Using Patent Data to Position Firms in Technology-intensive Environments

How managers can use the world's largest repository of technological information to select technologies

Master's thesis in Entrepreneurship and Business Design

DAVID GENELÖV
LILI YUN

Department of Technology Management and Economics
Division of Entrepreneurship and Strategy CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018
Report No. E2018:021

This page intentionally left blank

MASTER'S THESIS E2018:021

Using Patent Data to Position Firms in Technology-intensive Environments

How managers can use the world's largest repository of technological information to select technologies

DAVID GENELÖV
LILI YUN

Department of Technology Management and Economics
Division of Entrepreneurship and Strategy
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2018

Using patent data to position firms in technology-intensive environments
How managers can use the world's largest repository of technological information to select technologies

DAVID GENELÖV
LILI YUN

Supervisor, Chalmers University of Technology: Bowman Heiden
Supervisor, Volvo Cars: Dawan Mustafa and Linnéa Claesson

Master's Thesis E2018:021
Department of Technology Management and Economics
Division of Entrepreneurship and Strategy
Chalmers University of Technology
SE-412 96 Gothenburg, Sweden
Telephone +46 31 772 1000

Typeset in LaTeX
Chalmers Reproservice
Gothenburg, Sweden 2018

Abstract

Many industries are currently being disrupted by new technology. Perhaps one of the clearest examples is the automotive industry, where autonomous driving, electric vehicles, and new business models disrupt the very core of what used to be a more linear development process (McKinsey, 2017). In the midst of this, incumbent automotive companies need to re-evaluate and update their technology base to stay relevant. Technology managers need to choose which technologies to include in their products and services, and such decisions are getting harder as the speed of change accelerates (Kurzweil, 2001). To exclude human biases from the equation, strategy consultants (McKinsey 2017; Roland Berger, 2014) believe that incumbents will have to use technology indicators and make deliberate decisions based on probabilities.

The purpose of this study was to create and test a framework for how technology managers can use patent data to reach insights for positioning their firms within different technological fields. To do this, the viability of patent data as an information source was assessed, and its drawbacks and benefits were evaluated. A literature review about previous patinformatics studies was performed, and the most important technology indicators were evaluated. A framework was built based on current best practices in patinformatics research, and the framework was tested to extract insights for the sensing technologies lidar, radar, and sonar.

One of the main conclusions of this study is that patents contain a vast amount of information. Patent data can be used to predict things ranging from how fast a technology is improving to how much effort is being invested in inventive activities and how mature the technology is in the technology life cycle. Another conclusion is that a well-structured patinformatics process is key to reach valuable insights. Patent sets that accurately represents the technological field need to be retrieved, the right metrics need to be computed over a relevant period of time, and the extracted insights need to be communicated in a clear and concise manner. This was tested on the three case study technologies, where a number of insights were gained. For example, sonar seems to be the furthest along in its technology life cycle, and companies are decreasing their inventive activities within this field.

The study was based on a foundation of three existing theoretical frameworks. Petrusson's (2015) Intellectual Asset management (IAM) framework was used as an overarching framework for how knowledge-based (and thus also technology-based) businesses should position themselves to maximize the utilization of their resources and capabilities. Gregory's (1995) process framework for technology management complements this by putting technology selection into context as one of the key issues in technology management. Finally, Moehrle et al.'s (2009) framework for patinformatics research process served as a base for designing the patinformatics efforts of the study.

Acknowledgements

This master thesis is a result of a master thesis project conducted at Volvo Cars during the spring of 2018 in Gothenburg, Sweden. We would like to show our gratitude for the experience and the opportunity to learn from our client Erik Hjerpe at Volvo Cars, who introduced us to the research problem. Another shout-out of gratitude goes to our company supervisors, Dawan Mustafa and Linnéa Claesson, who continuously supported us throughout the process, and to Ivar Hammarstedt, for all your insightful guidance through the jungle of competitive analysis. It has been a pleasure working with a future-oriented automotive company, and we're excited to follow Volvo Car's future journey.

The master thesis was written as a part of a larger project, and the one person that has helped most deserves not just a thank you but a standing ovation. Our deepest gratitude, therefore, goes to our friend and fellow student Hannes Forssberg Malm. Much of the data science results have also been reached in collaboration with our group member Elmar Aliyev. Thank you, Elmar for bearing with us as we've entered into the for us unknown territories of data science.

We would like to direct a special thank you to our faculty supervisor, Bowman Heiden, for awesomeness, guidance, and feedback during the thesis process. This shout-out should also be extended to the rest of the CIP team and everybody contributing to our master education. Throughout the past two years, we've been challenged, stressed, and somewhat traumatized. But most of all, we've learned a great deal about the world. We can only hope that we have given back a fraction of this through asking the right questions and together reimagining what we've been doing and why.

This thesis would not have been possible without the help gained from academics and industry professionals. In particular, we'd like to thank Marcus Malek from the AI-startup Aistemos, Leonidas Aristodemou from the Centre for Technology Management (CTM) at the University of Cambridge, and Christopher Magee and Christopher Benson from Massachusetts Institute of Technology. The support from you have been outstanding, and we can not help but wonder if kindness and helpfulness are salient characteristics of IP professionals.

Finally, we would like to show gratitude to our friends and families for the motivation and support during this spring.

David Genelöv & Lili Yun, Gothenburg, May 2018

Contents

List of Figures

List of Tables

1

Introduction

This thesis focuses on how patent data can be used to help position companies in technology-intensive environments. Following chapter consists of a project background, a problem definition, a purpose and the research questions the study aims to answer. Taken together, these subjects provide an exposition of the underlying need to research the topic. A brief overview of the delimitations that have been made and an outline of the thesis disposition is then provided to give the reader a better picture of the scope of the study and the structure of the report. Finally, the chapter is concluded with a reading guide.

1.1 Background

This section presents the current situation in the industry being studied and the need for better methods for companies in technology-intensive environments to position their technological offerings. It further aims to bring clarity into why patinformatics may be a valuable tool for that purpose.

1.1.1 Technological Change

When reading technology management literature, there seems to be an agreement that the rate of technological change in society is increasing and that companies therefore need better ways to quantitatively assess technology characteristics to keep their technological offerings relevant. This is no new phenomena, as suggested already by Basberg (1987) who claimed that "How to measure technological change has concerned economists, economic historians, historians of technology and research analysts for a long time. However, no widely accepted method has been developed so far". While many consultancy reports (e.g. McKinsey, 2017) claim that the speed of technological change is increasing, academics have reached different conclusions about both the definition and the correctness of this statement (Magee & Devezas, 2011). While a full investigation of the actual changing rate in society is outside the scope of the study, this subsection aims to highlight the issue of exponentially improving technologies and how they affect technology-intensive companies.

One way to reason about technological change is that complexity increases both when the rate of technological change increases and when the number of things that change increases (Modis, 2002). Consequently, by devising metrics to measure the progress of individual technological fields, we can come a step closer to assessing the rate of technological change in society. One change metric is the technology improvement rate (TIR), which denotes the yearly rate of progress for a technological field. In his 1965 paper "Cramming more components onto integrated circuits", Gordon Moore noted that the number of components per unit cost integrated circuit had doubled just about every year since the introduction of such circuits in 1958. He further predicted that the of change would remain constant for at least ten years. Moore revisited his predictions in 1975 and then established that the actual time between performance doublings was closer to 18-months, and that future rates could be expected to be around two years (Moore, 1975). Nevertheless, the prediction turned out to be surprisingly accurate and has given rise to the well-known phenomena called Moore's law.

In a narrow sense, Moore's law asserts that the number of transistors in a dense integrated circuit doubles about every two years. The term is, however, often used in a broader meaning, to denote the speed of change for any technology that shows an exponential growth pattern. Since Gordon Moore, many different researchers (e.g. Sahal 1979 and Nordhaus 2014) have found that technological performance in a field increases exponentially with time and that the percentage change per year is constant. If p is the performance at time t and p_0 is the performance at a starting time t_0, then

$$p = p_0\, exp(k(t - t_0)$$

In these terms, while the percentage rate of change k is constant, the exponential nature of the equation leads to growing absolute changes in performance levels. Thus, the rate of technological change for transistors is constant on a logarithmic scale and increasing on an absolute scale. Famous futurist Ray Kurzweil describes this as an evolutionary process, where "Evolution applies positive feedback" and "the more capable methods resulting from one stage of evolutionary progress are used to create the next stage" (Kurzweil, 2001). Kurzweil also means that the human brain has been developed throughout thousands of years of relatively slow development and that exponential improvement is counter-intuitive to how we view the world. As an example of this, he tells an old story about invention and mathematical thinking (here recited by Brynjolfsson & McAfee, 2011):

"In one version of the story, the inventor of the game of chess shows his creation to his country's ruler. The emperor is so delighted by the game that he allows the inventor to name his own reward. The clever man asks for a quantity of rice to be determined as follows: one grain of rice is placed on the first square of the chessboard, two grains on the second, four on the third, and so on, with each square receiving twice as many grains as the previous. The emperor agrees, thinking that this reward was too small. He eventually sees, however, that the constant doubling results in tremendously large numbers. The inventor winds up with 2^{64}-1 grains of rice, or a pile bigger than Mount Everest. In some versions of the story the emperor is so displeased at being outsmarted that he beheads the inventor."

Studies show that Moore's law can be applied to other fields than just integrated circuits (e.g. internet data traffic (Coffman & Odlyzko, 2002) and magnesium refinement (Nagy et al., 2013)). If we accept the proposition that humans find it difficult to grasp exponential change, it follows that technology managers in the 21st century will need to become more data-driven in their assessment of technological change and in technology selection decisions as we move into the second half of the chessboard.

1.1.2 Automotive Industry in a Time of Change

While the pace of technological change across many industries is accelerating, many industries are on the verge of disruption. This is extra prominent in the automotive industry where technological fields like autonomous driving and electrical vehicles disrupt the core of what used to be a linear development process. At the same time, high-tech software players are moving into the sector by means of major efforts and investments. According to McKinsey (2017), the number of patents filed annually in autonomous technology has almost doubled since 2012 and patents filed by new tech players have increased by an average of 25 percent each year.

Experts deem that there is a clear set of factors that will disrupt the traditional automotive industry, including innovation, electrification, software and data (Butler & Martin, 2016). Along with the accelerating pace of innovation, companies have to innovate faster than their competitors to not risk losing current positions and getting lapped in one technology cycle. In the automotive industry, the new high-tech actors seem to be taking a different approach to cars than more traditional actors, and with changing user preferences it is hard to say what technological solutions and business models will ultimately prove successful. Companies like Apple, Tesla and Google are expected new car entrants concentrating on their development and design around electric vehicles (Butler & Martin, 2016). For traditional car manufactures to be able to keep up with these companies and their move from combustion engines, transmissions emission, carburetor and exhaust systems and fuel economy management to batteries, charging systems, engine controllers and power optimization, the core expertise of the traditional industry leaders will have to be revamped.

Furthermore, tech companies and automakers are striving to deploy fully automated driving to the roads within the next couple of years (Steinbaeck, 2017), which may carry the potential to completely change the industry (Butler & Martin, 2016). As cars achieve initial self-driving thresholds, some supporters insist that fully autonomous driving is around the corner, while others argue that the technology tells a different story (Butler & Martin, 2016). Although many of the required technologies already exist, there is without a doubt still challenges to overcome before autonomous driving can be practical as the industry has not yet determined the optimum technology archetype for autonomous vehicles (Roland Berger, 2014). Nonetheless, while analysts still debate the current state of autonomous driving technology, pace of change, and the dynamics between new entrants and incumbents, many strategy consultant companies seem convinced that autonomous driving is going to happen, and that the questions is *when* it is going to happen rather than *if* (Butler & Martin, 2016).

In the deployment of electrical and autonomous cars, car components are increasingly being operated by embedded coding systems. The move from hardware to software might perhaps be the most significant shifting of the automotive tectonic change (Butler & Martin, 2016). New interfaces between the driver and car are to be expected, and a large number of driving decisions based on lots of sensory data will need to be made to enable autonomous capabilities. The many layers of software cause lots of consternation in the traditional automotive companies since it has not traditionally been their main area of expertise and, therefore, further complicates the automotive industry's ability to rapidly innovate in their usual way. Instead, new actors take the opportunity to fight for dominance in the auto space to quickly move into position. Given their huge cloud-based computing resources, it is not a surprise that Google has emerged as one of the leading actors in the development of autonomous vehicles.

An additional reason for why incumbent automotive companies are unlikely to excel in these areas is data. To reach full autonomous driving the system first have to learn how to act and adapt in a safe manner to many different circumstances, and given that there is yet no existing set of rules that can be programmed into a car to prepare it to avoid and anticipate all dangerous situations it may eventually encounter, effective autonomous driving systems must use new techniques such as machine learning and deep learning to develop sophisticated models entailing the intelligence needed. In doing this, large data sets are required. Again, although Tesla is heavily focused on this, Google takes the lead in reaching performance levels others cannot match in terms of using data at scale (Butler & Martin, 2016).

To succeed in the changing automotive landscape, original equipment manufacturers (OEMs) will have to focus strongly on developing and producing market-leading technologies (McKinsey, 2017). According to Ernst (2003), the allocation of research and development (R&D) resources to different types of technology is one of the most important decisions that technology managers make. But while these companies are hurrying to develop their own autonomous vehicles, their efforts may be too late. The latitude to make the right strategic decisions before anyone else does, and the opportunity to make exploratory decisions, becomes more and more limited, and as cost, complexity, and rate of technological change increases, the issues of managers being able to deliver value and competitive advantage through strategic technology management becomes more critical (Phaal et al, 2001). To stand the best chance of keeping up with a successful transformation that results in sustainable profitability, strategy consultants (McKinsey, 2017; Roland Berger, 2014) believe that incumbents will have to find a way to obtain strong navigators and learn how to to make deliberate decisions based on probabilities.

1.1.3 Sensor Fusion in Autonomous Vehicles

One of the most critical puzzle pieces of developing autonomous cars is replacing human senses with sensory functions using various technologies simultaneously. To enable full autonomy, one fundamental problem that needs to be solved is the visual perception problem of providing cars with the ability to sense their surroundings (Wolcott & Eustice, 2017). Companies operating in the autonomous driving space are thus continuously trying to perfect self-driving cars by achieving reliable levels of perception, mapping, and localization with the smallest number of test and validation miles needed. To do this, there are two approaches that appear in regular practice by the majority of autonomous vehicle players;

1. The first approach uses radars, sonars and camera systems to perceive vehicles and other objects in the environment. The environment is not assessed on a deeply granular level using this approach, but it requires less processing power.
2. The second approach uses lidar in addition to the traditional sensor suit of radar and camera systems. This approach is more robust in various of environments (especially in tight, traffic-heavy ones) but requires more data-processing and computational power.

Lidar, short for light detection and ranging, is the latest development in surveying technology, advancing from its predecessors; radar (radio detection and ranging) and sonar (sound navigation and ranging) sensors. The three technologies aim to fulfill similar purposes, but instead of using radio waves or sound to scan its environment, lidar utilizes laser light pulses. Because of its ability to map its surrounding at the speed of light, lidar systems can achieve a much finer scan accuracy than other existing technologies can allow (Schwarz, 2010). Thus, lidar is the preferred technology of Google's self-driving car company, Waymo. Scholars claim that Waymo's development in lidar has become so advanced that it is able to create an accurate three-dimensional image that is almost as good as human eyes (Häne et al, 2017). Radar may not be as smart, but it is reliable, affordable and has a longer "eyesight". Among the autonomous car development companies, Tesla is fairly unique in forgoing lidar entirely in its development. Tesla's CEO and founder Elon Musk does not believe that lidar is a critical component in autonomous driving systems, and Tesla instead uses a passive camera, sonar, and radar systems to provide a complete set of sensor capabilities (Tesla, 2018).

Many thought leaders seem to have reached the conclusion that lidar augmentation will become important for many autonomous car players. But due to the cost of lidar still being unreasonably high for the technology to be incorporated in cars (Wolcott & Eustice, 2017), it has motivated the development and investments in other solutions and the question of which technology that will become the industry standard remains open (Steinbaeck, 2017). The interesting technology dynamics and fast-changing market conditions for autonomous cars therefore make lidar, radar, and sonar sensors exceptionally interesting cases for applying novel methods to gain new insights and aid managers in their technology selection decisions.

1.1.4 Patinformatics for Technology Positioning

The term technology positioning is used in this study to denote the sum of all technology selections made by an organization, resulting in a distinct set of technologies used in the creation of goods and services. A successful technology positioning thus means that the organization's technology selection decisions have been well-aligned with its overarching objectives.

With perfect information and a complete understanding of different technology options and their trajectories, companies could consistently select optimal technologies. In the real world, however, perfect information is an illusion and companies have to be content with using various indicators and clues to aid their decisions. Archibugi & Pianta (1996) wrote that *"Within firms, detailed information about technological advance is needed to take the right decisions concerning the amount of resources to devote to innovation, to select the fields where innovation promises economic returns and to manage innovative strategies within companies"*. Similarly, Kassicieh and Rahal (2007) expressed their views as *"predictions of benefits from investment in a new technology is of great interest. Forecasting the success of future technology is key to decision makers. Because, knowing or predicting the success of invested technology provides important clues, such as the current technology life cycle of the technology under consideration, diffusion potential and technology scope. In technology and business, it provides planners to choose the right strategies for the future."*

Patent data is the world's largest repository of technological innovation (WIPO, 2015) and patinformatics is defined as *the science of analyzing patent information to discover relationships and trends* (Trippe, 2003). Since patents represent inventive activity, patinformatics can help companies to target their innovative efforts and decide how much of the R&D budget should be spent on different technological fields (Basberg, 1987; Ernst, 2003). Companies that fail to file patents and process patent information therefore risk losing market competitiveness (Trappey et al. 2012).

In recent years, the traditional use of patent information has evolved into a more strategic use of the information thanks to the development of more computerized databases (Jun & Lee, 2012). Today, analytical software products, electronic databases and private service providers with their own proprietary value-added technology or patent databases are available for assisting in the patinformatics process. This emerging technology reflects a change in emphasis from using computers and patent information to support record keeping toward using computers to facilitate decision-making. Policymakers, economists, and other professionals are increasingly making use of patent information to analyze for examples patent activities in a sector, technological field, or company, to forecast the direction of technical development and change, or to ascertain a company's relative technological position in a marketplace (WIPO, 2015). Hence, an expanded use of patent information can be seen across many different strategic and tactical businesses, research and policy-making activities at institutional, national and enterprise levels.

Despite this increased use of patents as an information source, there is a lack of an overarching understanding of how patents can be used to support technology selection decisions and technology positioning. Current research articles in the field mostly evaluate how well different patent indicators correlate with the phenomena they supposedly represent, but this makes research somewhat fragmented and creates a need for joining previous studies in a larger setting.

1.2 Problem Definition

The speed of technological change is increasing and companies have to make decisions about which technological fields to use in their goods and services. Specifically, in the context of autonomous driving, there is currently a technological development where different types of sensors; lidar, radar, and sonar, are used to detect objects and map out surroundings. Technology managers should use all information at their disposal to make optimal decisions regarding how to position their technological offerings, but due to fragmented research and the limited data processing capabilities of humans, patent data is currently underutilized as a source of information to guide decisions. There is thus a need for an overarching framework for how the world's largest repository of technological information can be used in technology positioning.

1.3 Purpose

The purpose of this study is to create and test a framework for how technology managers can use patent data to reach insights for positioning their firms within different technological fields.

1.4 Research Questions

Research questions have been formulated to progressively focus down the study from a general field of research to a researchable thesis subject (Bryman & Bell, 2011). They have also served as guidance for fulfilling the purpose and relieving the research problem. To this end, the main research question (MRQ) was first defined and then broken down into four sub research questions (SRQs).

1.4.1 Main Research Question

Due to fragmented research and low degree of industry integration, patent information is currently underutilized as an information source. Simultaneously, technology managers need to continuously position their organizations in fast-changing industries. Thus, there is a need for an overarching framework that merges insights from previous research. The main research question is:

MRQ: How can a patinformatics framework be constructed to help technology managers gain valuable insights for selecting which technologies to invest in?

1.4.2 Sub Questions

At the core of patinformatics lies patents, and one can not build a good patinformatics framework without understanding patents characteristics and the drawbacks and benefits of using patents as an information source. The first sub research question is thus:

SRQ1: What information can be found in patent data?

Previous research includes a plethora of different metrics that have been used and tested as technology indicators. To build a comprehensive framework, there is a need to understand what those metrics are and how they have been used. The second sub research question is therefore:

SRQ2: What patent data metrics can be used to reach insights about technologies?

Having understood the individual metrics and their properties, a selection has to be made regarding which are most relevant for the problem of technology positioning. This leads to the third sub research question:

SRQ 3: How can a patinformatics framework be constructed for use in technology selection?

