Informant 3*

Growing the business is one objective in entering new markets but as a product portfolio company diversification is another point of view in expanding the offering. The case company acquired another cybersecurity product company in the spring 2021. The greater part of the revenue of the acquired company is generated from one product and thus it carries a risk of the changes in its underlying market and product competitiveness. The acquisition diversifies the risk thus benefitting the acquired and the case company.

> *"The growth expectations drive us to the new markets. Also, our [acquired company] product portfolio consisted of four products, with email encryption bringing a major part of the total revenue so it's a concrete risk. For example, if something happened suddenly in the email market globally it would be a huge risk for us. It would have basically meant bankruptcy for us. [...] But after this acquisition, the one product risk is more mitigated, and we can even cross-sell our products to others installed base." – Informant 6*

It's a well-known dilemma in the cybersecurity industry that the attackers are developing their methods and using novel technologies. For a cybersecurity vendor it means that they can either enhance some current solution or develop a new layer of cybersecurity function. On the one hand the current markets can be seen to be highly competitive while on the other hand developing something completely new and differentiating requires likely years of work thus meaning that the novel solutions must be aimed at far into the future where the final demand can be challenging to model.

> *"The B2B world is moving towards passwordless authentication to certificates and Just-in-Time Zero Trust. For example, Microsoft has pushed all kinds of fingerprint authentication and others. We were even too early on the move to this market back in the days but now when there is demand for that type of product, we have something to offer rather than beginning to develop it for years." – Informant 5*

4.3.2 Data on running the operations after a market entry

Just the data on the market top line potential isn't enough when conducting a market entry analysis. Data and estimates on the developing phase and running the day-to-day operations enhance aligning the company's resources in an efficient way. Data ownership on operating a different business model would reduce the risk of failure and accurate the business case estimates when conducting a strategy process,

Changing a business model requires time and effort but also learning new ways of working and aligning company's operations in a different way which might be challenging. That's also the case in the IT world in changing from on-premises to SaaS models. Having a SaaS operation running requires support and other costs for a vendor.

> *"If your SaaS is active 24/7, it requires at least 5 FTE in support which is a big cost. [...] Big corporations are unwilling to purchase SaaS. [...] If a small company can afford to pay a few hundreds per month for SaaS, the break-even isn't quickly accomplished as the selling organization must be aligned to sell to many small SMEs. [...] In addition to paying peanuts, SMEs require the same amount of software maintenance as the big ones." – Informant 5*

Naturally, developing the code required in the cybersecurity products is a manual operation which means that internal capabilities could limit a vendor's possibilities to start building novel solutions with the most advanced technologies.

> *"Never has somebody said we can't do this so that's why we shouldn't do it. So internally we have these kinds of capabilities. [...] We haven't developed a strategy purely on*

> *what our current capabilities are, so that has not been the driver [to enter a market]." – Informant 2*

4.3.3 Post-entry data leverage as part of a strategy process

Strategy processes are iterative ones especially in nascent, rapidly growing and changing market environments. Thus, the strategy process should not end with the decision whether to enter a market. The challenges may arise from internally aligning the business operation to match the required operating model after the market entry or the customer demand may differ or change from the earlier analysis.

Market entries are iterative and long-lasting processes meaning that the strategy work should continue after initially entering the market. Nascent and rapidly growing markets are under continuous change and therefore companies in these markets must evaluate their business environment and strategies regularly while having a wide range of strategic tools in the toolbox analysed beforehand.

> *"We had a strategy round in the beginning of the summer but already in the autumn we adjusted the strategy to another direction. The most important thing is that the counterparts have already had the discussions if the change of direction is required. Thus, everyone should know why the direction is changed as the discussions are already held." – Informant 1*

The development of a product might begin years before the demand rises. Therefore, the overall development process needs to be iterative and the value of the first paying customers are important.

> *"The biggest mistake in developing products just based on inhouse capabilities is to not verify the market, especially if we are talking about a completely new market. [...] Nobody*

> *knows the exact customer need in the beginning but having the discussions with the first customers helps at how the product should be developed further." – Informant 1*

Also, other participants in the market than customers should be followed. If the entered market or a new product can provide excess returns, the other parties in the market should be also interested to enter the market.

> *"If someone invents a breaking technology or a completely new product, often a good increased competition is a good validation [that there exists demand]." – Informant 5*

5. Discussion and Conclusions

In this section this thesis is concluded, and the key findings are discussed. First, the three research questions are answered. After answering the research questions, theoretical and managerial implications are conducted. In the end, limitations and possibilities for further research are discussed.

5.1 Key findings

The key findings of this thesis are presented in this chapter by answering the research questions. The findings are based on both case company's internal and external semi-structured expert interviews and the insights shared by the 16 interviewed professionals.

RQ1: What is the role of data ownership when formulating a market entry strategy?