Finally, the constructed framework should be applied to real cases and evaluated. The technological fields of sonar, radar, and lidar are at the epicenter of the disruption of the automotive industry and no dominant design has yet emerged. The fourth and final research question thus reads:

SRQ4: What insights can the constructed framework give about the technological fields of lidar, radar, and sonar?

1.5 Delimitations

Due to the beforehand stipulated resource constraints, the study will be delimited by a number of factors. First, the project will be studied through an Intellectual Property (IP) perspective studying only patent data, meaning that financial, regulatory and market considerations will be left outside of the research scope. Secondly, the study is limited to patinformatics for technology selection. Hence, it will not deal with analyzing competitive positions for competitive intelligence or mergers and acquisitions (M&A:s), or comparing innovative activity between actors, countries or other entities. While new patents and innovations might affect how technologies should position themselves, the framework created here is to be used at discrete points in time when evaluating different technology options, and patent monitoring will thus be left outside the scope of the study.

Moreover, technology selection is a complex issue involving a number of factors, such as assessing the economic and technical feasibility of different options, evaluating the company's current research position in the different fields and that of its competitors, and determining how well the different options align with the company's overarching strategy and business objectives. Naturally, not all information for these decisions can be found in patent data, and this study is not an attempt to build an all-encompassing framework for technology selection. The focus here is instead exclusively on patinformatics and what technology insights can be gained by analyzing patents.

1.6 Thesis Outline

This study is divided into seven chapters according to the following disposition:

Chapter one serves as an *Introduction* to the study. It includes an account for the background leading up to the problem being studied, a problem definition, the purpose of the study, the research questions to be answered and the delimitations of the study.

The second chapter, *Theoretical Foundation*, entails the fundamental frameworks from previous research that this study has built upon. The three most prominent frameworks are then merged to construct the theoretical framework that has been applied in this study.

The third chapter comprises the *Methodology* of the study. This includes an exposition of the research strategy, research design, and research method applied, as well as an outline of what measurements have been taken to assure a high quality of the research.

In chapter four, the Framework Construction is presented. This includes a background to patents, the patent system, what information that can be found in patent data, and the advantages and disadvantages of using patent information. A patinformatics framework for technology selection insights is then created.

In chapter five, *Findings and Analysis*, the previously created framework is applied to the technological fields of lidar, radar, and sonar. First, an introduction to the technological fields is presented. Further, the results from the analysis of the three technological fields are visualized, and possible business implications and insights are extracted.

Chapter six, *Conclusion*, presents the key findings of the study, including the answers to the research questions.

Finally, chapter seven, *Discussion*, addresses the theoretical and practical implications of the study, its limitations, and a call for further research in the field.

1.7 Reading Guide

This study addresses a well-defined problem in technology management and its main target audience is, therefore, technology managers that want to make more well-informed and data-driven technology selection decisions. For such readers, chapters four, five and six are recommended.

Apart from technology managers, actors within academia, and especially within the fields of intellectual property and data science may find the study interesting. Such readers are advised to start with chapter one and then review the thesis outline above to choose areas of further interest.

2

Theoretical Foundation

This chapter introduces a selection of the existing technology positioning and patinformatics frameworks. A new framework for use in this study is then created by combining the frameworks.

2.1 Existing Frameworks

This section introduces two technology management and one patinformatics framework that has been used as a foundation for creating the new framework used in the study. The choices of frameworks build on the underlying belief that technology can be understood as a special kind of applied knowledge. This belief is reinforced by Floyd's (1997) definition of technology as *"the practical application of scientific or engineering knowledge to the conception, development or application of products or offerings, processes or operations"*. By defining technology in this way, tools from knowledge management can be applied to effectively manage it. In line with this reasoning, the first framework introduced is Petrusson's (2015) Intellectual Asset Management (IAM) framework, covering a structured procedure for claiming, positioning and utilizing intellectual phenomena. Secondly, Gregory's (1995) process framework for technology management and its usefulness as a well-defined process for companies developing or reevaluating their technology base is presented. Finally, Moehrle et al.'s (2010) framework for how patinformatics research can be performed in a structured way is explained.

2.1.1 Intellectual Asset Management

The IAM framework used in this study emerged as a result of the *Knowledge Management Platform* (KMP) programme conducted jointly by the Swedish government agency for state funding of research and development (VINNOVA) and the University of Gothenburg. Building on previous models created by the Center for Intellectual Property (CIP) in Gothenburg, the framework was originally built to aid people in academic environments to increase the utilization of their research results (Petrusson, 2015). The framework has since been adapted to also suit the need of technology-based companies and has been successfully deployed as a technology management tool in companies like SKF (Konsert Strategy & IP, 2018). There is, however, still a need to distinguish between IAM for technology-based businesses and IAM for academic environments, and where necessary, such distinctions have been made in the following paragraphs.

The foundation for the IAM framework is the degunkification of the otherwise elusive term "know-how" into Intellectual Assets. By applying a structured process for how different kinds of know-how can be created and managed, organizations can achieve both a better alignment of their internal value creation efforts with their external environment and a larger degree of utilization of those results. As shown in Figure 2.1 below, the framework is built around four key processes that characterize value creation processes in knowledge-based organizations, namely 1) Claiming 2) Positioning 3) Deciding and 4) Organizing. Furthermore, the framework offers a support system including procedures, sources of information and process support tools for each key process (Petrusson, 2015).

Claiming Support system for identifying, analyzing and claiming knowledge assets and intellectual property assets	**Deciding** Support system for decision-making in utilization and collaboration contexts
Positioning Support system for positioning organizations' technology efforts	**Organizing** Support system for organizing intellectual assets in organizations

Figure 2.1: Intellectual Asset Management framework (*adapted from Petrusson, 2015*)

Key process 1), claiming, concerns how organizations should identify, analyze and eventually claim intellectual assets that exist or are created within their organizational contexts. The support system for this process includes a set of procedures for mapping out intellectual assets, analyzing their characteristics and value, and claiming that value through the use of different control mechanisms. The support system for key process 2), Positioning, advocates a more proactive approach by focusing on what is happening in the outside world to develop and utilize the organization's relative position. By identifying relevant knowledge fields and creating knowledge trees, the organization can visualize the strengths of weaknesses of itself and others, which in turn can lead to a roadmap for future developments and collaborations. Key process 3), Deciding, covers the decision-making process concerning how to best utilize the organization's intellectual assets for value creation. The support system for this process includes a set of tools for how organizations can license their intellectual assets, utilize collaboration potentials for research diffusion, and templates for managing contractual relations. Finally, key process 4), Organizing, concerns the administrative governance of intellectual assets (Petrusson, 2015).

Since this study mainly focuses on key process 2), Positioning, it is worthwhile to identify some key concepts for this stage. The IAM framework should not be seen as a linear process with well-defined starting and ending points. In most cases, a proactive approach starting with considerations about what is happening in the external world is more relevant than starting by claiming knowledge assets (Petrusson, 2015). Thus, understanding the world outside the organization and aligning technology initiatives accordingly is of utmost importance for technology-intensive companies. From this reasoning, it also seems fair to deduce that a continuous reevaluation of the current positioning strategy becomes of great importance in fast-changing industries.

With the framework being adapted from academia, it is also worthwhile to reflect on the difference between the positioning process in the context for which Petrusson originally developed the framework and a corresponding process in the technology-intensive companies that are the current subject of study. For one, the IAM tool for academic environments was developed to assist in societal value creation (Petrusson, 2015), whereas companies generally use commercialization to maximize the appropriation of the value they create. Furthermore, academic research is mainly focused on the creation of new knowledge and making academic contributions, whereas a technology-intensive company might consider it sufficient to acquire off-the-shelf technologies in certain technological fields, focusing their efforts on acquiring a competitive advantage through others. The positioning process therefore becomes more a process of deciding which areas the company wants to lead in, which areas to follow in, and which possible technologies to acquire or develop for each area.

2.1.2 Process Framework for Technology Management

While the IAM framework can be seen as a research policy or technology transfer-based view of technology management, Gregory's (1995) process framework for technology management instead approaches the subject from an operations background. Gregory reacted to the narrow scope and low degree of industry integration of existing frameworks and aimed to create a new framework that could help technology managers understand the status and implications of technology across their business. To make the framework easily assimilated, it centers around natural labels and a logical language for characterizing the key processes and issues that occur in managers' everyday working situations. According to the framework (see Figure 2.2), the five main processes in technology management is; 1) Identification, 2) Selection, 3 Acquisition, 4) Exploitation and 5) Protection.

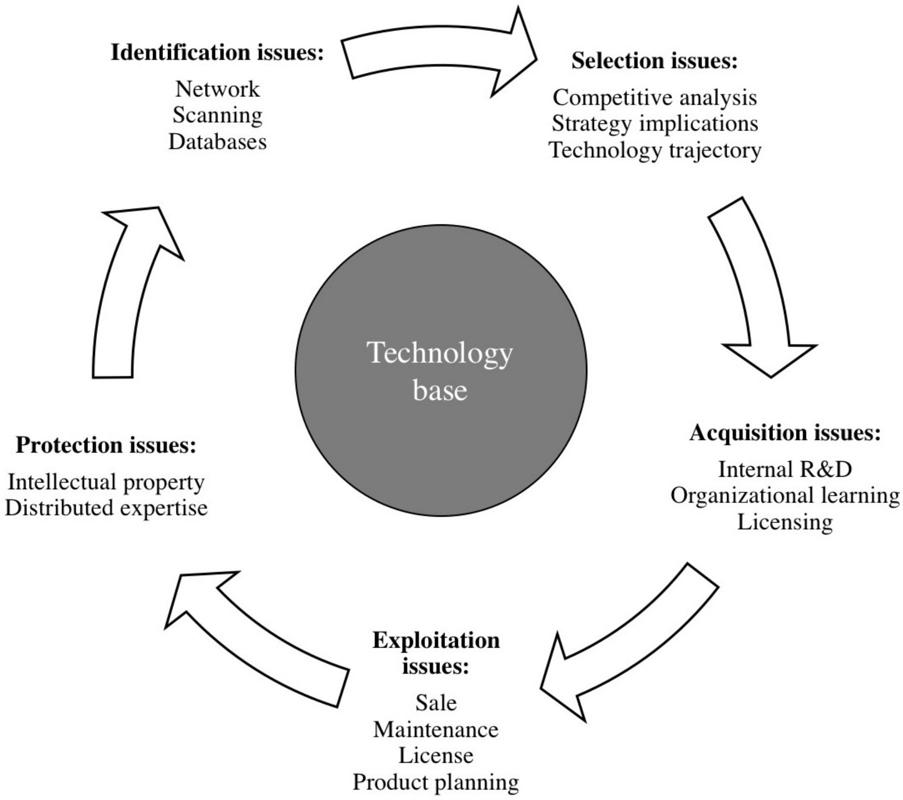

Figure 2.2: Process framework for technology management (*adapted from Gregory, 1995*)

Key process 1), Identification, concerns developing an awareness of the technologies that are important to the business. This may include the systematic scanning of which technologies exist or are emerging from outside the organization, as well as acknowledging the business's own technology development efforts. Once the organization understands which technologies exist in their context, they need to select the optimal ones, according to how well the technologies' characteristics align with the technology strategy and the overarching business strategy. Key process 2), Selection, is the process closest to the focus of this study, and the main goal of this process is to assess the relative importance of already identified technologies to the business and select how it should position itself in the technological landscape (Gregory, 1995). With this assessment done, the business can go ahead to key process 3), Acquisition, to purchase, license or otherwise obtain the right technologies. For this process, each organization should create a framework that explicitly outlines the processes needed to choose what acquisition routine to perform and how to execute the chosen approach. Having acquired the technologies, key process 4), Exploitation, concerns their systematic conversion into marketable products, services, or value extraction through other commercialization efforts such as technology licensing or joint ventures. Finally, key process 5), Protection, involves how to best protect the performed technology investment and includes the creation of processes for e.g. licensing contracts and patenting (Gregory, 1995).

2.1.3 Patinformatics Research Process

The last framework introduced in this section has been chosen for its clarity into how a patinformatics research process can be carried out. Moehrle et al.'s (2009) framework (see Figure 2.3) divides the process into three distinct "core processes"; 1) Patent pre-processing for easy access, 2) Data analysis to reach business insights and 3), Discovered knowledge (the utilization of analysis results for business decision-making).

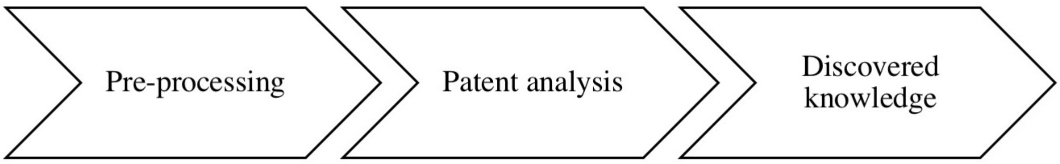

Figure 2.3: Patinformatics research process (*adapted from Moehrle et al. (2009)*

The core process of pre-processing focuses on preparing patent documents for other tasks. From the time when the patent is registered at the patent office, it must be converted to digital format and stored in a database. Once this is done, further value might be gained from generating metadata about documents, e.g. extracting keywords and summarizing content. The analyst may then choose to limit the number of documents by certain criteria and finally extract the resulting data from the database. Today, patents in a digital form with a limited selection of metadata can be easily obtained from publicly available databases (Altuntas, Dereli & Kusiak, 2015). The pre-processing task for the modern patinformatics researcher is, therefore, mainly focused on limiting the range of search through retrieving patent sets that are representative of the company, industry, technology or country being studied. This can be done by managing the trade-off between precision and recall, which means to include a large enough portion of the desired documents while excluding the undesired documents.

When limiting the range of search, the choice might be between using single patents, groups or classes of patents, or using data for total patenting activity (Basberg, 1987). Furthermore, the researcher must choose whether to study only granted patents, pending applications, or a combination. While granted patents might have higher average quality due to the screening performed by patent authorities, they also suffer from the additional time lag of the patent prosecution procedure, making them less well suited for studying recent developments and fast-evolving technologies. Document selection may also be based on regular patent search methods like binary word searches in patent titles (Altuntas, Dereli & Kusiak, 2015) or using the four-digit-codes of the patents' IPC classification (Gao et al., 2013).

The core process of pre-processing is shown in Figure 2.4 below.

Figure 2.4: The core process of pre-processing (*adapted from Moehrle et al. 2009*)

Once the relevant patent data is available in a digital database, the next core process is the actual patent analysis. An overview of this process can be seen in Figure 2.5 below.

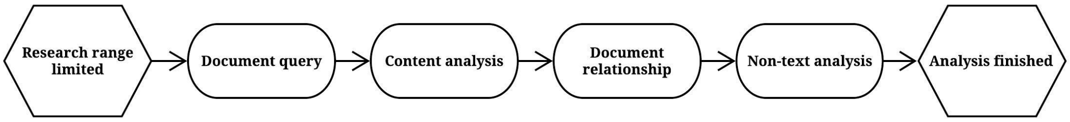

Figure 2.5: The core process of patent analysis (*adapted from Moehrle et al. 2009*)

The patent analysis starts with querying and retrieving the previously generated patent sets. With a relevant set of patent documents, the information in the patents may be analyzed. This can either be done by studying the content on a per patent basis (content analysis and non-text analysis) or by studying the relationships between documents (document relationship analysis). Aggregated metrics, such as the average number of citations per patent family or the number of new applications per year can be derived for different points in time, to show both the historical situation, the current situation and the current trajectory.

Having performed the patent analysis, the final step of a patinformatics research process is to evaluate and refine the results in order to support operational patent decisions or strategic decision-making (Moehrle et al. 2009). This process is seen in Figure 2.6 below.

Figure 2.6: The process of discovered knowledge (*adapted from Moehrle et al. 2009*)

Visualization of analysis results provides a quick overview of the results and helps managers and analysts to reach insights from a single impression. Moehrle et al. propose that the two methods of patent mapping and network-diagrams are well-suited for this purpose, but there is a plethora of other possible methods available. Having visualized the results in a suitable way, the implications of the research results can be evaluated and business insights can be extracted. Finally, the analysis results should be documented to allow for later revisitation or modification of the analysis.

2.2 Constructed Framework

The three presented frameworks together form the theoretical foundation upon which this study was based (see Figure 2.7). The IAM framework presents an overarching framework for how knowledge-based (and thus also technology-based) businesses should position themselves to maximize the utilization of their resources and capabilities. The framework also introduces patents and patent data as an important source of information. The process framework for technology management complements this by putting technology selection into context as one of the key issues in technology management, and further proposes important considerations to be made regarding how selection should be performed, including a competitive analysis, predicting technology trajectories, and understanding the strategic implications of the decisions on the overall business.

An understanding of how technology-based businesses should think about positioning and technology selection was crucial to design a patinformatics research process that could provide managers with valuable insights. For the more technical aspects of designing the patinformatics research process, Moehrle et al.'s framework was employed and the process was split into pre-processing, patent analysis and discovered knowledge.

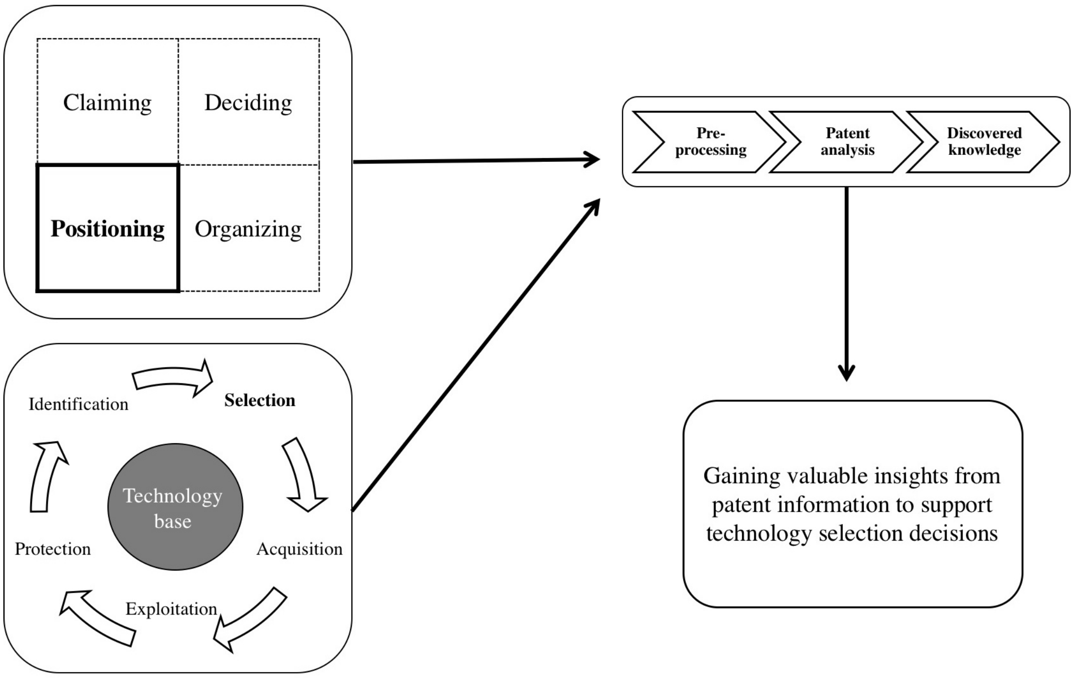

Figure 2.7: The theoretical framework constructed for this study

3
Methodology

This chapter outlines the underlying methodology that has been applied in the study, including the research strategy, research design, research process and a discussion about the quality of the research.

3.1 Research Strategy

According to Bryman and Bell (2011), the research strategy is dictated by the nature of the research and is further guided by different considerations made by the researchers. The following subsections aim to explain the most important considerations made for this study, including how theory has been linked to research and how epistemological and ontological issues have been handled. With these considerations as a fundament, the chosen research strategy is explained.

3.1.1 Linking Theory and Research

To recap, the purpose of this study was to build and test a framework for using patinformatics for technology selection insights.

Raw patent data concerns individual patent documents, and to study technological fields as a whole, metrics need to be aggregated for each technological field. The researcher thus has to go through the inherently deductive process of choosing what features to include and how to represent them. At the same time, a number of previous studies have investigated the efficiency of different aggregated patent metrics as technology indicators, and setting up hypotheses about all these metrics and testing them again was deemed unfeasible. In this study, a subset of existing metrics was used to gain insights about the field of study, and the existing theories about the connection between the metrics and the phenomena they supposedly represent were accepted on an “as-is” basis in the creation of the framework.

Since this study concerned both the creation and the testing of a framework, it is useful to consider how significant each of those two parts was. The theory creation part was inherently inductive, while the subsequent testing of theory was more deductive. Since the framework was tested on fields without an existing "ground truth", no real insights about the deductive power of the model could be reached. The framework creation was therefore considered to be of greater importance than the testing. The primary mode of reasoning employed in this study can thus be described as inductive, with the ultimate objective of merging previously fragmented research together into a coherent study of what insights patinformatics can bring to the technology selection process in organizations.