Data-based decision-making can overcome the strategic decisions of competitors if at least one of two value-creating data analysis conditions are met. Firstly, the decision-makers can have capabilities to analyse the available data better and faster than their competitors as Miller and Mork (2013) suggest that the value of data springs from the analysis. Secondly, strategic decisions can outpace alternative strategic decisions if the data analysis is conducted with data that is unavailable for other decision-makers producing data into a rare and imitable asset (Akhtar et al., 2019; Harris & Graig, 2010). Having the data unavailable for the competitors, a decision-making organization must either own the data or have access to it within an ecosystem where some other entity owns the data but doesn't share the data with the competitors of the decision-making organization.

In addition to having rare and challenging to imitate data with possibly advanced data analysis capabilities, the quantity of the valuable data needs to be sufficient but not abundant in order to avoid suffering from the reduced quality of the data. The analysed data must represent the underlying phenomenon and the selected data sources to be used in strategic analysis must be unbiased. Otherwise, the data will mislead the analysts and decisions-makers. (Harris & Graig, 2010)

Considering the market entry strategies, the data ownership aspect becomes more relevant the closer the market to enter is company's current business or as Siggelkow (2002) calls this to be thickening the portfolio. If the market to enter is close to the company's current portfolio and thus a company is thickening its offering, it can leverage the data and information it has gathered from the nearby market with possibly overlapping customer base. On the other hand, if an entree has a limited number of touching points at market to enter or as Siggelkow (2002) calls this as patching, the organization must either analyse data better than incumbents or it has to secure to have the valuable data in hand from the other stakeholders in its network. This study accumulates the understanding on Siggelkow's (2002) work by describing the process behind the strategic choices of thickening, patching, coasting and trimming and how a strategic decision-maker gains access to data in-house or via stakeholders while Siggelkow describes the strategic choices given that the needed information is already in decision-makers hands.

So, as a result of the research question, owning data provides a parallel option with having access to third party's data to conduct data analysis when organizations work on market entry strategies. Having the ownership over the data used in a strategy process may enable rarity and low imitability of the data but it doesn't solely secure a strong strategy without third party data that a company has access to. Finally to be mentioned, market entry strategies differ from day-to-day strategies as market entry strategies to some specific markets are typically done once or at most a few times whereas day-to-day strategies are more continuous operation thus meaning different approach on the data sources used in a process.

RQ2: What kind of data is valuable in a strategy process in the cybersecurity industry?

Shatrevich and Gaile-Sarkane (2015) define strategy to be formed based on the competitive environment that the corporation is doing business in. This is also the case in the cybersecurity industry as the market related data was found to be important, value-creating, and centric in a strategy process. In the interviews conducted for this thesis, market data was referred to general market factors such as growth rates and sizes of the markets, competition, and the offering of the competitors. Although this information was found useful, it's largely available for everyone or the data is publicly for sale such that every player in the market can reach it which means that building competitive advantage with market intelligence can be challenging to achieve without advanced analysis.

From technical point of view, interviewees found that there exist three types of data for a cybersecurity product company: analytics data, debug data and customers' production data under the security. The latest of these three is a type of data that a cybersecurity vendor is unwilling to possess as that data describes customers' operations meaning no or limited value in the cybersecurity vendor's strategic decision-making. Access to debug data in turn facilitates cybersecurity vendor's day-to-day research and product development operations. In the cloud and SaaS world debug data would possibly be easily available for a vendor than in on-premises models.

Finally from the technical data sources, the analytics data represents the most valuable form of technical data for a cybersecurity vendor strategic-wise and thus it can be identified as a data asset that can enhance companies' value creation (Beckwith, 2020). Analytics data of the products would enable vendors to track for example how much and what parts of their products have been used among the customers. This data can be utilized in further developing existing products or entering a new market with possibly new products. This analytics data accumulates vendors' understanding of the customers enabling vendors to steer their strategies to target the offering more towards the end user demand. For analytics and debug data, the choice over cloud or on-premises delivery model determines possible real time data availability through an access or data ownership for a vendor leaving the cloud or on-premises decision to a strategic choice for a vendor when offering and a customer when purchasing cybersecurity products and solutions.

Besides form the product-centric technical data, customer relation data both generates first-hand information about the market, customers' IT architecture and the end user decision-making processes. This information gives a cybersecurity vendor better understanding on the status of the customers' demand and the state of their IT environments. Also, having the data on the customers' decision-making processes and decision-making authorities enables targeting the marketing approach with a right balance between technicality and commerciality. This customer related data typically represents data that the vendor owns, and the form of this data is confidential thus meaning that the customers are unwilling to share this data more than the minimum amount required which make this type of data rare and a source to leverage in strategy processes.

RQ3: How could the case company leverage data ownership as part of its strategy?

The findings suggest that the case company should leverage the data ownership in two ways. Firstly, enhancing data collection capabilities are the basis of the data-driven decision-making process. This can be done by introducing features in the product portfolio that collects analytics data on the usage of the products and their features. This type of data would accumulate understanding through data analysis on the customer behaviour with the products.