3.1.2 Epistemological and Ontological Considerations

Decisions about what should be regarded as acceptable knowledge (epistemology) and the nature of existence (ontology) have impacted both how the research questions were formulated and what the most logical way of answering them was. The epistemological and ontological considerations have been based in the mode of reasoning described in *Linking theory and research* in line with academic best practices (Bryman & Bell, 2011).

Since patents are the result of human creation, the phenomena being studied can be considered ontologically subjective. Thus, it would seem logical for a study like this to employ a constructivist position, acknowledging that social order is in a constant state of change and that people continuously create and recreate their surrounding social reality. There is no law of nature that generates patents and patent data, but rather people and their everyday decisions. Nonetheless, valuable insights can be gained by applying scientific methods to find patterns in human behavior. For instance, advanced statistical methods such as artificial neural networks can be used to predict how likely a customer is to quit his or her wireless carrier contract (to churn) given characteristics of that person's cell phone usage (Pendharkar, 2009). Thus, a considerable value can be extracted through applying scientific methods to the study of human behavior, even while acknowledging the fact that there is no one true social reality. To balance the constructivist nature of the phenomena being studied with the apparent value creation possibilities for positivistic research methods, this study uses a pragmatic research paradigm. This allows for a mixed-method research design where a constructivist worldview can be employed in a scientific study to answer questions in a way that can be practically useful (Feilzer, 2009).

3.1.3 Quantitative and Qualitative Research Strategies

While the distinction between quantitative and qualitative research may be somewhat ambiguous, it is continuously used as an umbrella term to denote two different strategies concerning a range of issues within the practice of business research (Bryman & Bell, 2011). The terms build on separate epistemological and ontological orientations, as well as different views on how theory should be linked to research. While qualitative research is typically associated with inductive reasoning, constructivism, and interpretivism, quantitative research is more commonly connected with deductive reasoning, positivism, and objectivism (Bryman & Bell, 2011).
Building on the stance taken in *Linking theory and research* and *epistemological and ontological considerations*, and in particular, on the explained need for positivist methods to study human behavior, this study employs a mixed-methods research strategy where quantitative methods are used to generate patent data metrics that

can be evaluated qualitatively.

3.2 Research Design

This study adopted a comparative research design where identical methods were applied to examine multiple cases (Bryman & Bell, 2011). The reasoning for using a multiple-case study design was based on the logic that a multiple-case study helps improve theory building, and by comparing two or more meaningful cases, researchers can generate theoretical insights uncovered through the comparison (Yin 1984; Eisenhardt 1989). In this case, the comparative design was used to study developmental trends of patent data metrics as technology indicators across different technological fields. However, rather than building entirely new metrics for how to evaluate technologies, the goal with this study was to concentrate on using existing metrics in a new framework to give a comprehensive image of the development of different technological fields. According to Bryman & Bell (2011), this approach is relatively popular for business and management research, but a general objection towards studying multiple cases is the risk of losing attention to the specific context and therefore the study poses a high risk of being generalized. To mitigate this risk the study concentrates on allowing the distinguishing characteristics of the cases to act as a springboard for theoretical reflections about the findings (Bryman & Bell, 2011).

3.2.1 Research Methods

According to Bryman & Bell (2011), research methods constitute different techniques and instruments for collecting data. This section presents the data that was considered necessary for conducting the study in line with the constructed theoretical framework. Furthermore, it also presents the research process and the data collection techniques used in the data collection process.

3.2.2 Required Data

To answer the research questions, different types of data was required. Since the study was centered around patent data, it was deemed necessary to get a solid grasp of patents and the patent system. To exhaustively answer what information can be found in patent data, both the information fields on the patent documents and information aggregated by other actors needed to be considered and an evaluation of drawbacks and benefits helped shine the light on the limitations of patents as an information source.

The second research question focuses on which different patent data metrics can be constructed to provide insights about technologies. Two separate data collection efforts thus needed to be undertaken. First, an overview of different possible patent data metrics needed to be built, and then each metric needed to be evaluated based on how well it functioned as a technology indicator and thus provided insights.

The third research question concerns how to build a framework for using patinformatics for technology selection insights. In line with the patinformatics research process framework, this question can be seen as an optimization problem where data is needed on three key issues; how to best pre-process the data to create a good representation of the technology being studied, what metrics to use to best analyze the chosen data, and how to best communicate the discovered knowledge to help decision-makers reach insights and ultimately perform better technology selection decisions.

The fourth and final research question concerns the testing of the constructed model on the technological fields of sonar, radar, and lidar. To be able to evaluate the analysis results and make sense of the information being presented, a basic understanding of the technological fields and their historical development was deemed necessary. Furthermore, patent data sets corresponding to the different fields needed to be constructed, the analysis needed to be performed, and the results needed to be evaluated. The different research questions and a summary of the data needed to answer each question can be seen in Figure 3.1 below.

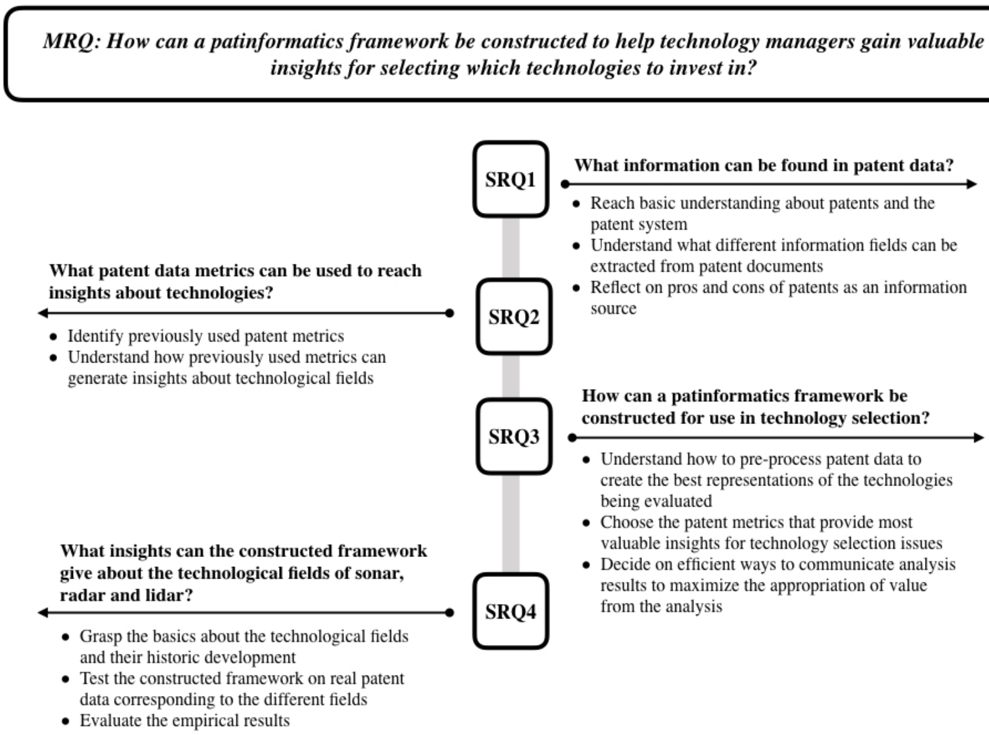

Figure 3.1: Data required to answer the research questions

3.2.3 Research Process

Formulating the research process includes defining the different stages of the research and their timing in the overarching process (Bryman & Bell, 2011). This study was completed through three separate phases, as shown in Figure 3.2 below.

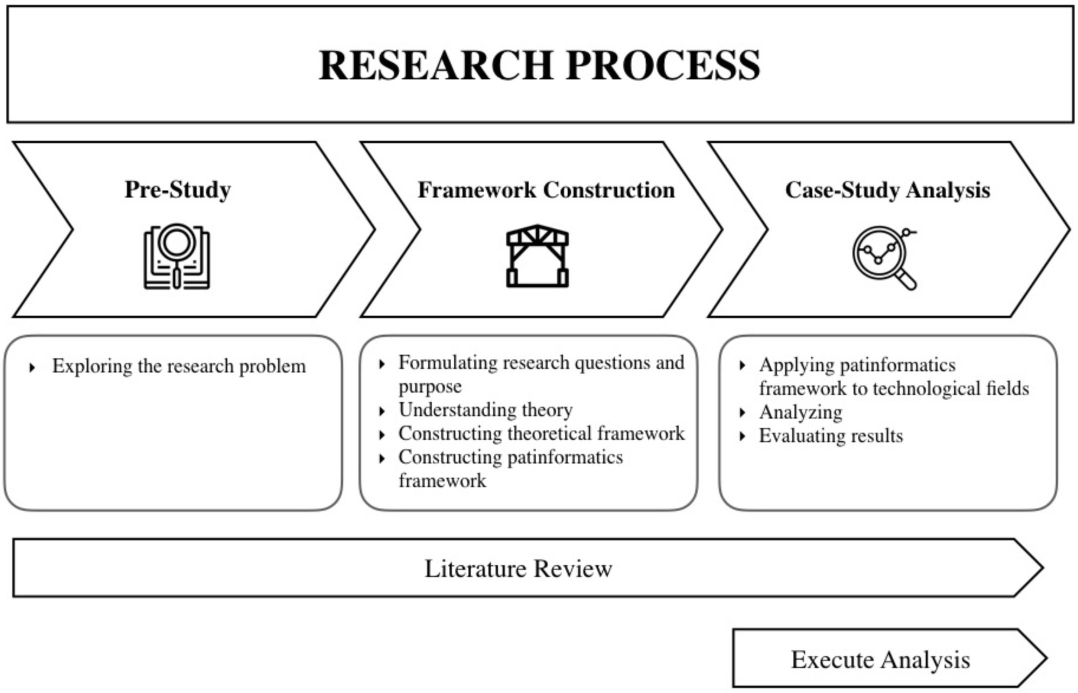

Figure 3.2: Research process

A literature review was conducted iteratively during the entire study to facilitate for new theories to be discovered and formulated. Reviewing literature also gave a more in-depth understanding of the different concepts, so that the right theories to use could be selected.

Given a research topic from a commercial automotive actor, the initial phase included an extensive background study to create a problem definition and to formulate a research problem of scientific relevance. By reviewing prior research, the research problem could be formulated to ensure that the research topic had not yet been addressed in literature, that it had an academic contribution as well as to aid an industry actor to build a solid foundation for their new patinformatics initiative.

The second stage started with the formulation of purpose and research questions based on the research problem. By scanning theories and models relevant to answering the research questions a framework with the purpose of providing scientific adjudication and guidance to the study could then be constructed. When constructing the framework, valuable insights and understanding of the research field were gained. To explore the research field even further, a chapter summing up the rest of the literature study and the basic understandings of different key concepts was created. During this step, patent metrics likely to be related to technology selection insights were identified. Next, an evaluation of which metrics to test was performed. In doing so the feasibility and implication of using each metric in the technology selection process were assessed. When an understanding of all relevant key concepts had been obtained, the analysis part of the study could be executed.

In the third and final stage, the constructed patinformatics framework was applied to three case subjects (lidar, radar and sonar) that the patent metrics were tested on. The initial step of this process was to extract patents representing the three different technological fields. By interpreting all the findings from the previous steps and from examining the different technological fields using the patent metrics, different technology selection insights could be identified. These insights were then analyzed to arrive at a final conclusion that satisfied the research questions and overall purpose of the study.

3.2.4 Data Collection

The research method used in this study was two-pronged with an equal emphasis on a literature review and a retrieval and analysis of patent data.

3.2.4.1 Literature Review

A comprehensive literature review using a variety of sources was conducted to create a theoretical foundation while providing a deeper understanding of the different concepts included in the study. Using Chalmers University Library and Google Scholar as the main search engines, relevant publications were collected, whereof articles from academic journals held for the majority of sources. To ensure high validity on these, the publication year and the number of citations were considered.

3.2.4.2 Case Study

Yin (1984) defines the case study research method as *"an empirical inquiry that investigates a contemporary phenomenon within its real-life context; when the boundaries between phenomenon and context are not clearly evident; and in which multiple sources of evidence are used"*. As already mentioned, literature clearly emphasizes the risk of generalization when performing case studies (Bryman & Bell, 2011). Therefore, rather than creating a generalized theory from the studied case subjects, the case study concentrated on identifying the differences between the cases to then take all the pieces and bits of seemingly unrelated information from each case together, to synthesize into a new understanding and insights.

The goal of conducting this case study was to understand what insights the previously constructed patinformatics framework could give for the task of technology selection. In deciding which technologies to proceed with in the case study, three technological fields appeared to be exceptionally interesting cases for applying novel methods to gain new insights; lidar, radar, and sonar. First, basic knowledge about the technological fields and their historical development needed to be understood.

This data was collected using company websites, articles, and reports. To test the constructed framework on real patent data corresponding to the different fields, patent data sets corresponding to the fields needed to be constructed. The analysis then needed to be performed, and finally the results needed to be evaluated.

3.2.4.3 Patent Retrieval and Analysis

The data collection process for the case study involved a) retrieving relevant sets of patent documents that represented each of the three technological fields being studied and b) analyzing the retrieved documents to create the aggregated patent metrics to be used in the analysis. The process of data collection for the case studies is visualized in Figure 3.3 below.

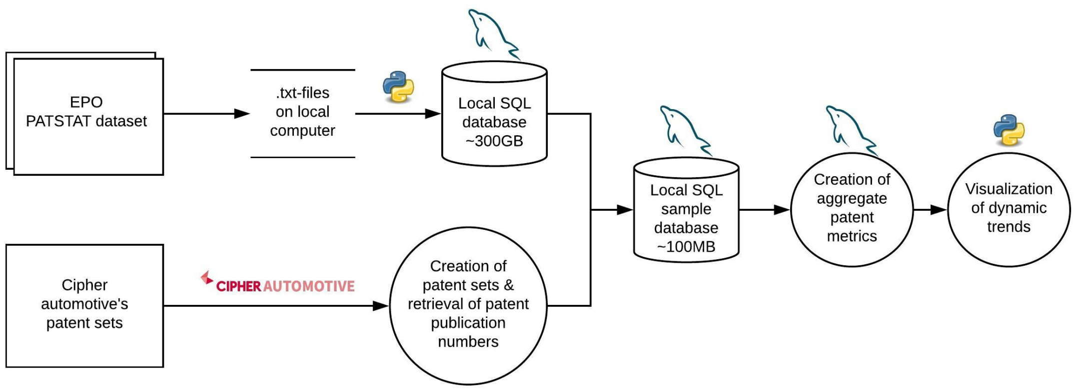

Figure 3.3: Data collection process for case studies

First, the IP Business Intelligence tool Cipher and the Cipher Automotive platform (red logo) was used to generate representative patent sets. Since Aistemos, the company behind the platform, has already performed an extensive work in categorizing patents into technologies together with automotive actors, the patent sets generated this way were both more relevant and more complete than what the researchers would have been able to generate with their limited industry understanding. This process resulted in a list of patent numbers that represented each technology, and some additional data about each document that Cipher had already gathered.

Concurrently with the generation of the patent sets, EPO's PATSTAT[1] patent datasets with more than 100 million unique patent applications was purchased. The rationale for this was twofold. First, the PATSTAT dataset gave access to a number of data features not present in the Cipher dataset and thus allowed for creating a larger amount of different patent metrics. Secondly, the PATSTAT dataset allowed for a larger study to be performed outside of the one currently being explained. Once purchased, the data was downloaded as .txt-files to the local computers. The programming language Python (yellow and blue logo) and its interactive development

[1]More information and sample data of the PATSTAT dataset can be found through the following link: https://www.epo.org/searching-for-patents/business/patstat.html#tab-1

environment Spyder was then used to upload the patent data into a local MySQL database instance. MySQL (blue logo) is an open-source relational database management system and in this case, the unified visual database tool MySQL Workbench was used for easier construction of queries and evaluation of results.

With the limited processing power of the computers used and the large size of the full PATSTAT database in SQL (300GB), regular data analysis tasks and methods such as joining or querying tables became painfully slow. All information concerning the three case study patent sets were consequently extracted into a second database (the "sample database") for efficient management. This database was then used to analyze the patent metrics under research for different points in time for each of the three technologies.

Finally, the numerical values for all different patent metrics were imported into Python and their dynamic trends were visualized using the open source 2D-plotting Python library Matplotlib.

3.3 Quality of Research

To assess the quality of the study, Bryman and Bell's (2011) framework for evaluating quantitative research has been employed. In line with this framework, this section concerns the reliability, replicability, and validity of the study.

3.3.1 Reliability

Reliability refers to the consistency of a measure or a concept, and in particular whether the measures used would remain stable when replicated. To put it simply, the measurements of a reliable study can be expected to show a high correlation with the measurements of a replication of the study. According to Bryman and Bell (2011), three factors should be considered when assessing whether or not a measure is reliable.

Stability refers to whether or not a measure is stable over time, and thus whether or not the researchers can trust that the results for a certain sample do not fluctuate. When doing social research, and in particular when performing interviews, this can prove difficult since social settings can be hard to replicate and people tend to change their opinions over time. *Internal reliability* concerns whether or not the indicators that make up an aggregated measure are consistent. If researchers want to use several simple metrics as attributes to describe a larger feature, the metrics should be expected to correlate with each other. *Inter-observer consistency* is important when different people use their subjective judgment, e.g. when grouping data into categories or recording observations. In such settings, measures should be taken to assure that all observers use a consistent methodology.

Since the data used in this study may be considered ontologically subjective, steps have been taken to assure the reliability of measurements. First, clear guidelines have been created for how to extract data. A reliable online patent analytics platforms have been used and an interview was conducted with the head of strategy of the company behind the platform to assure that it was used in a correct way. Since all patent documents are marked with a timestamp that indicates when the documents were made available to the public, future researchers should be able to limit their searches to only obtain patent documents that were available at the time of this analysis. The use of patent data thus makes for a fairly easy extraction of data that gives stable and reliable results.

3.3.2 Replicability

The replicability of a study concerns its capacity to be replicated by other researchers. To be replicable, the procedures used in a study must be spelled out in great detail. By making their study replicable, researchers enable others to assess the reliability of a measure of a concept (Bryman & Bell, 2011).

Measures have been taken to ensure the replicability of this study. An effort has been put into explaining data extraction procedures and the procedures that make up the different measurements that have been used.

3.3.3 Validity

Validity is a multi-faceted criterion that is concerned with the integrity of the conclusions generated from business research. According to Bryman and Bell (2011), the main types of validity generally distinguished are *measurement validity, internal validity, external validity and ecological validity.*

Measurement validity applies primarily to quantitative research and evaluates whether or not a measure that is devised of a concept really reflects the concepts it is supposed to denote. For instance, if IQ tests really measure intelligence. Internal validity is concerned with causal connections and if researchers can be certain that the direction of causality is as claimed. External validity concerns whether or not the results of a study can be generalized beyond the specific research context, and highlights the importance of choosing samples that are representative of larger populations. Finally, as the Hawthorne studies have shown, while some research may be technically valid, they may have little to do with what actually happens in people's lives. Ecological validity therefore concerns the degree to which social scientific findings are applicable to people's everyday, natural social settings.

This study has employed various measures to assure a high degree of validity to the findings. To increase measurement validity, any measure that was constructed was tested on previously established positive and negative examples. External validity was considered when choosing the samples of cases to be studied, and effort went into choosing cases for which literature could provide insights, while still assuring an interesting study on a current topic. Nonetheless, the limited sample of studied cases means that a generalization to the general patent landscape cannot be guaranteed.

Internal validity has been deemed to be a minor issue in this study since the aim of the study is not to establish any causal relationships but to build a framework of patent data metrics with previously established correlations. Likewise, ecological validity has been considered adequate since the study is performed on static data rather than by interfering in people's everyday lives.

4
Framework Construction

The following chapter presents the construction of a new patinformatics framework for how patents can be used to gain insights for technology selection, and the knowledge needed in the process. The chapter is divided into four different sections. The first section introduces the foundation upon which the whole thesis is built, namely *patents and the patent system*. Section two presents an in-depth analysis of *patent information*, including which different data features can be studied and their drawbacks and benefits. The third section presents the findings about *patinformatics*, and in particular which aggregated patent data metrics previous researchers have used to study technological fields. Finally, the fourth section brings the knowledge from the previous sections together in the creation of a new *patinformatics framework for technology selection insights*.

4.1 Patents and the Patent System

A patent is a set of exclusive rights granted by an intergovernmental organization or a sovereign state to an inventor or assignee to exchange detailed public disclosure of an invention for a limited period of exclusivity (WIPO, 2015). Implicit in this, the patent protection confers its owner the right to prevent third parties from using, making, selling, or importing the claimed invention for up to 20 years (WIPO, 1994).

In order for an invention to be accepted as patentable under the World Trade Organization's (WTO) TRIPS agreement (ratified by the majority of the world's nation-states) it must, in general, satisfy the following three requirements; *(1) the invention must show an element of novelty, (2) it must show an inventive step, and (3) be capable of industrial application* (WIPO, 1994). For patents to be granted, a generalized procedure is visualized in Figure 4.1.