Secondly, as all the data utilised in a strategy process can be challenging to gather by yourself and thus supporting the data-sharing attitudes in the business ecosystem would broaden the data source base that could be leveraged in the strategy work. This procedure enables opening external data streams to be used that would otherwise be left into shadow.

From a data ownership point of view, being an attractive partner in an ecosystem where data is shared among the counterparts one must have something to offer whether it may be an additional data source or something else that creates value to the ecosystem.

5.2 Theoretical implications

This thesis has several contributions by extending the theories and aligning with existing literature. Firstly, this study creates novel understanding on the importance of owning data as part of the strategy process. Previous research on data-driven decision-making has highlighted the importance of gathering a wide range of data from multiple sources when conducting data analysis but has lacked the ownership point of view (Miller & Mork, 2013). Also, this study develops an understanding on how the rarity and low-imitability of data should be considered in a strategy process while the previous work by Akhtar et al. (2019) and Harris and Graig (2010) has focused more generally on data-driven decision-making rather than focusing on strategy processes.

This thesis offers a new layer to be considered in the iterative market entry strategies and strategy-by-doing approaches in nascent markets (Ott et al., 2017; McDonald & Eisenhardt, 2020; Chen et al., 2021; Helfat & Peteraf, 2003). These theories rationalize the phased strategy processes with agilely developing the strategy as time goes by. However, this study pinpoints some areas for further research when it comes to data ownership in an iterative strategy process and strategy-by-doing. When an organization owns its own data itself, it can leverage the insights in the strategy processes the organization is conducting. The ownership of data enriches the strategy process such that it allows the organization to gather a strategy that is based on the insights and intelligence that its competitors are unable to conduct. Also, data ownership acts as a catalyst in scenario-based strategy processes where accumulating own data that is valuable, rare, and challenging to imitate can offer an organization support in its scenario work. The data structure in a strategy process with self-owned internal, ecosystem and public data sources is presented in Figure 16.

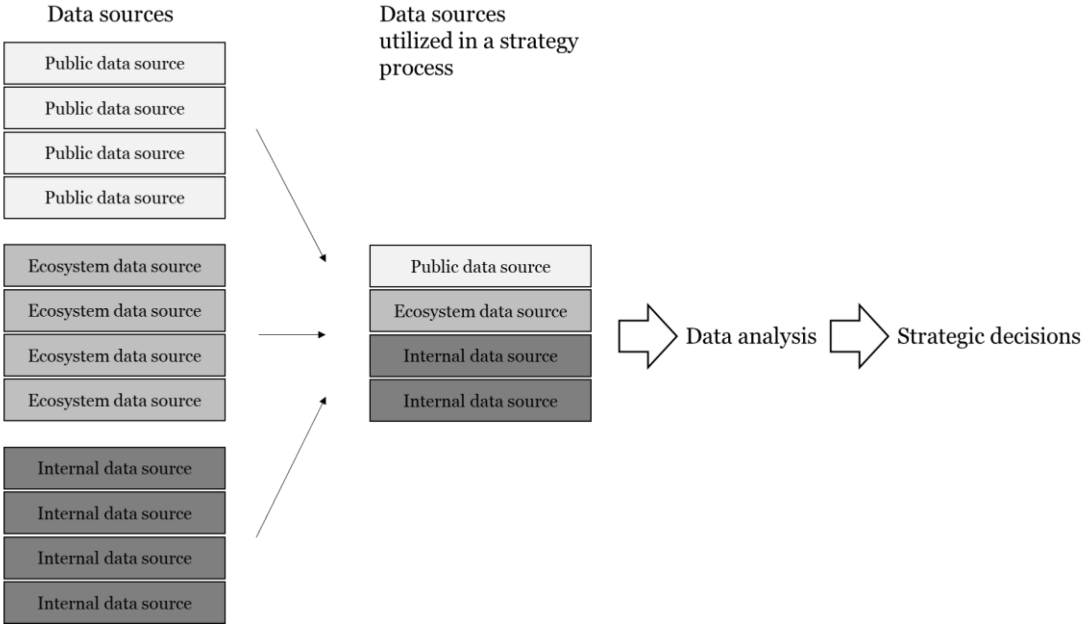


Figure 16: Data sources in a strategy process

Secondly, this study provides a breakdown on how the data ownership structure should be arranged in the cloud services and what data types should be shared with the cloud vendors. The findings support that the users should own their ‘own production data’ in the platforms. This structure somewhat differs from the biggest technology giants’ approach where they might own all the data within their platforms (Birch, Cochrane, & Ward, 2021). Other types of data, such as debug and analytics data, can or in some cases should be shared with the cloud service provider in order to secure the confidentiality, integrity and availability of the service (Soltys, 2020; Herrmann & Pridöhl, 2020). End users’ allowance to share data with the vendor enables enhanced product development and especially in the cybersecurity landscape to investigate possible vulnerabilities based on the analytics and debug data generated from the product.