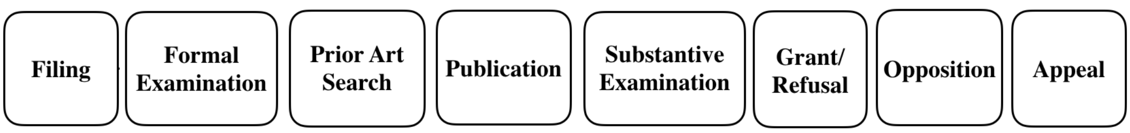

Figure 4.1: Patent granting procedure

The first step in securing a parent is to file an application at either a national, regional, or international patent office. This initial filing is considered as the "priority filing" from which additional national, regional, and international filings can be made within one year (priority period). Using the above criteria for patentability, one or more patent examiner(s) from the patent office is then assigned to ensure that all the administrative formalities have been complied with, and furthermore check the invention's susceptibility of industrial application. To assess the subject matter's degree of novelty and inventive step, the examiner(s) conducts a prior art search. Regardless of the outcome of the examination process, the patent application is, in most countries, published 18-months after its priority date. Once the substantive examination has been completed, the applicant has the opportunity to amend the application if the patentability requirements are not met. If the examiner(s) decide to grant the patent, many patent offices allow third parties to oppose the granted patent within a specified time period to assure that the patentability requirements really are satisfied. Generally, patent grant or patent refusal decisions can be challenged before a court or an administrative body (WIPO, 2015).

The main rationale for adopting a patent system is to *provide incentives* to individuals by recognizing their creativity and offering the possibility of reward for their inventions. A patent carries two important functions. First, it serves as a protection allowing the patent holder to exclude others from commercially exploiting the invention covered by the patent in a certain country or region in which the patent was granted and for a specific period of time (generally not exceeding 20 years from the filing year). The patent prevents competition and thus gives the patent holder the opportunity to profit from the patent by selling the invention for a higher price than would have been the case without a patent. The second important function of the patent system is disclosure. By giving the public access to information regarding new technologies, available to any individual or organization worldwide, anyone is allowed to learn and build on the knowledge contained in the patent document. This enriches the total body of technical knowledge in the world, while simultaneously providing valuable information and inspiration for future generations of inventors and researcher, stimulating innovation and contributing to economic growth (WIPO, 2015).

4.2 Patent Information

Patent information comprises all information which has either been published in a patent documents or can be derived from other complementary documents. This includes technical information but also legal information, business-relevant information, and public policy-related information. With some databases containing more than 100 million patents (EPO, 2018), patent information constitutes the largest, most up-to-date and well-classified collection of technical documents on new and innovative technologies in the world (WIPO, 2015).

4.2.1 Primary Patent Information

A patent document is divided into several sections which provide different types of information regarding the invention. The sections comprised in the patent document is shown in Figure 4.2 below and then explained further in Table 4.1.

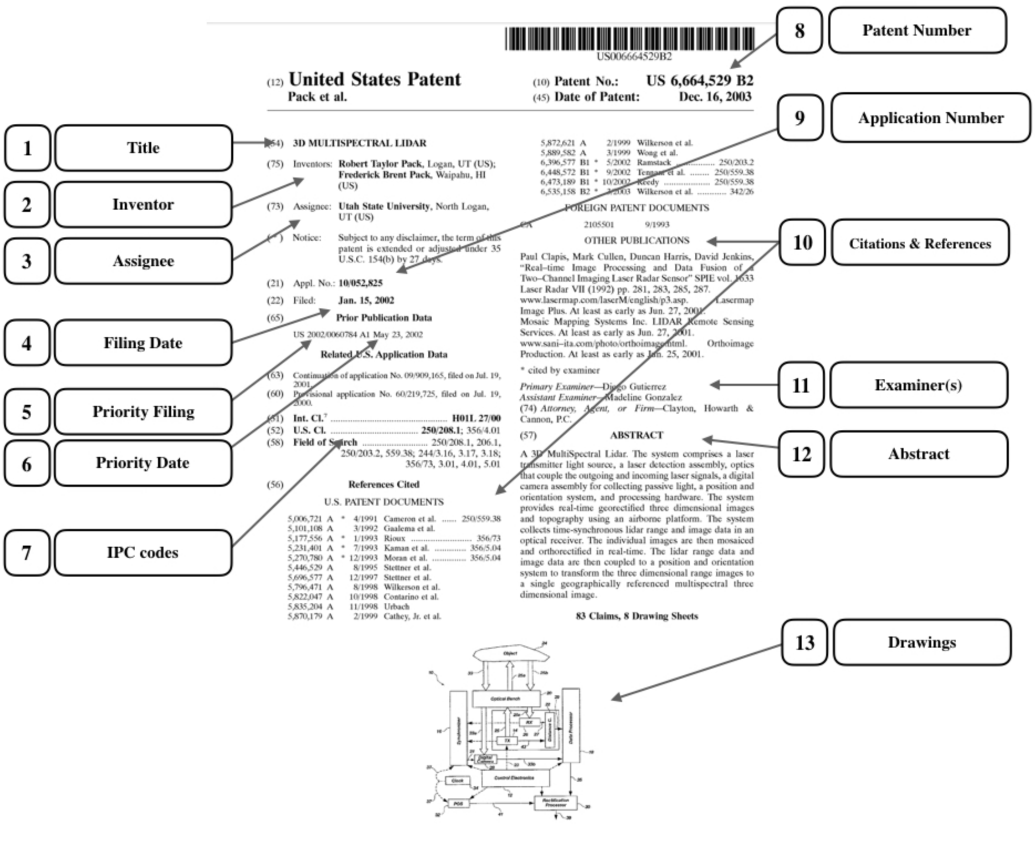

Figure 4.2: Sample patent document front page.

Table 4.1: Patent features contained in a patent document

No.	Primary Patent Feature	Description
1	Title	Brief one-paragraph description of the invention covered by the patent.
2	Inventor	Name of the person(s) who invented and developed the new technology.
3	Assignee	Name of the individual or company assigned to have ownership of the protected invention.
4	Filing Date	Date when an individual patent application was submitted at a particular patent office.
5	Priority Filing	Original first filing number of the patent application, from which further national, regional and international filings can be made within the priority period of one year.
6	Priority Date	Date of the first filing from which the one-year priority period for further applications starts.
7	IPC Codes	Code classifying the technological characteristics of the invention in a uniform manner according to hierarchical classification systems.
8	Patent Number	Identifier following a set standard typically consisting of six to eight digits, to distinguish published patent documents according to type and status. Assigned by the Patent Office.
9	Application Number	A unique identifier assigned by the Patent Office once the patent application is filed.
10	Citations and References	Prior art references to related invention uncovered by the applicant or the examiner.
11	Examiner(s)	Names of the examiner(s) involved in examining the patent.
12	Abstract	A summary of the general nature and core of the patent's subject matter.
13	Drawings	Some patents contain drawings illustrating the invention, its embodiments or prior art.
14	Background	Describes the general nature of the problem to be solved by the invention and the state of prior arts.
15	Description	A detailed explanation of known existing technologies related to the new invention, specific embodiments of the new invention and explanation on how this invention could be applied to solve problems not addressed by the existing technologies.
16	Claims	The legal definition of the subject matter for which protection is sought or granted. The invention and its unique features are defined in the claims, where each claim is one single sentence in a clear, concise and legalistic form, fully supported by the description.

Patent features can be classified into structured and unstructured feature groups. Feature 1-11, generally referred to as a patent's bibliographic data, appear on the front page of a patent document and follows a uniform format across each patent and are thus know as structured features. The same semantic structure follows in the rest of the document where the information is divided into sections with fixed headings constituting the background, description, and claims of the invention. This information is, however, formulated in free text, varying in content and length, and is consequently considered to be unstructured data containing the technical information

To facilitate the access to the information contained in patent documents every patent is classified to a technological field using classification schemes, i.e a system of codes that groups patented inventions according to technical areas. The International Patent Classification (IPC) system comprises approximately 70 000 IPC codes for different technical areas and is the primary classification scheme used by patent offices worldwide (EPO, 2017). The primary purpose of the IPC system is to obtain an internationally uniform classification of patent documents to establish an effective search tool for the retrieval of patents. Based on the IPC system, the Cooperative Patent Classification (CPC) system has been jointly developed by the European Patent Office (EPO) and the US Patent and Trademark Office (USPTO). The CPC is a more detailed classification including an additional Y section that monitors new technological developments and cross-sectional technologies that do not fit in any section of the IPC (EPO, 2017).

In brief, the IPC system follows a hierarchical structure of classification, where whole bodies of technical knowledge are broken down using the hierarchical levels of classification symbols. Each patent is assigned to at least one classification symbol indicating which subject the invention relates to, but additional classification symbols and indexing codes can be assigned to the patent to give further details of the content. Figure 4.3 below shows a classification symbol of the form "G01S 13/00" (IPC code representing sonar systems), according to the classification system's layout.

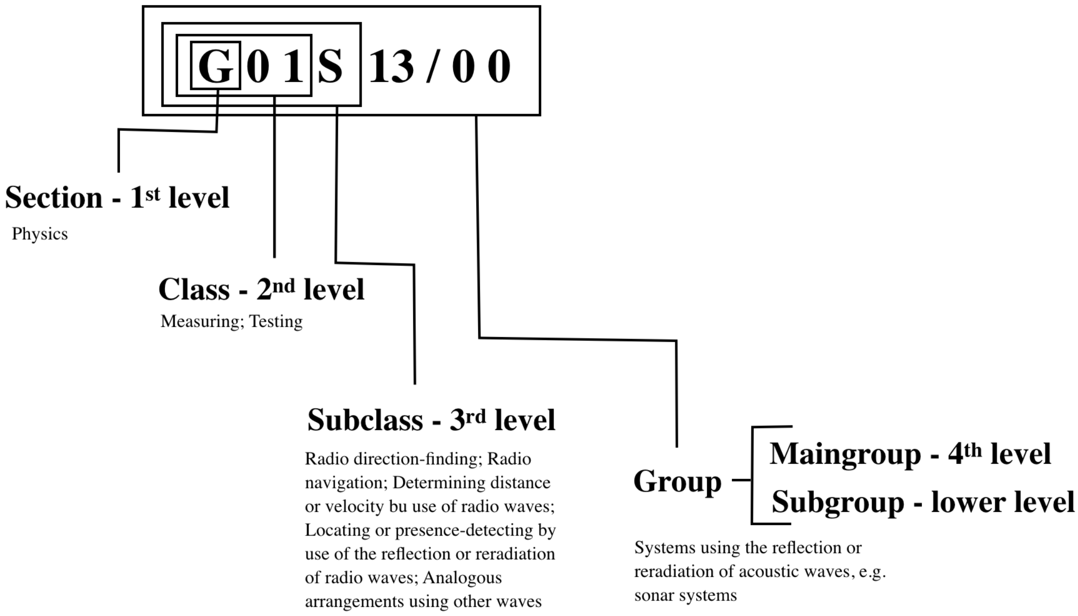

Figure 4.3: A complete classification symbol (IPC code), representing sonar systems.

Sections, represented by the first letter in the IPC code, are the highest hierarchical classification level and are to be considered as a very broad indication of the patent's content. Each section is then subdivided into *classes*, consisting of the section symbol followed by a two-digit number. Classes are the second hierarchical level, and each class can be broken down into subclasses comprising the third hierarchical level of classification. These are represented by the class symbol followed by a capital letter. Finally, *subclasses* are broken down into *groups*, which are either main or subgroups depending on the main group of the classification (WIPO, 2018).

The embodiments or technical features of a patented invention are comprised in *the claims* of the regarded patent, further defining the scope of the legal protection of the invention conferred by the patent or patent application (WIPO, 2015). These are stipulated by the applicant who is required to point out and distinctly claim the subject matter which he or she considers as his or her invention (EPO, 2017). Two main types of claims exist; (1) independent and (2) dependent claims. All patents contain one or more *independent claims*, which are stand-alone claims directed to the essential embodiments and features of the invention (WIPO, 2017). There are two types of independent claims; a product claim (such as for a composition of matter, machine, apparatus or device) and a process claim (for a process/method, activity or use) (EPO, 2017). An independent claim is followed by one or more *dependent claims* concerning particular embodiments of the invention (EPO, 2017). A dependent claim may refer back to one or more independent claims, to one or more dependent claims, or to both independent and dependent claims (EPO, 2017).

Furthermore, published patents can build or be built on previous work (prior art). The phenomena of referencing to a prior art that is considered relevant to a current patent application refers to the term *backward citations*. A cited document can either be a patent document or an item of non-patent literature (WIPO, 2015). These may be added by the patent examiner and/or the applicant of the patent and has the general purpose of helping the examiners assess the patentability of a claimed invention (WIPO, 2015).

4.2.2 Complementary Patent Information

Except the information written on the patent document itself, other sources aggregate information about patents over the course of their lifespan. Previous literature suggests that complementary patent information can be used as a source of information about patents and patenting activity and thus provide some business insights about companies' technology investments (Lanjouw et al, 1998). A list of the main categories of complementary patent information can be seen in Table 4.2 below.

Table 4.2: Complementary patent features

No.	Complementary patent feature	Description
1	Designated states	Countries which the rights may be extended if the application is regional or international.
2	Legal status	Indicates whether the patent application has been granted or not, and if so, which countries and regions the patent has been granted, and whether the patent still is valid, or if it has been expired or invalidated in a particular country or region.
3	Forward citations	Additional patent documents citing a particular patent after publication.
4	Renewal fee payments	In most countries, the patent holder(s) must pay periodic renewal fees to maintain granted patents.
5	Litigation data	Document relating to legal processes which unfold when a patent owner enforces their right by suing another for patent infringement.
6	Prosecution data	Data revealing correspondence between the patent applicant(s) and the examiner(s) throughout the patent prosecution process for the invention.

4.2.3 Advantages and Disadvantages of Using Patent Information

Since patents contain a large amount of technological information, patent information can be used to support technology management (Ernst, 2003). As for any technology indicator, the use of patent information has both advantages and disadvantages (Archibugi & Pianta, 1996).

Bonino et al. (2010) distinguish between three main classes of patinformatics tasks: patent search, analysis, and monitoring. Trippe (2002) also emphasized the difference between searching and analyzing, when he wrote that patent searchers try to find a needle in a haystack on a microscopic level of patent data, while patent analysts try to “identify a haystack from space” on a macroscopic level. In accordance with section 1.5 *Delimitations*, this thesis focuses on the tasks within patent analysis.

Archibugi & Pianta (1996) mention four distinct advantages:

1. Since patents aim to capture the inventions which are expected to have a commercial impact, they are a direct result of the inventive process and succeeds in capturing the commercial and proprietary aspects of technological development.
2. Since patents are divided into different technological fields, they contain information on both the rate of inventive activity and its direction.
3. Patents are available for long time series and in large numbers.
4. Patents are public documents, and no information is covered by confidentiality.

Furthermore, Asche (2017) means that a large portion of the information disclosed in patents is never published anywhere else and that patents thus provide a unique source of information. One of the patent requirements is that the invention must be explained in sufficient detail for a “person skilled in the art” to be able to carry it out. Thus, patents carry much more detailed information about the technology than other types of technical or scientific publications. Basberg (1987) writes that *“one of the advantages of patent statistics used as a technology indicator, is the possibility of constructing long and complete time-series.”* and that *“It is possible to maintain that in using aggregated patent statistics, sheer quantity will secure meaningful mean values”*. Furthermore, Ernst (2003) means that patents, unlike information published on e.g. company websites, have been examined and granted by a patent office, and is therefore a more objective measure of R&D activities.

Among the disadvantages of using patent data is that the propensity to patent differs between companies, industries, and countries. Not all inventions can be patented, and not all inventions that can be patented are. For instance, the software industry has historically relied on copyright protection, and firms often choose to protect inventions by the use of trade secrets (Archibugi & Pianta, 1996). Secrecy might be particularly preferred when the expected life of the invention is very short and an invention might be obsolete before the patent is granted, or when the expected life of the invention far exceeds the usual 20 year protection period of patents. Each national patent office has its own characteristics which affect the propensity of inventors to file for patents. Furthermore, the economic and technical value differs greatly from patent to patent, and using simple patent counts may consequently not be a complete indicator of larger trends (Ernst, 2003). Finally, most jurisdictions publish patent applications 18-months after they are filed. As such, there is always a time lag between the point in time the patent application is being published and the time at which the invention is completed. One could, however, argue that in many industries, R&D times for new products and services exceed 18-months and the products based on the patents are therefore introduced to the market behind the patent is published.

The advantages and disadvantages of patent information are summarized in Figure 4.4 below.

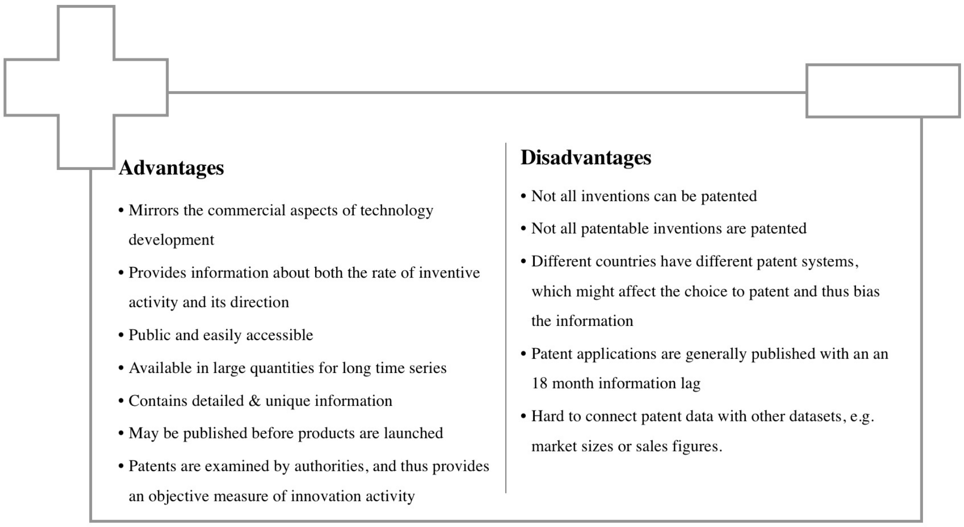

Figure 4.4: Advantages and disadvantages of using patent information as a technology indicator.

As a conclusion, the American economist Jacob Schmookler seemed to have understood a great deal about patent information fifty years ago when he wrote that

"We have a choice of using patent data cautiously and learning what we can from them, or not using them and learning nothing about what they alone can teach us." (Schmookler, 1966).

4.3 Patinformatics

Previous researchers have used different definitions to describe the process of using patent information to support business decision-making. Trippe (2003) used the term patinformatics, which he means encompasses all macro-level kinds of analyzing patent information to educe business decisions. The term thus includes all three steps in the framework presented in section 2.1.3. *Patinformatics research process.* Aristodemou et al. (2017) instead suggested that the term "patent analytics" could be defined as *"the science of analysing large amounts of patent information to derive meaningful insights to support decision-making, which constitutes of the deployment of different technologies, techniques, and approaches"*. This study has adopted Trippe's definition for two reasons. First, the term patent analytics suggests a focus on data analysis tasks and techniques, while the term patinformatics instead emphasizes the information aspect and thus edges closer to the actual value creation. Secondly, chosen definitions should allow for precise and meaningful distinctions between different concepts, and the term "patent analytics" might be confusing if used to denote a larger concept than just the data analysis tasks included in "patent analysis" or "patent data analysis".

The following subsections are structured according to the patinformatics research process presented in section 2.1.3. Relevant findings from the literature review are therefore presented about the key issues of patent data pre-processing, patent analysis, and discovered knowledge. Furthermore, the section ends with the creation of a framework for how patinformatics can be used for technology selection, where the previously presented findings are put in the context of the current research problem and their relevance is discussed.

4.3.1 Patent Data Pre-processing

To recap from the theoretical framework, patinformatics researchers today have access to publicly available databases with most of the pre-processing work already performed. The pre-processing task is therefore focused on limiting the range of documents to be analyzed through patent retrievals which generate patent sets that represent the phenomena (technology, company, country) being studied. A good patent set should be both complete (i.e. the set should include as many as possible of the relevant patents), relevant (a large share of the patents in the set should indeed be relevant) and possible to replicate with a new search (Benson & Magee, 2012). This subsection will present the most common patent retrieval methods along with newly developed methods that might be useful in the study at hand.

4.3.1.1 Boolean Search Methods

The most basic patent retrieval method is the boolean keyword and classification search (Benson & Magee, 2012). This method relies on keyword search terms, technology classifications and boolean operators (like AND, OR and NOT) to generate relevant and complete sets of patents. This method has been relatively unchanged since its introduction over thirty years, ago, and getting the right search results still require understanding set theory and exclusion. The process is often iterative where each iteration represents an evaluation of the previously generated results and decisions to include or exclude certain new keywords or technology classes from the search query. Extensive domain knowledge of the phenomena being studied is often a prerequisite to making these kinds of judgments. For instance, Traijtenberg (1997) used knowledge from his previously performed case-study on the field to generate search terms for his well-known study of computer tomography. Traijtenberg also read the abstracts of every patent in his result set to exclude inappropriate patents.