In addition to data-sharing with the cloud service provider, this study supports Goasduff’s (2021) “must share data unless” mindset towards data-sharing with certain refinements after the word ‘unless’ that DalleMule and Davenport (2017) defines to be defensive data in their data-strategy spectrum. More offensive data with less legal, financial, compliance, and IT concerns could be shared within the ecosystems that the data-sharing organization operates. This type of data-sharing enables companies to create value to its ecosystem and thus benefiting from it also through not only via the network effects that the other ecosystem participant can generate but also as the data-sharing is not one-directional, an organization can leverage the data that the other participants own within the ecosystem. The visualisation of the data-sharing ecosystem with categorization into offensive and defensive data is presented in Figure 17.

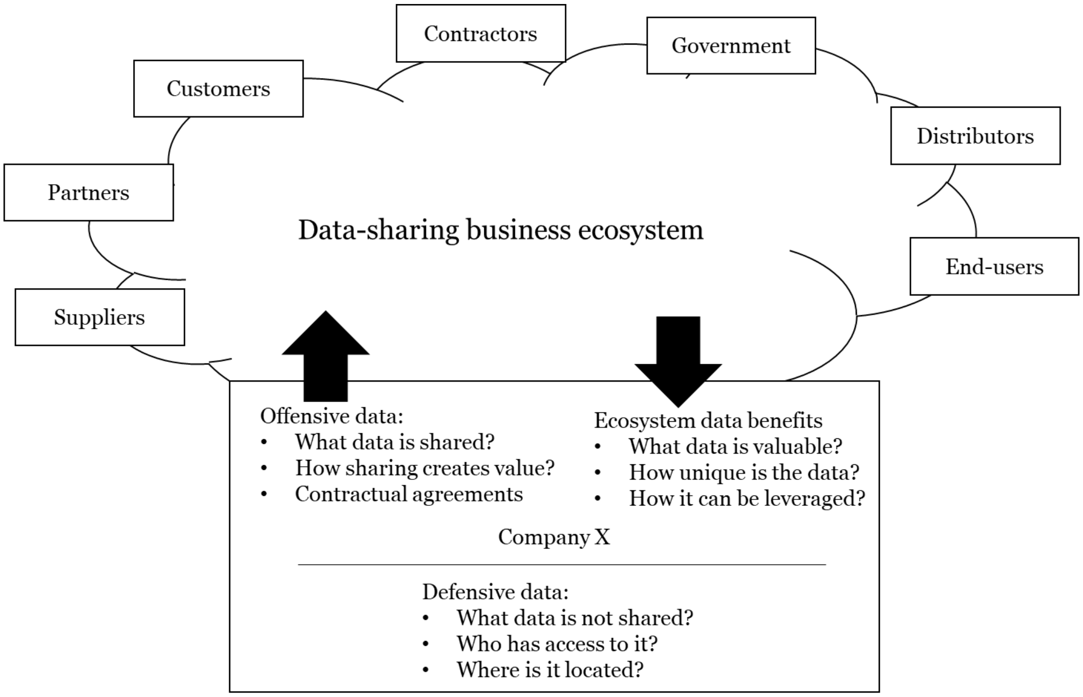


Figure 17: Data-sharing ecosystem

5.3 Managerial implications

The following implications are targeted at managers and executives within the fields of strategy and information security. Firstly, in order to benefit from data ownership in a strategy process, the data must be owned whether the data is located on-premises or in the cloud. The discussions in the interview sessions on multiple occasions began with pondering who owns the data when we use various business and consumer cloud applications which indicates that the ownership angle is often neglected when using different software.

As the organization owns the data, it can leverage not only the value created from the data analysis generated from multiple sources of raw data but also the organization is better able to handle the data-sharing and access permitting operations. When controlling the data-sharing scheme within the ecosystems, organizations should define their defensive data that remains solely in the original ownership and more offensive data that would enrich the ecosystems of the organization and its role in those ecosystems. On a practical level, defining offensive data and sharing it means that accesses and copies of the original data is shared rather than the ownership of the whole data or the data source. Therefore, sharing and allowing access to data differs from operating with tangible assets where sharing something means that the sharing entity abandons the asset whereas the dynamics in the data-sharing can be narrated to be copying the intangible asset.

5.4 Limitations and possibilities for further research

This thesis offers a point of view on data ownership's role in a strategy process. However, there still exists three main limitations in this thesis with each of which creating avenues for future research to be done. Firstly, other industrial contexts than cybersecurity might bring different points of views on data ownership. Particularly, future research could target data ownership studies on more data offensive industries as the stakeholders of the case company of this thesis lean more towards data defensive organizations according to the framework by DalleMule and Davenport (2017).

Secondly, the case company is a challenger in a large and rapidly growing cybersecurity market. Therefore, different companies with different customer and stakeholder bases might provide a different weighting on owning and sharing data, especially when talking about the largest cloud giants.

Thirdly, the interviewees were mostly Finnish, or they have lived in Finland for years. The use of cloud computing in the Finnish enterprises in 2021 was 75% which is clearly above the EU average of 42% (Eurostat, 2021). Therefore, the attitudes towards cloud services might vary with people from different geographical settings and thus future research on the topic with different geographical approaches would validate and generalize the findings.

6. References

Adner, R. (2013). *The Wide Lens: What Successful Innovators See That.* New York: Penguin.

Adner, R. (2017). Ecosystem as Structure: An Actionable Construct for Strategy. *Journal of Management, 43*(1), 39-58.

Adner, R., Puranam, P., & Zhu, F. (2019). What Is Different About Digital Strategy? From Quantitative to Qualitative Change. *Strategy Science, 4*(4), 253-261.

Akhtar, P., Frynas, J. G., Mellahi, K., & Ullah, S. (2019). Big Data-Savvy Teams' Skills, Big Data-Driven Actions and Business Performance. *British Journal of Management, 30*(2), 252-271.

Amit, R., & Zott, C. (2001). Value Creation in e-Business. *Strategic Management Journal, 22*, 493-520.

Anand, J., & Singh, H. (1997). Asset Redeployment, Acquisitions and Corporate Strategy in Declining Industries. *Strategic Management Journal, 18*(7), 99-118.

Ansoff, H. I. (1957). Strategies for Diversification. *Harvard Business Review, 35*(5), 113-124.

Atapour-Abarghouei, A., McGough, A. S., & Wall, D. S. (2020). Resolving the Cybersecurity Data Sharing Paradox to Scale Up Cybersecurity via a Co-production approach towards Data Sharing. *IEEE International Conference on Big Data*, 3867-3876.

Atkins, S., & Lawson, C. (2021). Cooperation amidst Competition: Cybersecurity Partnership in the US Financial Services Sector. *Journal of Cybersecurity, 7*(1), 1-11.

Ayoub, K., & Payne, K. (2016). Strategy in the Age of Artificial Intelligence. *Journal of Strategic Studies, 39*(5-6), 793-819.

Barney, J. (1991). Firm Resources and Sustained Competitive Advantage. *Journal of Management, 17*(1), 99-120.

Beckwith, F. (2020). Business Analytics Revisited: A Gap Analysis of Research and Practice. *ACIS 2020 Proceedings* (pp. 1-12). Association for Information Systems.

Bejtlich, R. (2004). *The Tao of Network Security Monitoring*. Boston: Addison-Wesley Professional.

Beynon-Davies, P., & Wang, Y. (2019). Deconstructing Information Sharing. *Journal of the Association for Information Systems, 20*(4), 476-498.

Bhatia, J., Breaux, T. D., Friedberg, L., Hibshi, H., & Smullen, D. (2016). Privacy Risk in Cybersecurity Data Sharing. *Proceedings of the 2016 ACM on Workshop on Information Sharing and Collaborative Security* (pp. 57-64). Vienna: ACM.

Birch, K., Cochrane, D. T., & Ward, C. (2021). Data as Asset? The Measurement, Governance, and Valuation of Digital Personal Data by Big Tech. *Big Data & Society, 8*(1), 1-15.

Black, J., Myles, g., & Hashimzade, N. (2017). *A Dictionary of Economics* (5 edition ed.). Oxford University Press.

Boudreau, K. J., Jeppesen, L. B., & Miric, M. (2021). Competing on Freemium: Digital Competition with Network Effects. *Strategic Management Journal*. From https://ssrn.com/abstract=2984546

Brilingaité, A., Bukauskaus, L., Juozapavicius, A., & Kutka, E. (2022). Overcoming Information-sharing Challenges in Cyber Defence Exercises. *Journal of Cybersecurity, 8*(1), 1-9.

Brittain, J. W., & Freeman, J. H. (1980). Organizational Proliferation and Density Dependent Selection. In J. Kimberly, & R. Miles, *The Organizational Life Cycle* (pp. 355-378). San Francisco: Jossey-Bass.

Campos, J., Sharma, P., Jantunen, E., Baglee, D., & Fumagalli, L. (2016). The Challenges of Cybersecurity Frameworks to Protect Data Required for the Development of Advanced Maintenance. *Procedia CIRP, 47*, 222-227.

Chen, L., Wang, M., Cui, L., & Li, S. (2021). Experience Base, Strategy-by-Doing and New Product Performance. *Strategic Management Journal, 42*, 1379-1398.

Cilluffo, F. J. (2017). The Finance Sector and Countering Cyberthreats: Lessons from the Front Lines. *RSA Conference*. San Francisco.

Cser, A., Balaouras, S., Pesa, K., & Dostie, P. (2018). *The Forrester Wave™: Privileged Identity Management, Q4 2018*. Forrester.

DalleMule, L., & Davenport, T. H. (2017). What's Your Data Strategy. *Harvard Business Review, 95*(3), 112-121.