Needless to say, not all patent searches can be performed with the same amount of domain-specific knowledge and resources as Traijtenberg's. The boolean search method is especially hard to apply when creating many different independent patent searches, since the resources needed for each search is high, and the time needed scales linearly with the number of searches to be performed.

4.3.1.2 Hybrid keyword-classification method

To combat the drawbacks of the classic boolean search method, Benson & Magee (2012) proposed a less work-intensive patent retrieval method specifically to retrieve patent sets that represent different technological fields. Their hybrid keyword search method was tested on granted US patents and consists of three steps:

1. A pre-search, where a two-word keyword search is made to find the most relevant documents for the technology being retrieved. Benson & Magee exemplifies this by searching for “Solar Photovoltaic” in the titles and abstracts of the patents in their database.
2. A ranking of IPC and UPC classes on the retrieved patents, where the most commonly occurring IPC and UPC classes are listed first.
3. Selecting the overlap of the most representative IPC class and UPC class. That is, searching the database for patents that are classified both in the most representative IPC class and in the most representative UPC class.

The intuition behind this search method is that the patent classification is based on the expertise of the patent examiner, and if he or she classifies the documents into the most representative class in two separate classification systems, the chance is high that the patent is indeed representative of the keyword being searched. With only three simple steps, the hybrid keyword-classification (HKC) method takes substantially less time to carry out than an iterative boolean search method. This enables a single patent analyst to create hundreds of patent sets in just a single day. Furthermore, the use of classification overlaps makes the method more robust than a simple two-word keyword search. For instance, the search queries “photovoltaic electricity” and “solar power” share the same most representative IPC and UPC classes, and while a simple keyword search gives very different resulting patent sets, the HKC method returns the same set.

Since the HKC method builds on specifically using both IPC and UPC classes, a limitation of the method is that it can only be applied to patents classified by both systems, e.g. US patents. While one could make the case that the same method could be applied to the overlap of CPC and IPC classes as well, this idea has not been tested empirically, and since the CPC system is basically an extension of the IPC system, the information gained from considering the overlap between those two systems may not be as high. While the HKC method results in more robust, complete, and relevant patent sets than a simple keyword search, it is important to consider that it is not intended to be a replacement for expertly selected sets of patents (Benson & Magee, 2012). When trying to reproduce Traijtenberg’s (1987) patent set, the HKC method reached a relevancy of 26% and a completeness of 30%.

4.3.1.3 Cipher Automotive Method

In 2017, Aistemos launched the Cipher Automotive extension, which they call *"the first sector analytics platform to focus on the automotive sector"*. They further write that *"Cipher Automotive includes a comprehensive taxonomy of technologies developed in collaboration with a number of OEMs and Tier 1s. It is the latest on AI and machine learning."*.

Cipher Automotive encompasses a machine-learning based search engine that lets users search for patents and tag the results as positives or negatives. This trains a so-called classifier and the search results are refined continuously through a supervised learning process, where patents that are marked as positive generates new, similar documents into the set, and patents that are marked as negative excludes other similar documents from the set. The similarity between documents is based on their claims, abstracts, CPC codes, and citations. According to Aistemos[2], this leads to much higher levels of both precision and recall than traditional boolean searches. Marcus also mentions that the search engine is less prone to simple errors, such as missing characters, than boolean searches, and that boolean searching is limited to semantic searching and does not have the same deep ability to find links that machine learning has. In addition, the classifier ranks every document in the set, and lets you see not only which patents were included in the set, but also which were excluded and how narrowly. In boolean searches, you only see the patents in your search result and there is no way of knowing what was left out.

For users that want to look at existing automotive technologies, Cipher's team of artificial intelligence, intellectual property, and automotive industry professionals, has worked together to train classifiers to retrieve patent sets for more than 200 automotive technologies. These pre-defined patent sets combine the extensive knowledge of the industry experts with the benefits of the supervised learning based search engine, and thus offers patent sets that are very representative of the technology being studied. One important point that should be expressed is that Cipher has limited the range of assignees on the automotive patents to only automotive companies (OEMs), their first-tier suppliers, and non-practicing entities (NPEs). This creates a risk that patent searches through Cipher might miss out on non-automotive companies patenting in the field. At the same time, the list of assignees covers hundreds of companies (including autonomous drive leaders like Waymo, Uber, Tesla, and GM) that together own close to seven million patent families, which can be considered a reasonably large number for most studies.

4.3.2 Patent Analysis

The main task of the patinformatics process is to analyze the retrieved patent sets in order to gain insights that can support business decision-making. In line with Basberg (1987), this can be done by defining certain patent data metrics as technology indicators, where an indicator is as a proxy that may help discover trends.

[2]Interview performed 2018-05-10 with Marcus Malek, Head of Strategy at Aistemos

This approach is based on the underlying assumption that patents do indeed reflect the underlying inventive activity in the fields being studied.

This subsection starts with a review of previously used patent data metrics. After presenting the previously used metrics, the section concludes with a subsubsection about best practices within patinformatics research.

4.3.2.1 Previously Used Patinformatics Metrics

Metrics used in previous research are here presented in order according to what patent data feature they are based on, meaning that for instance all metrics based on citations are presented together. For each metric, a formula for how to calculate the metric is given together with the underlying rationale of the metric (what phenomena it is supposed to indicate), and a review of how well it actually correlates with that phenomena, as shown by previous research. To put the metrics into the context of this study, their relevance for technology selection decisions were also assessed.

4.3.6.1.1 Metrics Based on Backward Citations
Previous research indicates fragmented views about the implications of backward citation as a patent metric. On one hand, fewer backward citations might indicate a larger degree of novelty and thus a higher economic value. On the other hand, fewer patent citations may indicate a less thorough prior art analysis and a higher chance of patent invalidation in court proceedings.

Harhoff et al. (2003) found that the number of backward citations correlated positively with the economic value of patents. Lai & Che (2009) found that the sum of all backward NPL citations (*strength of patentability*) showed a small but significant correlation to the damages awarded plaintiffs in US patent litigation cases. Benson & Magee (2015) used the metric backward citation immediacy (or *average citation lag*), which they defined as the average age of backward citations for each patent (averaged over the domain) at the time of the citing patents publication, and showed that this metric was strongly correlated with the technological improvement rate of a domain. This metric built on the rationale from Price's (1965) study, which showed that fast improving scientific fields are close to a research frontier that relied on recently cited papers. They also investigated a joint metric which they called *recency and immediacy*, which they defined as the average date of publication for backward citations from patents in a domain.

Table 4.3: Patent metrics devised from backward citations

Metric	Definition
Average number of backward citations	Total number of backward citations in a patent set / total number of documents
Breakthrough inventions	Total number of foreign patent citations, domestic patent citations and non-patent citations / total number of documents
Average citation lag	Average age of backward citations for each patent (averaged over the patent set) at the time of the citing patents publication
Recency and immediacy	Average publication date for backward citations from patents

4.3.6.1.2 Metrics Based on Forward Citations

As presented in section 4.2, assignees are obliged to cite the inventions and scientific research that has been fundamental for the creation of their invention. The citations given to a patent document after its publication (forward citations) are often considered the most important indicator of the technological impact of a patented invention (Archibugi & Pianta, 1996). Harhoff et al. (1997) found that the economic value of individual patents, as measured through a survey with the assignees, rose with the number of forward citations. For the most valuable patents, a single citation implied an average of more than 1 million USD in value. Hall et al., (2005) also showed that the average number of forward citations per patent correlated strongly with the market value of firms, and concluded that “If a firm’s quality of patents increases so that on average these patents receive one additional citation, the firm’s market value would increase by 3%”. Furthermore, forward citations are positively correlated to patentees decisions to pay renewal fees (Hegde & Sampat, 2009), which indicates the economic value of individual patents.

In accordance with the indications that forward citations indicate both technical impact and economic value, the first metric introduced here is *the average number of forward citations*. This is simple to calculate, yet gives a good overview. The drawback with this metric is that older patents have, on average, more citations than newer ones, and the metric thus becomes distorted when used for comparing technologies of different age. One way to solve this is to group patents according to their application or publication year, and then rank different technologies according to their mean percentile score of citations. For instance, Squicciarini (2013) defined what he called *breakthrough inventions* as the top 1% cited documents for each year. Benson & Magee (2015) instead devised a metric they called *immediate importance* as the average number of citations that a patent receives within 3 years of publication. Both rank percentile scores and age-based citation metrics relieve some of the problems with comparing patent citations across technologies and over time.

Table 4.4: Patent metrics devised from forward citations

Metric	Definition
Average number of forward citations	Total number of citations in a patent set / total number of documents
Breakthrough inventions	Number of breakthrough patents in a patent set / total number of documents in patent set, where breakthrough patents are defined as the top 1% most cited patents in the same cohort
Immediate importance	Average number of citations that a domain patent receives within 3 years of publication

4.3.6.1.3 Metrics Based on NPL Citations

Non-patent literature (NPL) citations are backward citations from patent documents to other kinds of documents, of which a majority is scientific papers (Callaert et al., 2006), but also conference proceedings and databases (Squicciarini, 2013). Callaert et al. (2006) found that NPL citations were consequently a good indicator of the 'science intensity' of the inventions contained in the patents. Branstetter (2005) also found that the number of NPL citations of patents was correlated to certain quality indicators, like the number of forward citations or the generality index. On this point, Harhoff et al. (2003) contradicted Branstetter in their findings, only finding statistically significant relationships between NPL citations and patent value within the fields of pharmaceuticals and chemicals. NPL citations can also be indicative of the current stage of the technological life cycle of a technology since technologies tend to be more science-intensive in early stages of the cycle.

An easy way to track *science intensity* in a technological field is the average number of NPL citations. This metric has been used by several previous researchers (e.g. Trappey et al., 2013; Squicciarini, 2013). Benson & Magee (2015) instead proposed to use the *NPL Ratio*, that is, the ratio of scientific citations to patent citations in a field.

Table 4.5: Patent metrics devised from NPL citations

Metric	Definition
Science intensity	Total number of NPL citations / total number of documents
NPL ratio	Ratio of scientific citations to total citations

4.3.6.1.4 Metrics Based on IPC Classes

IPC classes indicate the technological breadth of a patent. If a patent set contains documents from many distinct IPC classes, the set covers a broad range of technologies. The number of different IPC classifications on a patent is sometimes referred to as the "patent scope", and can be calculated either on an individual patent basis or as the scope of a range of patents, like a company's patent portfolio. Lerner (1994) found that the scope of companies' patent portfolios was highly correlated with the value of biotechnology firms in venture capital funding rounds and that one standard deviation larger patent scope was associated with a 21% increase in company value.

Squicciarini (2013) defines patent scope as the number of distinct 4-digit IPC subclasses listed in a patent document. An aggregate metric, *average patent scope* can be constructed by calculating the average patent scope for each patent. Another aggregate metric was proposed by Altuntas et al. (2015), who call the total number of distinct IPC classes in a patent set its *expansion potential* and *the average expansion potential patent power.*

Table 4.6: Patent metrics devised from IPC classes

Metric	Definition
Average patent scope	Average number of IPC classes assigned to a patent
Expansion potential	Total number of distinct IPC codes
Patent power	Total number of distinct IPC codes / total number of patents

4.3.6.1.5 Metrics Based on Number of Patents

Many previous studies have constructed metrics based on the number of patents in a field. The simplest way to do this is to just count the number of patents per field. Since patenting and R&D spending are strongly correlated (Basberg, 1987), the total amount of patents in a field indicates how much money has been spent on R&D efforts. Similarly, to estimate the current R&D efforts in a field, the yearly number of patents or patent applications can be counted.

One of the most prominent use cases for metrics based on the number of patents is estimating so-called *technology life cycles* (TLC) or s-curves. Altuntas et al. (2015) used the cumulative number of alive patents in a field as a boolean indicator for TLC; if the number of patents was increasing, the technology was considered to be in the growth stage of the TLC, whereas fields with a decline in the cumulative number of patents were considered to be in the saturation or decline phases. Benson & Magee (2015) instead chose to refer to the cumulative number of patents as *effort*. Likewise, if the time period being studied is recent and few documents have been granted, the *cumulative number of patent applications* or the *yearly number of patent applications* can be assessed. Ernst (2003) assessed the *relative patent growth*, meaning the patent growth in recent years related to the total number of patents in the field, with the rationale that technological fields with high relative patent growths are more attractive than those with low relative patent growth. Finally, Benson & Magee (2015) used the term *recency* to denote the average publication year for all patents in a domain and found that more recent technological fields were developing faster.

Table 4.7: Patent metrics devised from the number of patents

Metric	Definition
Technology life cycle	Boolean; if cumulative number of patents increase, then technology is in the growth stage
Effort	Number of issued patents
Cumulative number of patent applications	Total number of patent applications
Yearly number of patent applications	Number of patent applications in a given year
Relative patent growth	Yearly number of patent applications / cumulative number of patent applications
Recency	Average publication year for all patents

4.3.6.1.6 Metrics Based on Patent Family Size

The patent family size (also called international scope or geographic family size) of a patent family is often defined as the number of distinct jurisdictions that share the same priority document (e.g. Ernst, 2003; Squicciarini, 2013). This shows the areas of exploitation of an invention, with the intuitive meaning that companies will aim to get broader protection for those inventions they consider more valuable. Lanjouw et al. (1998) found that the international scope of patents does correlate with payment of patent renewal fees, which indicates that companies assign greater value to those patents.

Patent family size, or *average patent family size* in aggregate studies, has been used as an indicator of patent value in many different studies (e.g. Ernst, 2003; Lai & Che, 2009; Archibugi & Pianta, 1996).

Table 4.8: Patent metrics devised from patent family size

Metric	Definition
Average patent family size	Sum of all patent family sizes in patent set / total number of priority documents

4.3.6.1.7 Metrics Based on Assignees

According to Basberg (1987), the number of assignees patenting in a field indicates the interest in the technology. This can be used as a cumulative metric (*cumulative number of assignees*) or on a year to year basis (*yearly number of assignees*). Furthermore, the *yearly number of new entrants* in a field can be calculated by assessing how many companies are filing their first patent applications in a field in the given year. This number can be compared to the total number of applicants in a field to obtain its *relative assignee growth.* Bass & Kurgan (2009) also used the metric "*in top assignees*", defined as a boolean variable assigned to 1 if the patent has an assignee in the 95th percentile of citations, and found that this metric was strongly correlated to the number of citations patents got. For aggregate studies, the *number* *of top assignees* in a field can be assessed.

Table 4.9: Patent metrics devised from assignees

Metric	Definition
Cumulative number of assignees	Count all distinct assignees having filed patent applications
Total number of assignees	Count all distinct assignees
Yearly number of assignees	Count all distinct assignees filing patent applications for a given year
Number of new entrants per year	Number of companies filing their first patent application in the patent set on the given year
Relative assignee growth	Number of new entrants per year / total number of assignees up to that year
Number of top assignees	Sum of all patents where the assignee has an average incoming citation count greater than the 95th percentile

4.3.6.1.8 Metrics Based on Claims

Claims are the sole determinant of what aspects of a patented invention that can be legally protected and enforced (Squicciarini, 2013). A larger number of claims therefore makes patents harder to invalidate and gives the right a broader scope, which is the rationale behind some researchers referring to the number of claims as the strength of the property right. Archibugi & Pianta (1996) also mean that the number of claims gives information about the range of novelties in a patent.

Previous studies (e.g. Lai & Che, 2009) have sometimes made a difference between independent claims and dependent claims, with the rationale that the independent claims are more important than dependent claims. Lanjouw & Schankerman (2001a) found that patents with more claims were much more likely to be involved in US lawsuits, and meant that this indicated value. Furthermore, patent renewal rates increase with the number of claims in a patent (Trappey et al., 2012), and the average number of citations a patent receives decrease with the length of claims (Okada et al., 2016). Thus, it can be useful to study the *average length of claims* as a proxy for patent strength. Squicciarini (2013) instead proposed to simply use the *average number of claims* as a measure of patent strength, whereas Lai & Che (2009) also investigated the *average number of independent claims*.

Table 4.10: Patent metrics devised from claims

Metric	Definition
Average length of claims	Sum of all characters in all claims in all patents / total number of patent documents
Average number of claims	Total number of claims / total number of patent documents
Average number of independent claims	Total number of independent claims / total number of patent documents

4.3.6.1.9 Metrics Based on Patent Renewal Fees

In most countries, patent protection is limited to 20 years from filing the patent application. Simultaneously, most countries have progressively increasing annual (or semi-annual) renewal fees (Basberg, 1987), meaning that not all patents are kept alive for the full 20-year period. Decisions from assignees to pay renewal fees thus indicates that the assignees consider the legal protection granted by the patent more valuable than the fee, and technological fields where patents are being kept alive for a longer period can be thought to be more economically viable than other fields. Svensson (2012) found that renewed patents were indeed commercialized to a larger degree than those allowed to expire.

Squicciarini (2013) proposed the use of simple metrics such as the *renewal time*. This metric is, however, skewed by design, as it is prone to exhibit higher values for older technological fields than younger, but can be corrected for by splitting the data into year-based cohorts. Archibugi & Pianta (1996) also proposed that the *average cumulative renewal fees per patent* should be calculated, but this task grows very complex with large patent sets from different geographies and application years.

Table 4.11: Patent metrics devised from patent renewal fees

Metric	Definition
Average renewal time	The total time from application to lapse (or current date if still alive) for all granted patents / total number of granted patents
Average renewal fees	The total cost of renewal fees paid to maintain the legal value of the patents

4.3.6.1.10 Metrics Based on Patent Grants

Since only granted patents are enforceable in judiciary systems, it follows that companies may want to speed up the examination process for the inventions they deem most valuable. Harhoff & Wagner (2009) performed a study of over 200 000 randomly chosen patent documents, where they found that companies accelerate prosecution for their most valuable patents, but also that they prolong the battle for those patents if they are likely to be refused. Nonetheless, the overall conclusion was that short grant proceedings were a good indicator of patent value.

Ernst (2003) proposed using the *ratio of granted to filed patents* as an indicator of patent value, and Squicciarini (2013) measured the *grant lag* in the number of days between patent applications and grant dates.

Table 4.12: Patent metrics devised from patent grants

Metric	Definition
Ratio of granted to filed patents	Number of granted patents / number of filed patent applications
Average grant lag	Sum of the number of days between application and grant date for all granted patents / total number of granted patents

4.3.6.1.11 Metrics Based on Inventors

Previous literature is sparse when it comes to using inventor data for predicting valuable technologies. However, Bass & Kurgan (2009) used a number of inventor-based features in a regression model trying to predict the most cited patents. The most predictive feature of more than 40 metrics used in this study proved to be *"in top inventors"*, defined as a boolean variable with a value of one if at least one of the inventors had an average citation count greater than the 95th percentile. Furthermore, the *inventor's average inventing time*, that is, the average difference between the first and last filed patent application by the same inventor, added further predictive power to this feature. The rationale given by the authors was that inventors that have been active for a long time, and also filed impactful patents, were more likely to do so again.

Table 4.13: Patent metrics devised from inventors

Metric	Definition
Number of top inventors	Boolean variable; assigned to one if at least one of the inventors has an average incoming citation count greater than the 95th percentile
Inventor's average inventing time	Sum of all inventors' time between first and last patent application / total number of patents in set

4.3.6.1.12 Metrics Based Jointly on Several Features

Some of the most predictive factors in patent analysis are not based on assessing a single patent data feature but rather on assessing the relationships between several features. This section thus serves as a catch-all for such metrics before moving onto the next key process in the patinformatics research framework.

To assess the *degree of maturity* of a technological field, Basberg (1987) proposed using a joint boolean metric based on the number of assignees and the number of patent applications. His rationale was that the number of assignees reflects the interest in the technology and the number of patent applications reflects the actual technological activity. If both of these rise, he considered the technological field to be in the developing stage, if both decrease he considered it to be in the maturity stage, and combinations of positive and negative values were considered to be in the R&D stage.

Another much used joint metric is the so-called *originality index*, first proposed by Trajtenberg et al. (1997). Patent originality refers to the breadth of the technological fields that a patent relies on according to its backward citations. The originality index can thus be seen as a metric for assessing the knowledge diversity behind a patent, and a high value of originality means that the patent cites prior art from many different IPC-classes. Harhoff & Wagner (2009) found that patents with higher originality had longer grant lags. Furthermore, Kaplan & Vakili (2012) performed a topic modeling of patents, where they found that patents with higher originality to be less likely to generate new topics (i.e. create technical breakthroughs) but more likely to be well-cited. This could be interpreted as original patents having larger economic, but lower technical, value than less original patents. Hall (2002) showed that the originality measure by design would be biased against patent sets with fewer patents, and the definition shown in Table 4.14 below uses the adjusted Hirschman-Herfindahl index to correct for that bias.