Daniel Ani, U. P., He, H., & Tiwari, A. (2017). Review of Cybersecurity Issues in Industrial Critical Infrastructure: Manufacturing in Perspective. *Journal of Cyber Security Technology, 1*(1), 32-74.

Dubois, A., & Gadde, L.-E. (2002). Systematic Combining: An Abductive Approach to Case Research. *Journal of Business Research, 55*(7), 553-560.

Eisenhardt, K. M. (1989). Building Theories from Case Study Research. *The Academy of Management Review, 14*(4), 532-550.

Eurostat. (2021, December 9). *Cloud computing used by 42% of enterprises*. Retrieved February 10, 2022 from Eurostat:

https://ec.europa.eu/eurostat/web/products-eurostat-news/-/ddn-20211209-2

Fisher, T. (2009). *The Data Asset: How Smart Companies Govern Their Data for Business Success*. Hoboken, New Jersey: Wiley & Sons.

Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2012). Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology. *Organizational Research Methods, 16*(1), 15-31.

Giustiziero, G., Kretschmer, T., Somaya, D., & Wu, B. (2022). Hyperspecialization and Hyperscaling: A Resource-based Theory of the Digital Firm. *Strategic Management Journal*, 1-34.

Goasduff, L. (2021, May 20). *Data Sharing Is a Business Necessity to Accelerate Digital Business Necessity to Accelerate Digital Business*. Retrieved December 20, 2021 from Gartner: https://www.gartner.com/smarterwithgartner/data-sharing-is-a-business-necessity-to-accelerate-digital-business

Goodwin, C., Nicholas, J. P., Bryant, J., Ciglic, K., Kleiner, A., Kutterer, C., . . . Sullivan, K. (2015). A Framework for Cybersecurity Information Sharing and Risk Reduction. *Microsoft*.

Goyal, S. (2014). Public vs Private vs Hybrid vs Community - Cloud Computing: A Critical Review. *International Journal of Computer Network and Information Security, 6*(3), 20-29.

Gregory, R. W., Henfridsson, O., Kaganer, E., & Kyriakou, H. (2021). The Role of Artificial Intelligence and Data Network Effects for Creating User Value. *Academy of Management Review, 46*(3), 534-551.

Guo, Z., & Ma, D. (2018). A Model of Competition Between Perpetual Software and Software as a Service. *MIS Quarterly, 42*(1), 101-120.

Gupta, P., Seetharaman, A., & Raj, J. R. (2013). The Usage and Adoption of Cloud Computing by Small and Medium Businesses. *International Journal of Information Management, 33*(5), 861-874.

Harris, J. G., & Graig, E. (2010). How to Turn Data into a Strategic Asset. *Accenture*, 1-11.

Helfat, C. E., & Peteraf, M. A. (2003). The Dynamic Resource-based View: Capability Lifecycles. *Strategic Management Journal, 24*, 997-1010.

Herrmann, D., & Pridöhl, H. (2020). Basic Concepts and Models of Cybersecurity. In M. Christen, B. Gordijn, & M. Loi, *The Ethics of Cybersecurity* (pp. 11-44). Springer.

Holst, A. (2021, June 7). *Volume of Data/Information Created, Captured, Copied, and Consumed Worldwide from 2010 to 2025*. Retrieved October 20, 2021 from Statista: https://www.statista.com/statistics/871513/worldwide-data-created/#statisticContainer

Kazhdan, A. P. (1991). *The Oxford Dictionary of Byzantium*. Oxford University Press.

Kitsios, F., & Kamariotou, M. (2021). Artificial Intelligence and Business Strategy towards Digital Transformation: A Research Agenda. *Sustainability, 13*(4), 2025-2039.

Koepke, P. (2017). *Cybersecurity Information Sharing Incentives and Barriers*. Sloan School of Management. Cambridge: MIT University.

Kumar, N. S., Lakshmi, G. R., & Balamurugan, B. (2015). Enhanced Attribute Based Encryption for Cloud Computing. *Procedia Computer Science, 46*, 689-696.

Lanjouw, J. O., Pakes, A., & Putnam, J. (1998). How to Count Patents and Value Intellectual Property: The Uses of Patent Renewal and Application Data. *The Journal of Industrial Economics, 46*(4), 405-432.

Lassila, A. (2005). Moving from Product-based to Online Service Business. *IADIS International Conference e-Commerce*, (pp. 109-117).

Liao, C.-H., & Chen, C.-W. (2014). Network Externality and Incentive to Invest in Network Security. *Economic Modelling, 36*, 398-404.

Lilien, G. L., & Yoon, E. (1990). The Timing of Competitive Market Entry: An Exploratory Study of New Industrial Products. *Management Science, 36*(5), 568-585.