Following the rationale behind the originality index, Trajtenberg et al. (1997) also proposed a *generality index*. Whereas the originality index is based on backward citations and the breadth of knowledge applied to create the patented invention, the generality index is created from forward citations and indicates the range of later generations of inventions that benefit from a patent. The generality index will be high if the invention is cited by patents from many different IPC classes, i.e. the knowledge contained is more generally applied. As with the originality index, the generality index needs to be corrected for bias and the definition with the adjusted Hirschman-Herfindahl index is therefore shown below.

As previously explained, simple patent counts are often used to compare two companies' patent portfolios or to assess two different technological fields. Numerous studies have shown that the economic value of patents is very skewed, with a few patents amounting to a majority of many firm's patent portfolio values, and a long tail of low valued patents. Ernst (2003) proposed that this should be corrected for by multiplying the patents in a portfolio with some measure of patent quality (e.g. forward citations). In theory, this *adjusted technology value* would give a more accurate view of the value in the different patent sets.

Finally, there has been a number of studies where patent data has been used to predict technology improvement rates (TIR). Benson & Magee (2015) and Triulzi et al. (2017) both found robust and strong correlations between known technology improvement rates and different patent metrics for a wide range of technological domains. In fact, Triulzi et al. obtained a Pearson correlation coefficient (r) value of 0.795 (R^2 of 0.63) over 30 different technological fields using a single feature. The feature used (mean centrality cited ZRP) is a measure of how central the patents in the technological field are. This centrality measure is in turn calculated by the Search Path Node Pair (SPNP) values first introduced by Hummon and Doreian

(1989). SPNP is a joint metric based on both forward and backward citations and is calculated by multiplying the number of ingoing citation links to a patent with its number of outgoing citation links, both augmented by one. This is shown in Figure 4.5 below.

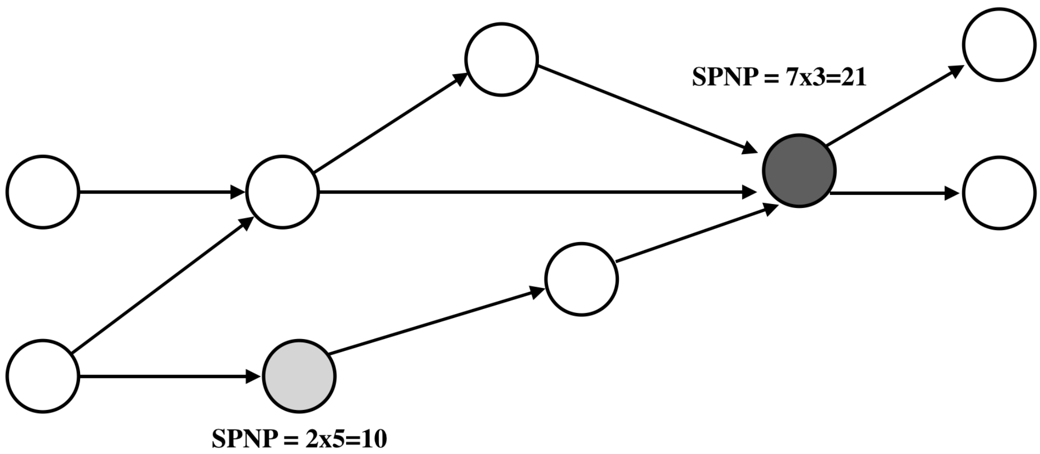

Figure 4.5: SPNP calculation

By defining centrality this way, it depends not only on the number of incoming and outgoing citations from the patent itself but also on the citations to and from the patents citing or being cited. Central patents cite other central patents and are being cited by other central patents in turn. Due to the dependency on forward citations, older patents have on average higher SPNP values than newer. To get a more informative measure of patent centrality, it is consequently useful to group and assess values in yearly cohorts. Triulzi et al. (2017) calculated the z-score (number of standard deviations from the mean value) for each patent's SPNP value, but patents could also be scored according to what percentile their SPNP values fall in.

Table 4.14: Patent metrics devised jointly from several features

Metric	**Definition**
Degree of maturity	If both the number of assignees and the number of patent applications increase, technology is in developing stage. If both decrease, the technology is mature, and combinations describe the R&D stage
Originality index	$$\frac{\text{number of patents}_i}{\text{number of patents}_i - 1} * (1 - \sum_j^{n_i} s_{pj}^2)$$ where s_{pj} is the percentage of citations made by patent p to patent class j out of the n_p IPC 4-digit (or 7 digit) patent codes contained in the patents cited by patent p.
Generality index	As for originality index but using forward citations
Adjusted technology value	Number of patents in a portfolio multiplied by some measure of patent quality
Average patent centrality (average SPNP value)	Average of ((number of links in backward citation chain + 1) * (number of links in forward citation chain + 1))

4.3.3 Discovered Knowledge

After the patent data metric section, it seems clear that patent data can be used in many different ways. A comprehensive patinformatics research process can be performed by analyzing all metrics and then leaving the reader to interpret the analysis results and link them to important insights for technology selection. However, communicating the results in this manner suffers from a number of drawbacks. First, it requires a thorough understanding of the different metrics and their implications, for anybody that wants to use the analysis results. Secondly, even with a basic grasp of patent data, understanding and communicating the results on a metric-to-metric level is very time-intensive and might be ill-suited for the C-suite managers that perform technology selection decisions. Thus, even though this communication strategy might maximize the absolute information available, the actual appropriation of information to knowledge and insights may be very low.

4.4 HELD patinformatics Framework for Technology Selection Insights

The HELD patinformatics framework for technology selection insights has been created as a communicative tool for maximizing the appropriated value of a patinformatics process. The main idea behind the framework was to optimize each part of the patinformatics process, with a special emphasis on the patent analytics process. The framework is shown in Figure 4.6 below.

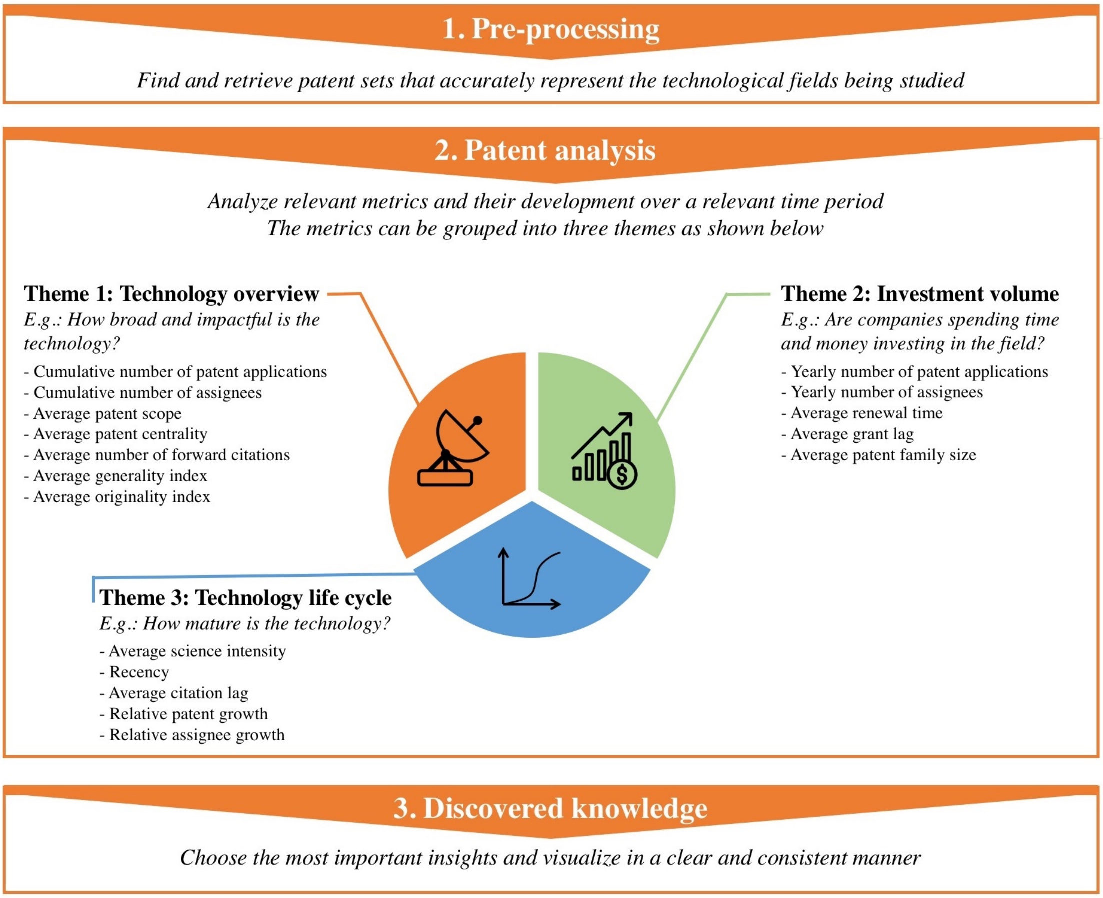

Figure 4.6: HELD patinformatics framework for technology selection insights

As has been previously mentioned, the task of pre-processing patent data is mostly about creating representative patent sets. The patinformatics researcher should therefore choose the best way to represent a technological field. Since this depends on the skills and knowledge of the researcher, there can be no one single best practice for this process. Researchers with in-depth industry knowledge of the field to be studied might be able to construct good boolean searches or manually tag documents in a supervised-learning based search engine, while less knowledgeable researchers might have to use pre-defined patent sets or consult external expertise to create complete patent sets with good recall.

The HELD framework groups patent data metrics together in “themes” based on what insights can be derived from them. Initially, the goal was also to do this in a MECE (mutually exclusive, collectively exhaustive) manner, but due to the multi-faceted nature of many of the patent metrics, this proved unfeasible. For instance, the total number of patent applications in a field might be an important factor for analyzing where in the technology life cycle a technological field is, but it can also be useful as a proxy for the amount of resources that has been invested in a field. Consequently, the framework consists of three themes that are not completely MECE, but where each field nonetheless might contribute valuable information about technological fields to a technology manager. Furthermore, each theme is built on a number of patent data metrics. Since the dynamic development over time might be as important as current figures (Ernst 2003), all metrics in the framework should be evaluated and visualized over time.

Finally, the insights gained from the patent analysis should be communicated in a clear and concise manner to allow for a high appropriation of the information to relevant stakeholders. How this communication should be done depends on both the target audience and the skillset of the patinformatics researcher, but can include videos, powerpoint decks, and posters.

The following three subsections present the three different themes for the patent analysis process and the rationale behind the presented clustering of the metrics.

4.4.1 Theme 1: Technology Overview

Theme one serves as an introduction to the technological field and an overview of its technical characteristics. First, the *total number of patent applications* filed in a field is introduced as an indicator of how much effort has been put on development activities within that field. Secondly, the *centrality* of the patents within the field is presented as a proxy for how fast the technological field is improving. Thirdly, the *average number of forward citations* indicates how technically impactful the inventions in the field is. The *patent scope* indicates how technologically niched and concentrated the inventions are, with higher scope meaning less niched inventions. *Originality* is a measure of the patent scope of the inventions being cited in a field, adding another level to the analysis of how technologically niched it is. Finally, the *generality* of a field is a measure of the patent scope of citing patents, thus indicating how broadly the inventions are being applied.

4.4.2 Theme 2: Investment Volume

The second theme deals with signals about the investment volume and economic value that the actors in a field assign to their patents. First, the *yearly number of patent applications* is examined as an indicator of how much resources are currently being put into inventive activities within the field. Secondly, the *yearly number of assignees* complements this by measuring the interest in the technology. Thirdly, *renewal time* is introduced as an indicator of how highly companies in the field value their intellectual property rights. Fourthly, the *grant lag* for patents within the field is examined, with the rationale that lower grant lags indicate that companies spend time and money to accelerate the patent prosecution process within the field.

Finally, the *average patent family size* is presented as an indicator of how widely companies try to protect their property right, where larger family sizes indicate that companies value their property rights higher.

4.4.3 Theme 3: Technology Life Cycle

One of the most important considerations when investing in new technology is understanding how mature the technology is (Gao et al. 2013). According to Altuntas et al. (2015), successful technology investments are most often performed when the technology is in the growth stage of the TLC.

A number of patent data metrics can be used as indicators of how mature a technology is, and the third theme of the HELD framework consequently focuses on technology life cycle considerations. First, *science intensity* can be measured, since technological fields are generally closer to science in earlier stages of the technology life cycle. Secondly, the *recency* in a field indicates the age of the field. Thirdly, *citation lag* indicates how old the knowledge used in the patent applications are. Thus, if a field is both more recent and has shorter citation lag than another, it is likely to be at an earlier stage in the TLC. Finally, the *relative growth* and the *relative assignee growth* indicate if the development expenses and interest in that field are growing or declining. High percentual increases indicate that a field is at an early stage in the TLC, whereas lower increases indicate that the field is more mature.

5

Findings and Analysis

This chapter presents the patinformatics research process applied to three case study technologies, lidar, radar, and sonar sensors. The chapter first gives a brief background to the technologies and their historical development. Further, the chapter is split into three parts in accordance with the HELD patinformatics framework for technology selection insights. First, a section explains how the pre-processing was performed and what patent sets ultimately came to represent the different technologies. Secondly, the three patent analysis themes from the HELD framework are applied to the task of patent analysis and the different metrics are evaluated. Finally, the insights gained from the analysis are communicated in the section about discovered knowledge.

5.1 Case Study Technologies

According to Altuntas et al. (2015), it is hard to quantitatively assess technologies in the growth stage of the TLC based on market data. This study has therefore used the three cases of lidar, radar, and sonar sensors for autonomous vehicles to test and evaluate the constructed patinformatics framework for technology selection. The following subsections will introduce the three technologies and thus give the basic understanding needed to conduct the case study.

5.1.1 Lidar

Light detection and ranging (lidar) sensors are commonly perceived as the key technology to enable full autonomous driving (Steinbaeck, 2017; McKinsey, 2017). The basic principle of the technology builds on using pulsed infrared light to measure distances that can be used to create three-dimensional models and maps of objects and environments in real-time (Schwarz, 2010). More specifically, lidar systems emit rapid pulses of laser light at a target, and based on the logic known as "time of flight" measurements, lidar calculates distances from the time it takes for the speed of light to hit an object or surface and reflect back (Himmelsbach, et al, 2008). Each measurement is constituted by a point in the resulting point-cloud which combines to create a virtual representation of the target object or area. Given the known and constant speed of light, distance calculations can be performed with high accuracy, and with the rich three-dimensional information contained in the returned point-cloud, autonomous machines are given enough amounts of data to sense their immediate environment and identify and react to obstacle and threats in sufficient time (Wang et al, 2017).

Laser ranging has been around since the early 1960s, where its high performance in mapping large-scale swaths of land was largely recognized. Thus, one of its most popular use cases has long been archeology (Wang et al, 2017). The concept of this sensing technology was first deployed in the automotive industry in 2005 when David Hall, the founder of Velodyne, entered the Defense Advanced Research Project Agency's (DARPA) grand challenge (a prize competition sponsored by the United States' Department of Defense to spur technological innovation in the field of autonomous vehicles) with a spinning lidar (Schwarz, 2010). Instead of using the typical single spinner laser that fires into a rotating mirror generating up to 200,000 points per second, David designed a system using a rotating unit encompassing 64 spinning lasers, each firing up to 20,000 times per second (Young, 2011). With its spinning unit and millions of data points generated each second, objects at distances over 100 meters and in 360 degrees field-of-view could be identified in both light and dark.

Since then, and especially during the DARPA Challenges in 2005 and 2007, lidar sensors have shown great potential for automotive applications (Wei et al, 2013). Many of today's prototype vehicles targeting fully autonomous driving are highly reliant on a spinning lidar sensor mounted on the top of the vehicle (Steinbaeck, 2017) supporting ADAS systems, such as lane change assist, autonomous emergency braking, blind spot detection, collision warning systems, cross-traffic alerts and adaptive cruise control. But due to its high cost and performance challenges in rain and fog, the research area remains open (Wang et al, 2016).

The need for advancements to make lidar systems adequate for automotive use and its many applications, has created an entire ecosystem containing many talented founders and teams engaged in the space. While the industry is marching ahead, the shared focus builds on the idea of *decreasing cost*, while *increasing range and resolution* (LeddarTech; LuminarTech; Velodyne LiDAR; Moebius, 2017).

5.1.2 Radar

Radar (radio detection and ranging) sensors use radio waves instead of light pulses to determine the angle, velocity, and range of objects (Steinbaeck, 2017). Electromagnetic waves are emitted towards an object, reflected, and returned to a receiver. The returned waves allow information to be given about the object's speed and location (Ragonese et al, 2009; US3604805A, 1971). Radar sensors were initially developed for military use before, and during, World War II, with the purpose of locating sea, ground and air targets (Militaryradar, 2018). Due to its possibility to directly measure the relative angular velocity of the detected targets (Doppler effect) with relatively little computational effort and its ability to work in bad weather conditions (Wei et al, 2013; Steinbaeck 2017), the technology has evolved into the automotive area and has become an established and fundamental component in assisted driving systems (Steinbaeck, 2017).

The large number of potential applications of automotive radar sensors are deployed in almost identical features as for lidar, including autonomous emergency braking, collision warning systems, cross-traffic alerts and adaptive cruise control. The application of radar to enable safety systems has increasingly become a standard equipment in even lower-cost vehicles (Ragonese et al, 2009), and according to Grand View Research (2017) radar accounts for more than 35% of the collision avoidance sensor market. Thanks to its potential use in a wide range of active safety features and its high probability of spreading downwards from what was once just used in high-end cars to being adopted in less expensive models the next coming year, scholars seem confident that the penetration and use of radar sensors in cars will grow at a substantial rate (Grand View Research, 2017; Steinbaeck, 2017; Lanchner, 2013).

Leading automotive electronics manufacturers such as Texas Instrument, Continental AG, and Robert Bosch GmbH are characterizing the radar sensor market today. But in spite of it being widely acknowledged as high performing, cost-effective and reliable, the sensors are not alone sufficient to provide a car with reliable environment information to navigate fully autonomous (Steinbaeck, 2017). In approaching the challenge of utilizing the full potential of radar, there are a four clear focus areas that seem to be prominent to improve the technology; *reducing the size and thereby the cost of radar systems, obtaining high resolution using higher frequencies* while *promoting the use of one single technology for all applications* in order to reduce the risk of mutual interference (Ragonese, 2009; Roselli 2005; Russell, 1997; Hasch 2012; Steinbaeck, 2015).

5.1.3 Sonar

Sonar, short for sound navigation and ranging, is a technique relying on sound waves to detect objects (Kim et al, 2005). In the same way as bats use sound for aerial navigation, sound-based echolocation was put in good use in various fields before the introduction of lidar and radar systems. Among other application areas, sound pulses with extremely high ultrasonic acoustic frequency were used to calculate the distance of an object by measuring the time of flight for reflected pulses (Carullo & Parvis, 2001; Kim et al, 2005).

Before the technology was recognized to have civilian uses, echo-ranging devices were initially developed to detect icebergs in 1906 and further developed during the World War I for military applications (Hill, 1962). Sonar systems are generally known to have a very poor range due to its short wavelengths, thus the sensor will not alone be able to figure out what pedestrians or other drivers are doing, or be able to respond to unexpected situations (Carullo & Parvis, 2001). But thanks to sound waves being comparatively slow, sonars are good for very near range three-dimensional mapping, as well as detecting differences of a centimeter or less and works regardless of light levels and conditions like rain, fog, or snow. The sonar sensor are thus being actively used in cars today for short-distance applications (usually below 10 meter distance) at driving speeds less than 10 km/hour (Park et al, 2008), primarily in the form of parking sensors as it has not yet proven to work at the type of ranges that an unmanned car demands (Kim et al, 2005).

Although their cars are not yet fully autonomous in all situations, Tesla has shown that sonar sensors, in combination with radars and camera sensors, can allow cars to drive themselves. With 12 ultrasonic sensors per sonar system, the system helps enable 360 degrees of visibility around the car allowing for detection of both hard and soft objects and any blind spot at up to 250 meters range (Tesla, 2018).

The performance of a sonar system's detection and localization depends primarily on two factors; the environment and the targeted object (Zhao, 2010). As ultrasonic signals are absorbed by the atmosphere, errors in the classification of surfaces may anticipate. The temperature, the air pressure, and the humidity may affect the accuracy of the measurements, and furthermore, the amplitude of the pulse can vary depending on the noise in the open-air condition and what type of surfaces it is reflected from (Carullo & Parvis, 2001). To reduce these errors, additional measurements given by sound attenuation correction as well as temperature and humidity sensors are often used to supply further information and support an adequate system (Bystrov et al, 2016).