Liu, S. (2021, October 22). *Information Technology (IT) spending on enterprise software worldwide, from 2009 to 2022*. Retrieved December 22, 2021 from Statista: https://www.statista.com/statistics/203428/total-enterprise-software-revenue-forecast/

Lynch, L., & Hayes, T. (2011). Cloud Morphing: Shaping the Future of Cloud Computing Security and Audit. In B. Halpert, *Auditing Cloud Computing: A Security And Privacy Guide* (p. Chapter 9). Hoboken , New Jersey: John Wiley & Sons, Inc.

Mäkilä, T., Järvi, A., Rönkkö, M., & Nissilä, J. (2010). How to Define Software-as-a-Service – An Empirical Study of Finnish SaaS Providers. (pp. 115-124). Springer Berlin Heidelberg.

Marshall, C., & Rossman, G. B. (1989). *Designing Qualitative Research*. London: Sage.

McDonald, R. M., & Eisenhardt, K. M. (2020). Parallel Play: Stratups, Nascent markets, and Effective Business-model Design. *Administrative Science Quarterly, 65*(2), 483-523.

McLilly, L., & Qu, Y. (2020). Quantitatively Examining Service Requests of a Cloud-Based On-Demand Cybersecurity Service Solution for Small Businesses. *International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 116-121). Las Vegas, Nevada: IEEE.

Mell, P., & Grance, T. (2011). The NIST Definition of Cloud Computing. *NIST Special Publication 800-145* (pp. 1-3). National Institute of Standards and Technology.

Miller, H. G., & Mork, P. (2013). From Data to Decisions: A Value Chain for Big Data. *IT Professional, 15*(1), 57-59.

Miller, N. (2018, May 8). *With More Than 1,200 Cybersecurity Vendors in the Industry, How Do You Stand Out?* Retrieved November 4, 2021 from McAfee: https://www.mcafee.com/blogs/enterprise/with-more-than-1200-cybersecurity-vendors-in-the-industry-how-do-you-stand-out/

Mitchell, W. (1989). Whether and When? Probability and Timing of Incumbents' Entry into Emerging Industrial Subfields. *Administrative Science Quarterly, 34*, 208-230.

Mlitz, K. (2021, August 4). *Public Cloud Application Services/Software as a Service (SaaS) End-User Spending Worldwide from 2015 to 2022*. Retrieved December 22, 2021 from Statista: https://www.statista.com/statistics/505243/worldwide-software-as-a-service-revenue/

Montgomery, C. A., & Wernerfelt, B. (1988). Diversification, Ricardian Rents, and Tobin's Q. *RAND Journal of Economics, 19*(4), 623-632.

Morgan, S. (2019, October 21). *Cybercrime Magazine*. Retrieved October 28, 2021 from Global Ransomware Damage Costs Predicted To Reach $20

Billion (USD) By 2021: https://cybersecurityventures.com/global-ransomware-damage-costs-predicted-to-reach-20-billion-usd-by-2021/

Naicker, V., & Mafaiti, M. (2019). The Establishment of Collaboration in Managing Information Security through Multisourcing. *Computers and Security, 80*, 224-237.

Ott, T. E., Eisenhardt, K. M., & Bingham, C. B. (2017). Strategy Formation in Entrepreneurial Settings: Past Insights and Future Directions. *Strategic Entrepreneurship Journal, 11*, 306-325.

Otto, B. (2015). Quality and Value of the Data Resource in Large Enterprises. *Information Systems Management, 32*(3), 234-251.

Peteraf, M. A. (1993). The Cornerstones of Competitive Advantage: A Resource-Based View. *Strategic Management Journal, 14*(3), 179-191.

Porter, M. E. (1980). Competitive Strategy: Techniques for Analyzing Industries and Competitors. *Free Press*.

Porter, M. E. (1985). Competitive Advantage: Creating and Sustaining Superior Performance. *Free Press*.

Ring, T. (2014). Threat Intelligence: Why People Don't Share. *Computer Fraud & Security, 2014*(3), 5-9.

Rose, S., Borchert, O., Mitchell, S., & Connelly, S. (2020, August). *Zero Trust Architecture*. Retrieved December 22, 2021 from National Institute of Standards and Technology: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-207.pdf

Ruohonen, J., Hyrynslami, S., & Leppänen, V. (2016). An Outlook on the Institutional Evolution of the European Union Cybersecurity Apparatus. *Government Information Quarterly, 33*(4), 746-756.

Sääksjärvi, M., Lassila, A., & Nordström, H. (2005). Evaluating the Software as a Service Business Model: From CPU Time-sharing to Online Innovation Sharing. *IADIS International Conference e-Society*, (pp. 177-186).

Safari, F., Safari, N., & Hasanzadeh, A. (2015). The Adoption of Software-as-a-Service (SaaS): Ranking the Determinants. *Journal of Enterprise Information Management, 28*(3), 400-422.

Sarker, I. H., Kayes, A., Badsha, S., Alqahtani, H., Watters, P., & Ng, A. (2020). Cybersecurity Data Science: An Overview from Machine Learning Perspective. *Journal of Big Data, 7*(1), 1-29.