The second factor highly affecting the correctness of a system relying on sound waves to visualize the environment is the shape, size, surface, and density of the object. Detection can be hindered in case of soft objects with strong sound absorption, when a flat object angled from the vertical deflects return sound waves away from the sensors, or if the object is insufficiently large to reflect sound (Zhao, 2010).

5.2 Pre-processing

This section explains how the patent sets for lidar, radar, and sonar were created, why this way was the best available, and what the resulting patent sets looked like.

5.2.1 Patent Retrieval Process

The main tool used for the patent retrieval process was the IP business intelligence platform Cipher from London-based Aistemos. Cipher Automotive's pre-classified patent sets, from the category "Autonomous and ADAS", the subcategory "ADAS components", and the three patent sets for "lidar sensors", "radar sensors" and "ultrasonic sensors" belonging to all automotive IP owners in Cipher's search engine. This search generated a list of patent numbers, publication numbers, and application numbers, along with some limited information that could be exported through the platform. To enable a study of the previously defined patent data metrics, the list of publication and application numbers was then used to create a sample database based on EPO's PATSTAT - 2017 autumn edition database.

The use of Cipher Automotive's pre-classified patent sets was motivated by the limited automotive knowledge of the researchers and the large amount of work already put into perfecting the Cipher patent sets by industry experts. Should the researchers instead have decided to create their own boolean search strings or manually tag documents, this would likely have resulted in smaller sets of patents with a poorer recall and many documents outside the automotive scope.

5.2.2 Retrieved Patent Data

After retrieving the patents from Cipher automotive, the resulting PATSTAT sample database contained data about a total of 12630 patent applications, of which the earliest were filed in 1862 and the most recent in 2017. The set covers a total of 132 distinct corporations, and since multinational automotive companies often have operations under several organizational entities, this amounts to 696 distinct company names.

As presented in chapter 4, patinformatics can yield the most insights about technological fields when the dynamic development of the different technologies over time is analyzed. With an 18-month publication lag and the 2017 autumn edition of the PATSTAT dataset containing data until the end of July 2017, some of the applications for 2016 and 2017 are missing, and the last full year for which the complete set of patent applications exists is 2015. This study has therefore focused on the 10-year window of 2006-2015, and all metrics have been evaluated based on patent applications filed in those years. Over the course of these years, the dataset contains 7029 distinct patent applications, of which 872 concern lidar sensors, 3124 concern radar sensors, and 3033 concern sonar sensors. The distribution of the data in year/technology cohorts can be seen in Table 5.1 below.

Table 5.1: Data distribution over technology fields and years

	2006	**2007**	**2008**	**2009**	**2010**	**2011**	**2012**	**2013**	**2014**	**2015**
Lidar	27	39	48	73	93	75	78	134	138	167
Radar	236	264	240	207	184	275	349	419	433	517
Sonar	236	211	279	202	264	344	403	405	363	326

As can be seen in the table, the lidar patent set contains much fewer patents than the radar and sonar sets. This lowers the statistical significance of the results from the lidar set analysis and creates larger year-to-year fluctuations in the metrics. On the total, a sample of 872 distinct lidar patent applications was still considered to be an adequate sample size, and the results have thus been trusted to be correct for the period as a whole.

5.3 Patent Analysis

This section presents the result of the patent analysis in three subsections according to the three themes outlined in the HELD framework from subsection 4.3.8. The underlying metrics for each theme are evaluated over the years of 2006-2015, and explanations for how the metrics were calculated can be found in subsubsection 4.3.6.1.

In some cases it proved meaningful to analyze and compare how the mean values for a metric differs between the three technological fields. In those cases, a one-way analysis of variance (ANOVA) test has been run. The ANOVA tests the null-hypothesis that the values observed for two or more groups were in fact drawn from groups with the same mean value (University of Kent, 2018). It is important to understand that the ANOVA test does not, however, say specifically which of the different means were statistically significantly different from each other. Showing a statistically significant difference was deemed adequate for this study, and p-values below 0.001 were considered to be statistically significant.

5.3.1 Theme 1: Technology Overview

In this subsection, the analyzed metrics for theme 1 are analyzed and visualized over the years of 2006-2015. This includes the cumulative number of patent applications, the cumulative number of distinct assignees, average patent scope, average patent centrality, average number of forward citations, average originality index, and average generality index.

5.3.1.1 Cumulative Number of Patent Applications

An initial overview of the three different fields can be gained from assessing the cumulative number of patent applications, as seen in Figure 5.1 below.

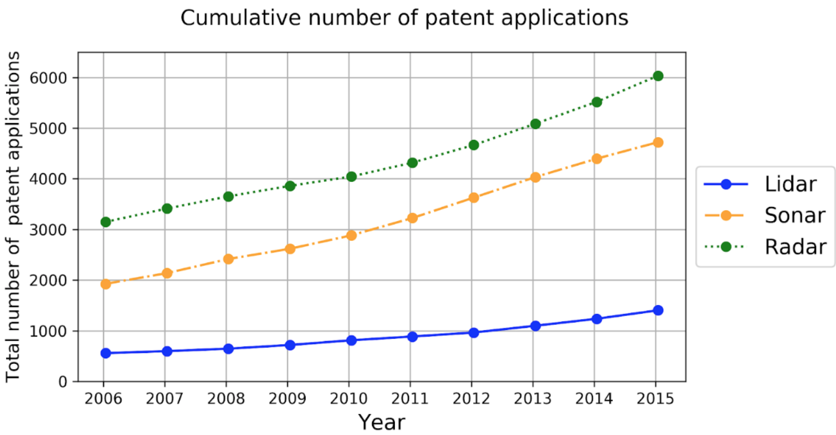

Figure 5.1: Cumulative number of patent applications

As can be seen in the figure, the largest effort has been put into radar technology, followed closely by sonar. The lidar technology contains less than a third of the patent applications as either that of radar or sonar, and all three fields have seen at least half of their total patenting activity occur in the past ten years. Since the number of filed patent applications is strongly correlated to the effort spent on R&D and inventing, this implies that much more resources have been spent on both radar and sonar sensors than on lidar sensors. It does not, however, say anything about the current investment levels in the different fields.

5.3.1.2 Cumulative Number of Assignees

The cumulative number of assignees has been calculated and is shown in Figure 5.2 below.

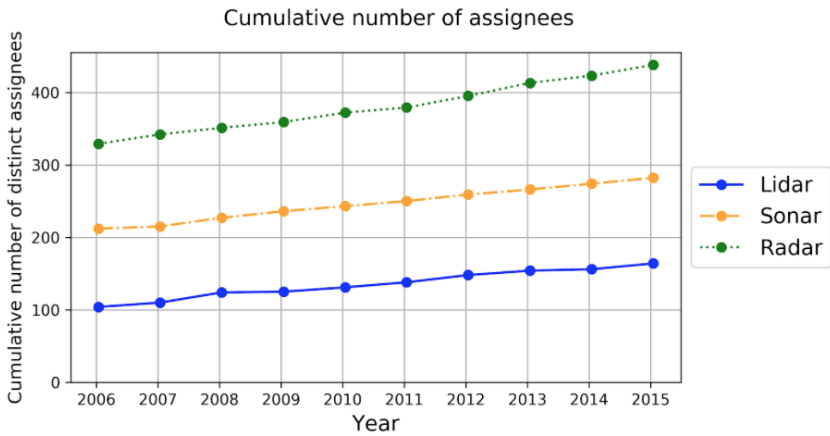

Figure 5.2: Cumulative number of assignees

The cumulative number of assignees follow a distribution similar to that of the cumulative number of patent applications. The growth rate of assignees is, however, lower than that of the number of patent applications. This indicates that not only is the number of assignees increasing but so is also the number of patent applications per assignee. Furthermore, the graph shows that many more companies have pursued inventions within radar than in sonar and lidar respectively.

5.3.1.3 Average Patent Scope

The third metric to be assessed in theme one is the average patent scope, meaning the average number of IPC classes per patent. The average patent scope is shown in Figure 5.3 below.

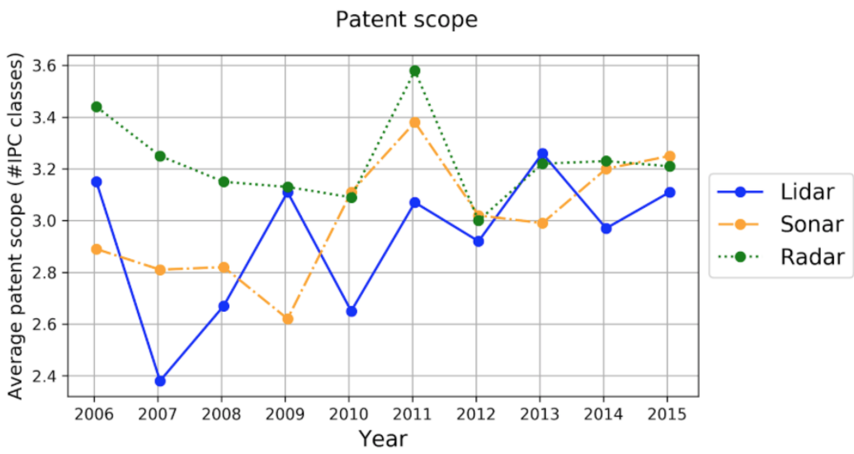

Figure 5.3: Average patent scope

While it may be hard to discern a clear trend from this image, the average patent scope for the whole time period is 2.99 for lidar, 3.04 for sonar, and 3.23 for radar. The difference in mean values is statistically significant ($p < 0.001$). This shows that patented radar inventions are somewhat less niched than lidar and sonar inventions. Although statistically significant, the difference is relatively small.

5.3.1.4 Average Patent Centrality

In accordance with the SPNP ranking process outlined in section 4.3.2, patent centrality has been assessed by computing SPNP values for all patent applications and then calculating percentile scores for those values in year-based cohorts. For instance, if the max SPNP score was 450 for a patent in 2008, then patents from 2008 with an SPNP value of 450 were assigned a score in the 100th percentile. The dynamic development of this metric over time can be seen in Figure 5.4 below.

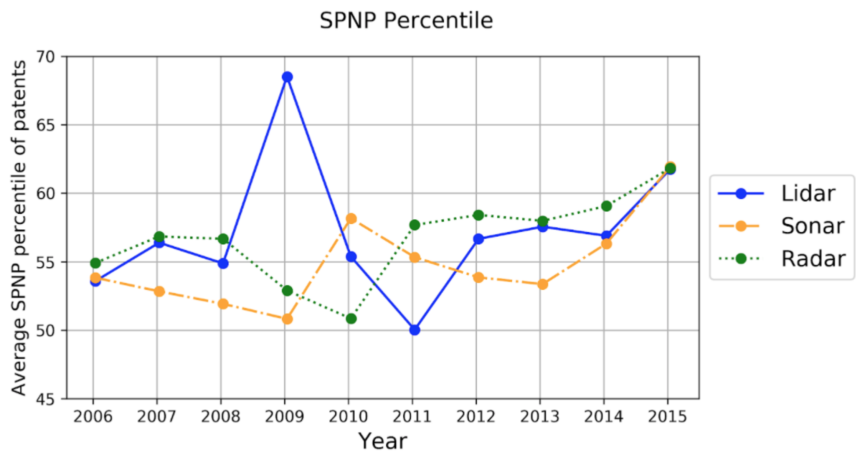

Figure 5.4: Average SPNP percentile score

Two things should be noted from the figure above. First, the SPNP percentile for lidar shows a big peak in 2009. When analyzing the data, the reason for this becomes clear; 20 separate patent applications from GM all cite each other. With the SPNP values being calculated based on all patents in the backward and forward citation chains, all those patents get the same SPNP value in the 99th percentile. With only 73 lidar patent applications for 2009, this is enough to skew the data. Secondly, all values seem to converge in 2015. This again is due to the design of the SPNP metric, where newly filed patent applications that have not yet received any forward citations all have approximately the same SPNP value. Despite these two shortcomings it is possible to calculate meaningful averages for all three technologies over the period. The average lidar patent was scored in percentile 57.8, while the average radar patent was scored in the percentile 58.7 and sonar patents averaged in percentile 56.2. The difference in mean values is statistically significant ($p < 0.001$). This indicates that the fields of radar and lidar may be developing at faster technology improvement rates (TIRs) than sonar.

5.3.1.5 Average number of Forward Citations

There are many different ways to calculate forward citations. In this study, the citations have been computed on a family-to-family basis, meaning that the number of distinct patent families citing each patent family in the database has been calculated, and the average for each technology/year cohort has been computed. The resulting graph can be seen in Figure 5.5 below.

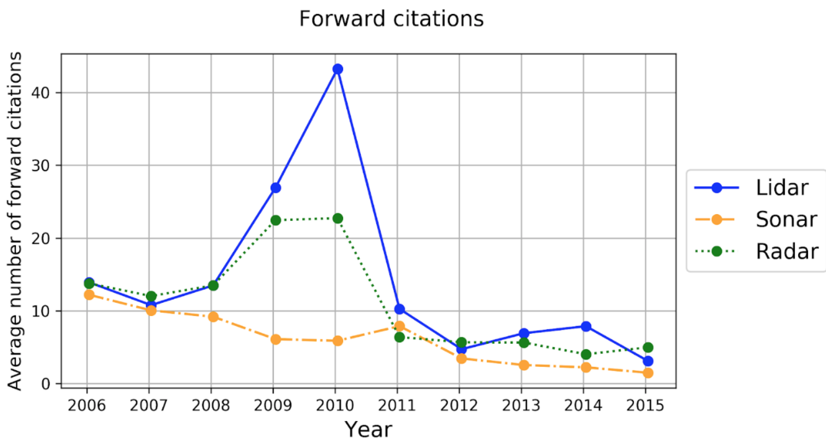

Figure 5.5: Average number of forward citations

As expected, older patents have more forward citations than newer patents. Another interesting point is that both lidar and sonar patents from 2009 and 2010 experience large surges in their numbers of forward citations, indicating that the inventions in these fields and years have been very important for the technical progress of their respective domains. Over the ten year period, the average lidar patent has gotten cited 15 times, the average radar patent has gotten cited 11 times and the average sonar patent has gotten cited 7 times. The difference in mean values is statistically significant ($p < 0.001$). This is a clear indication that lidar patents have larger technological impacts than their radar and sonar counterparts and also that their economic value is higher.

5.3.1.6 Average Generality Index

The generality index has been calculated and can be seen in Figure 5.6 below.

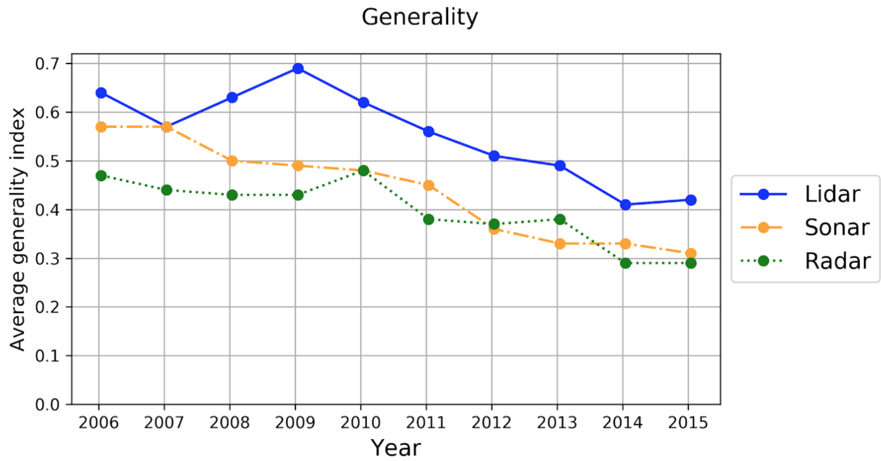

Figure 5.6: Average generality index

As seen in the figure, lidar patents are more general than sonar and radar patents, meaning that they get cited by other patents from a larger range of IPC classes. The mean value for lidar is 0.55, for sonar 0.46, and for radar 0.41. The difference in mean values is statistically significant ($p < 0.001$). This indicates that the knowledge contained in Lidar patents gets more generally applied than that in sonar and lidar patents and that the expansion potential for lidar inventions may consequently be larger. As expected, the metric drops with age due to its dependency on forward citations.

5.3.1.7 Average Originality Index

The originality index has been calculated and can be seen in Figure 5.7 below.

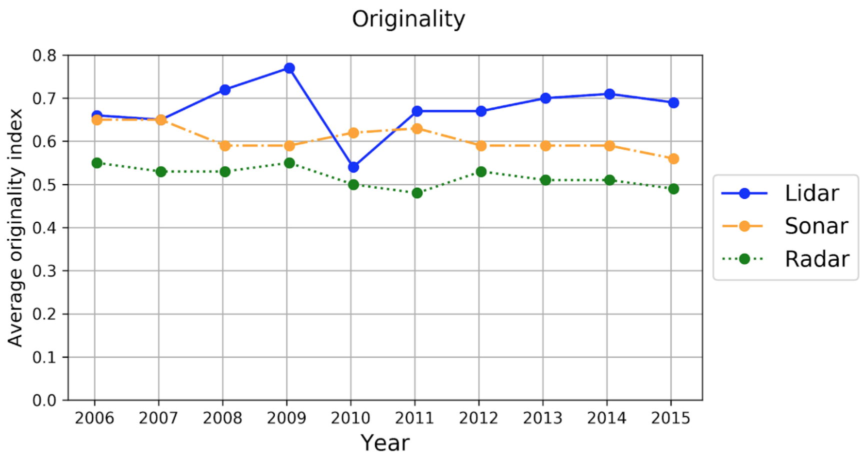

Figure 5.7: Average originality index

With the only exception of 2010, Figure 5.7 gives a clear and consistent picture of the fields' originality indexes. With an average originality index of 0.67, lidar patents are most original, followed by sonar patents (0.60) and radar patents (0.51). The difference in mean values is statistically significant ($p < 0.001$). This means that lidar patents cite other patents from a larger number of IPC classes, and indicates that a more diverse knowledge might be needed to carry out inventions in the field of lidar, whereas sonar and radar require more niched knowledge.

5.3.2 Theme 2: Investment Volume

In this subsection, the analyzed metrics for theme 2 are analyzed and visualized over the years of 2006-2015. This includes the yearly number of patent applications, the yearly number of assignees, average renewal time, average grant lag, and average patent family sizes.

5.3.2.1 Yearly Number of Patent Applications

The yearly number of patent applications for the three technology fields can be seen in Figure 5.8 below.

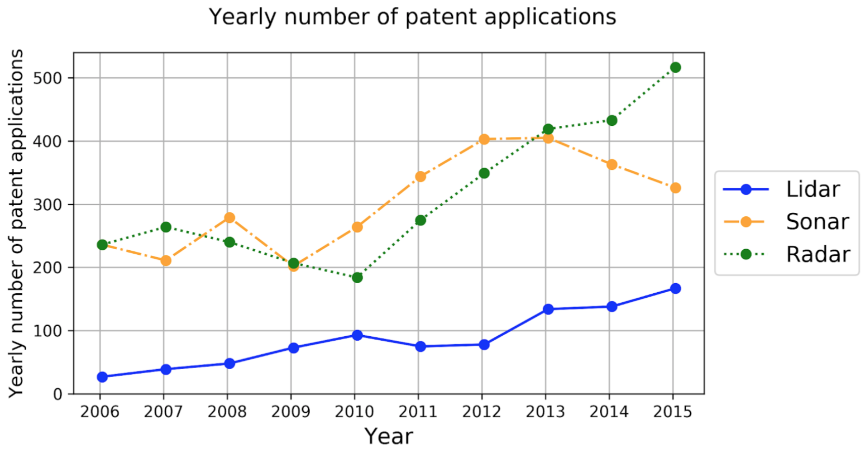

Figure 5.8: Yearly number of patent applications

The field of lidar has been growing the slowest over the
ten-year period with an average of 87 patent applications per year, followed by sonar at 303 and radar at 312. It is, however, worth noting that the yearly number of radar and lidar patent applications have been increasing steadily over the past few years, whereas sonar applications maxed out in 2013 and have been decreasing since. In total, this implies that radar technology attracted most investments as of 2015, followed by sonar and radar, but also that the investments in radar and lidar are growing while those in sonar are decreasing. If the trend from the years 2012-2015 continues linearly, lidar applications will outnumber sonar applications in 2018.

5.3.2.2 Yearly Number of Assignees

The yearly number of assignee figures were calculated based on how many distinct assignees filed patents in each technology/year cohort, and are shown in Figure 5.9 below.