Saunders, M., Lewis, P., & Thornhill, A. (2009). *Research Methods for Business Students* (Fifth edition ed.). London: Pearson Education Limited.

Shah, S., Horne, A., & Capellá, J. (2012). Good Data Won't Guarantee Good Decisions. *Harvard Business Review*, 1-4.

Shatrevich, V., & Gaile-Sarkane, E. (2015). A Strategic Fit Relation Model as a Tool for Organization Development. *19th World Multi-Conference on Systemics, Cybernetics and Informatics* (pp. 94-99). Orlando: Proceedings, Winter Garden: International Institute of Informatics and Systemics.

Sherman, J. R. (2011). Cloud-Based IT Audit Process. In B. Halpert, *Auditing Cloud Computing: Security and Privacy Guide* (p. Chapter 2). Hoboken, New Jersey: John Wiley & Sons, Inc.

Siggelkow, N. (2002). Evolution toward Fit. *Administrative Science Quarterly, 47*(1), 125-159.

Soltys, M. (2020). Cybersecurity in the AWS Cloud. *ArXiv*, 1-23.

Teece, D. J., Pisano, G., & Shuen, A. (1997). Dynamic Capabilities and Strategic Management. *Strategic Management Journal, 18*(7), 509-533.

Williamson, O. E. (1975). Markets and Hierarchies: Analysis and Antitrust Implications: A Study in the Economics of Internal Organization. *Free Press*.

Xin, M., & Levina, N. (2008). Software-as-a-Service Model: Elaborating Client-side Adoption Factors. *Proceedings of the 29th International Conference on Information Systems* (pp. 1-12). Paris: SSRN.

Xue, K., Chen, W., Li, W., Hong, J., & Hong, P. (2018). Combining Data Owner-Side and Cloud-Side Access Control for Encrypted Cloud Storage. *13*(8), 2062-2074.

Yin, R. K. (2017). *Case Study Research Design and Methods* (Sixth edition ed.). Las Angeles: SAGE: Thousand Oaks.

Yu, S. (2016). Understanding The Security Vendor Landscape Using the Cyber Defense Matrix. *RSA Conference*, (pp. 1-27). San Francisco.

Zuzul, T., & Tripsas, M. (2020). Start-up Inertia versus Flexibility: The Role of Founder Identity in a Nascent Industry. *Administrative Science Quarterly, 65*(2), 395-433.

Appendix 1

Preliminary question template for internal interviews

Profile

Q1: Could you describe yourself and your work? How is strategy, data and cybersecurity related to your work?

Section 1: Strategy process

Q2: What are the key drivers to enter a new market? How does internal capabilities and external market demand affect the market entry choices?

Q3: How would you describe your strategy or product development processes? How does the strategy or product development process differ in the cybersecurity industry compared to others or there at SSH?

Q4: What do the customers value in choosing cybersecurity vendors? What are the main sources of competitive advantage in the cybersecurity industry?

Section 2: Data ownership

Q5: What kind of data do you see valuable? How can data improve the decision-making or product development in the cybersecurity industry? What kind of data would help you to further develop your products?

Q6: What kind of differences there are between customer groups when it comes to ownership of (cybersecurity) data? How does customers value sharing the cybersecurity data with the vendor in order to further product development and better vulnerability protection?

Q7: What are the pros and cons in data sharing for third parties or a vendor?

Section 3: Cybersecurity

Q8: What kind of data do your products collect and how has this collected data been utilized in strategy formulation or product development?

Q9: What kind of cybersecurity data are the customers more/less willing to share? Why?

Q10: How the cloud and SaaS is/has been seen in the cybersecurity industry? Are there differences between customers, products etc?

Q11: What kind of changes have there been in data sharing in the past few years?

Section 4: Utilizing data

Q12: What kind of data ownership models do you recognize? What do you prefer and avoid?

Q13: How would you evaluate direct data ownership versus possibility to access the data on-demand?

Appendix 2

Preliminary question template for external interviews

Profile

Q1: Could you describe yourself and your work? How is strategy, data and cybersecurity related to your work?

Section 1: Strategy process

Q2: What do the customers value in choosing cybersecurity vendors? What could be the main sources of competitive advantage in the cybersecurity industry?

Q3: How would you describe cybersecurity vendors' strategy or product development processes from your point of view? How can you influence cybersecurity vendors' product development and strategic moves from your perspective?

Section 2: Data ownership

Q4: What kind of data do you see valuable in your business? How can data improve strategic decision-making? What kind of data would help you to further develop your products?

Q5: What are the pros and cons in data sharing for third parties or a vendor?

Q6: How would you elaborate the difference between owning the data by yourself versus having access to the data that a third party owns/hosts?

Section 3: Cybersecurity

Q7: How do you value sharing the cybersecurity data with the vendor to further product development and better vulnerability protection?

Q8: What kind of cybersecurity data are you more/less willing to share? Why?

Q9: How do you see cloud and SaaS in cybersecurity? What kind of changes have there been in data sharing in the past few years?