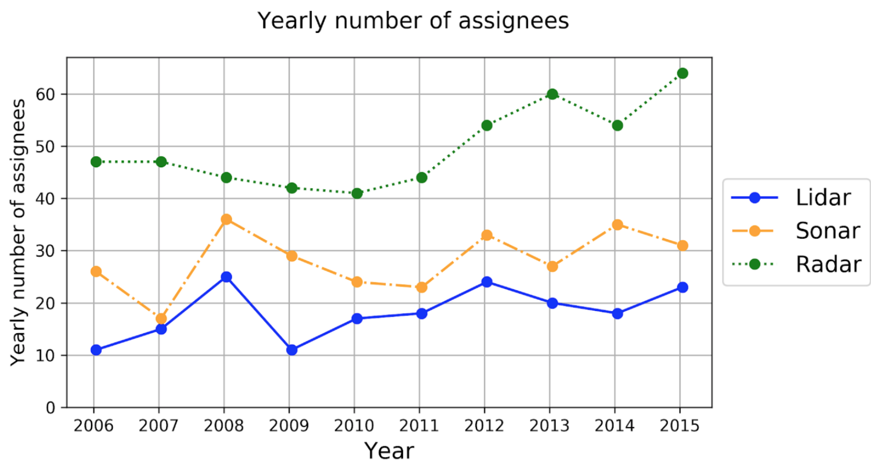

Figure 5.9: Yearly number of assignee

Among the three technologies, radar attracts interest from the largest number of assignees, followed by sonar and lidar. The number of assignees filing sonar and lidar patents has been quite constant over the last 6 years, whereas interest in radar technology has been growing.

5.3.2.3 Average Renewal Time

The average renewal time has been calculated as the number of years that an average patent from each technology/year cohort has been kept alive and renewed. The results are shown in Figure 5.10 below.

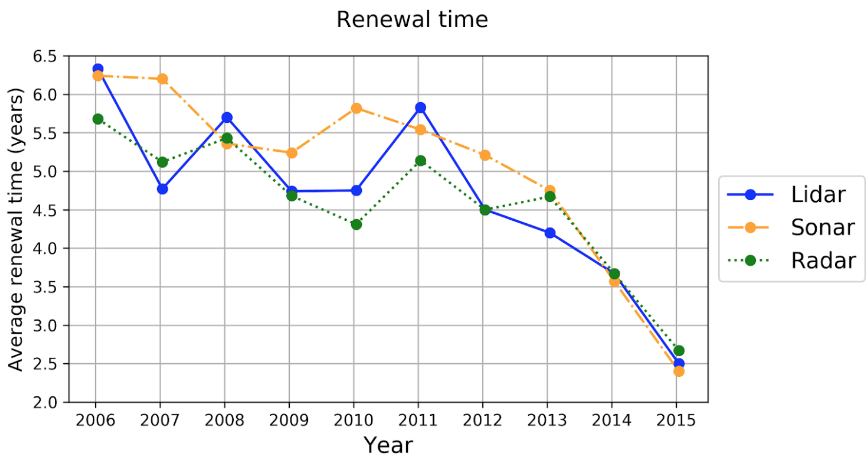

Figure 5.10: Average renewal time

It is hard to discern a general pattern from Figure 5.10 and the meaning of averages is distorted when performing dynamic analysis about time-dependent variables like the case here. It seems as though renewal time is not a very informative metric for the current analysis, and maybe there are better metrics to analyze and visualize companies' propensity to renew their patents.

5.3.2.4 Average Grant Lag

The average time between application and first grant dates has been calculated and is presented in Figure 5.11 below.

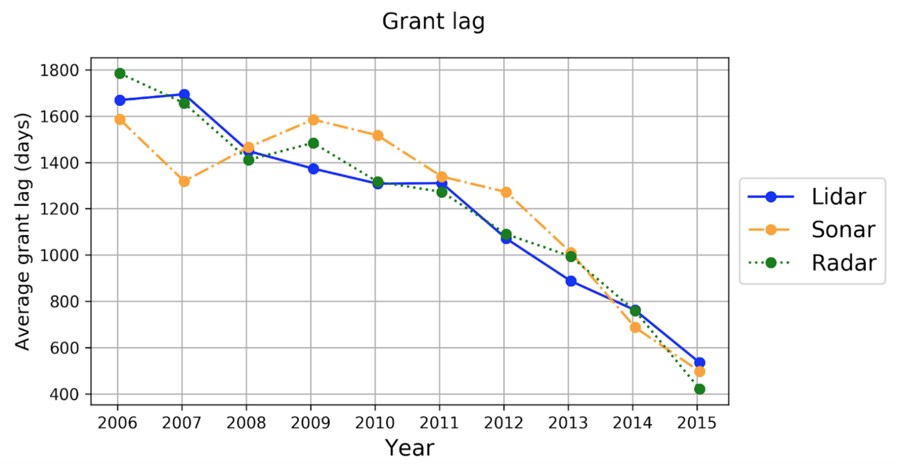

Figure 5.11: Average grant lag

Sonar has had the longest grant lags in five of the last seven years, which indicates that companies may not be putting as much resources into speeding up the prosecution of sonar patents as they do for lidar and radar. Since the grant lag is only calculated for patents that have been eventually granted, patent applications from newer years have lower grant lags than older applications. No average values have therefore been calculated for this metric.

5.3.2.5 Average Patent Family Size

The average geographic patent family size has been calculated and can be seen in Figure 5.12 below.

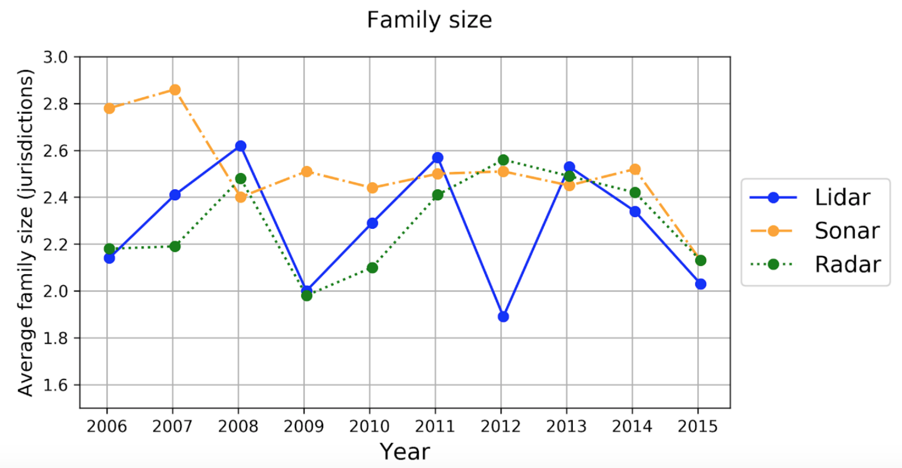

Figure 5.12: Average geographic patent family size

This metric suffers from large year to year fluctuations and it's hard to discern a general pattern.Numbers for 2015 are misrepresentative since not all applications that claim 2015 patent applications as priority documents have been published as of the creation of this database. An average radar family consists of 3.04 patents, a lidar family consists of 3.13 patents and a sonar family consists of 3.18 patents.

These results are however not statistically significant with a $p < 0.001$, and patent family size was therefore not considered an informative feature for the purpose of this analysis.

5.3.3 Theme 3: Technology Life Cycle

In this subsection, the analyzed metrics for theme 3 are analyzed and visualized over the years of 2006-2015. This includes the average science intensity, recency, average citation lag, relative patent growth and relative assignee growth.

5.3.3.1 Average Science Intensity

Science intensity has been calculated as the average number of NPL citations per patent application and is presented in Figure 5.13 below.

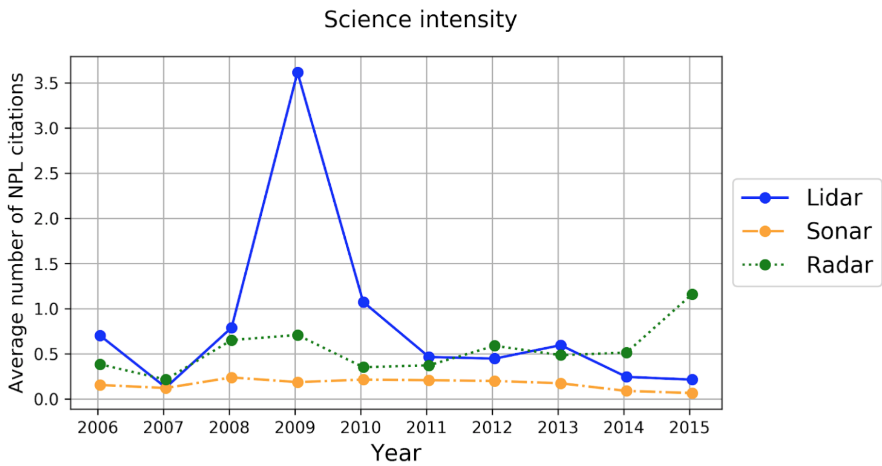

Figure 5.13: Science intensity

Two distinct anomalies should be pointed out from the figure. First, the science intensity for lidar experiences a strong peak in the year 2009, which is again due to the 20 GM patent applications that all cite the same 12 NPL documents. Secondly, the sudden peak for radar in 2015 is due to one single US patent (US9643605) that alone cites 502 different NPL publications. With those two years exempt, the average science intensity over the period is 0.55 for lidar, 0.46 for radar, and 0.18 for sonar. The difference in mean values is statistically significant ($p < 0.001$). This difference implies that sonar is further away from basic research, and thus at a later stage in the TLC than radar and lidar.

5.3.3.2 Recency

The recency of all fields has been calculated as the average age of all documents in the sets. For this figure to give an accurate representation of the whole field, all patent applications in the set have been used, and the analysis has thus not been limited to the years between 2006 and 2015.

Lidar is the most recent of the fields, with an average patent application being filed in January 2008. An average sonar patent application was filed in November 2006, while an average radar patent was filed in June 2005. The difference in mean values is statistically significant ($p < 0.001$). Lidar is thus the most recent field, which implies that it may also be earliest in the technology life cycle.

5.3.3.3 Average Citation Lag

Citation lag has been calculated as the average age of cited patent publications, at the time of the citing patents' publication. The results are shown in Figure 5.14 below.

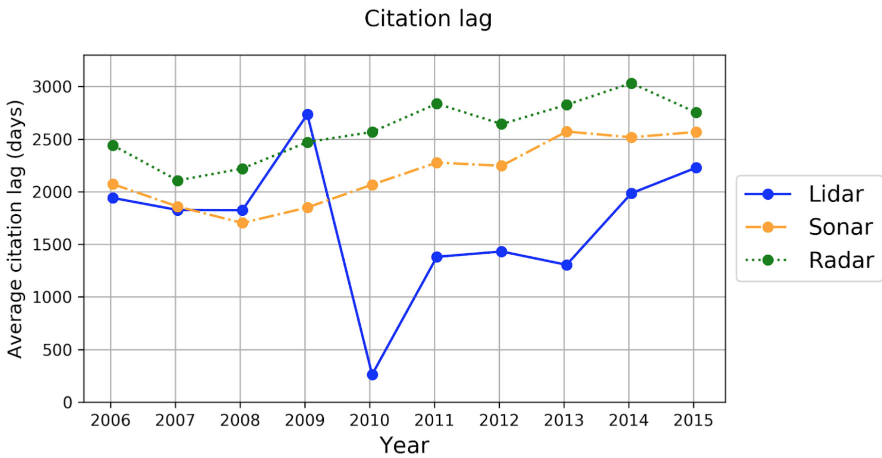

Figure 5.14: Citation lag

Once again, the values for lidar patents are skewed by GM's large patent family that was applied for in 2009 and published in 2010. When excluding patent publications from 2010, lidar patents have an average citation lag of 1212 days, sonar patents have an average citation lag of 2294 days, and radar patents have an average citation lag of 2632 days. This difference is statistically significant ($p < 0.001$). This means that the field of lidar builds on more recent knowledge and thus might be earlier in the TLC. It goes well in line with the fact that the lidar field is more recent, and citations to other documents within the field are therefore also likely to be more recent. Similarly, radar was identified as the oldest (least recent) field, and it also has the highest citation lag.

5.3.3.4 Relative Patent Growth

Relative patent growth has been calculated as the number of yearly patent applications divided by the total number of patent applications in a field and gives an indication of how much effort is being put into development in a field relative to its size. The results are shown in Figure 5.15 below.

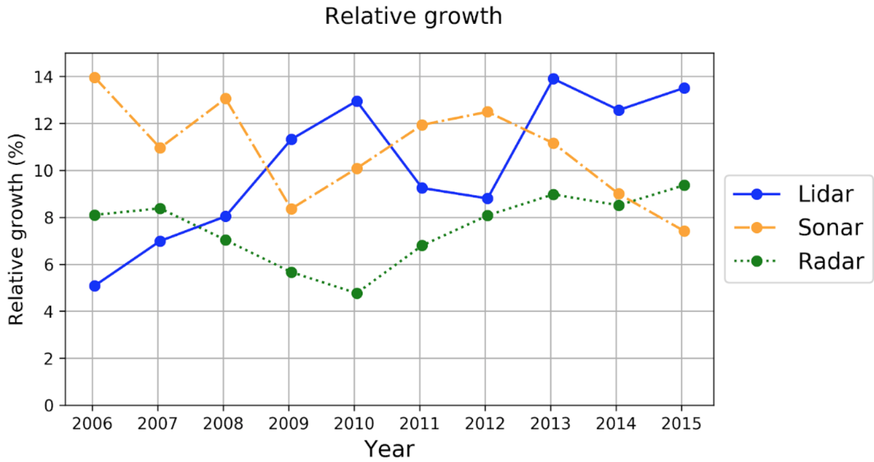

Figure 5.15: Relative patent growth

Except for a temporary decline in 2011-2012, the relative growth of lidar patent applications has increased steadily over the past ten years, with a current yearly growth of around 14%. Thus, the field is not only growing, but it is growing at an increasing rate compared to its existing knowledge base. The same goes for the field of radar over the past six years, though its growth rate is somewhat smaller at around 9%. The field of sonar is growing by about 8% per year, but this speed has been decreasing over the past four years. The decline in relative growth in the sonar field implies that it is at a later stage in the TLC, where companies are cutting back their development efforts. Similarly, high and increasing relative growth rates for lidar and radar implies that these fields are earlier in the TLC and companies are increasing their inventive activity within these fields.

5.3.3.5 Relative Assignee Growth

Similarly as relative patent growth, relative assignee growth has been calculated as the number of new assignees in a field divided by the total number of distinct assignees in a field and gives an indication of how many new entrants a field is attracting, relative to its size. The results are shown in Figure 5.16 below.

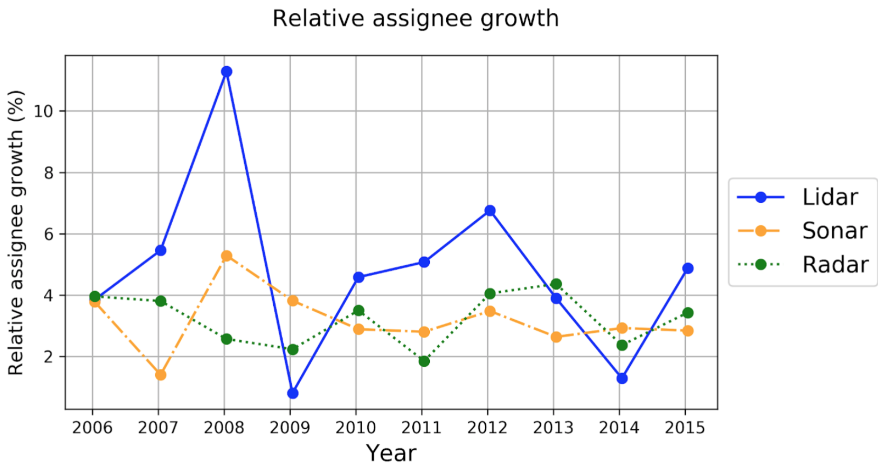

Figure 5.16: Relative assignee growth

Lidar is the smallest field and has attracted the largest relative share of new entrants for six out of the past ten years, with an average relative assignee growth of 4.7%. This rush of new entrants into the field indicates that it is earlier in the TLC than sonar and radar, which are both experiencing relative assignee growths of 3.2% per year.

5.3.4 Discovered Knowledge

Having analyzed all patent data metrics that could give insights, the results need to be interpreted and the most important insights communicated to decision makers. This subsection thus aims to show an example of how they could be communicated efficiently.

5.3.4.1 Discovered knowledge from Theme 1: Technology Overview

For this analysis, relative patent growth, relative assignee growth, number of citations, originality index and generality index proved to be informative features, while it was hard to draw any conclusive insights from the patent scope and patent centrality. The insights from theme 1 can thus be concluded as shown in Figure 5.17 below.

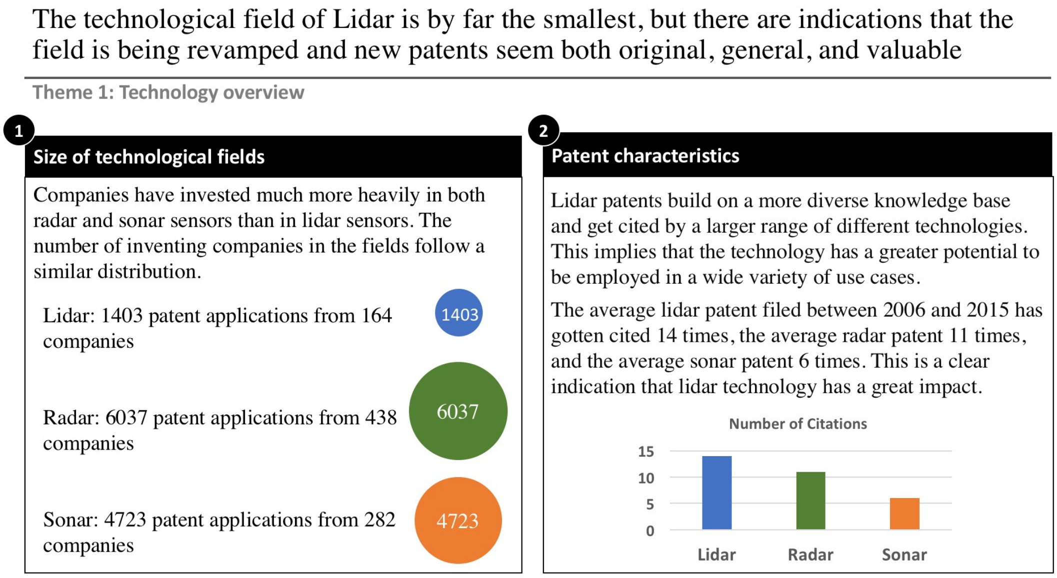

Figure 5.17: Discovered knowledge from theme 1

5.3.4.2 Discovered knowledge from Theme 2: Investment Volume

For this analysis, yearly numbers of patent applications, yearly numbers of assignees and grant lag proved to be informative features, while it was hard to draw any conclusive insights from the average renewal time and average patent family size. The insights from theme 2 can thus be concluded as shown in Figure 5.18 below.

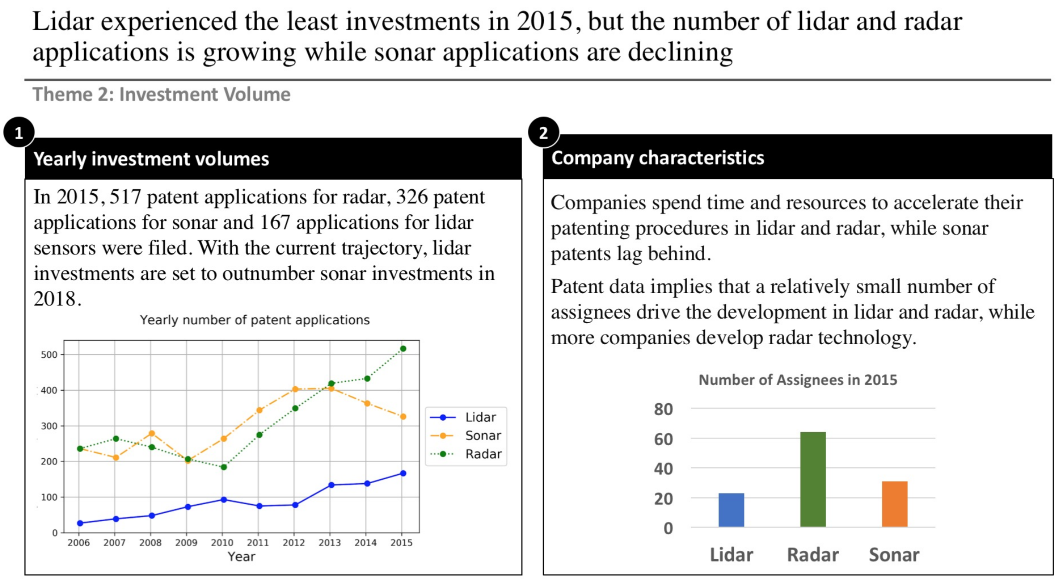

Figure 5.18: Discovered knowledge from theme 2

5.3.4.3 Discovered knowledge from Theme 3: Technology Life Cycle

For this analysis, recency, science intensity, citation lag, relative patent growth, and relative assignee growth all proved to be informative features. In total, the metrics for theme 3 give a good overview of the TLC situation for each one of the analyzed technologies. The insights from theme 3 can be concluded as shown in Figure 5.19 below.

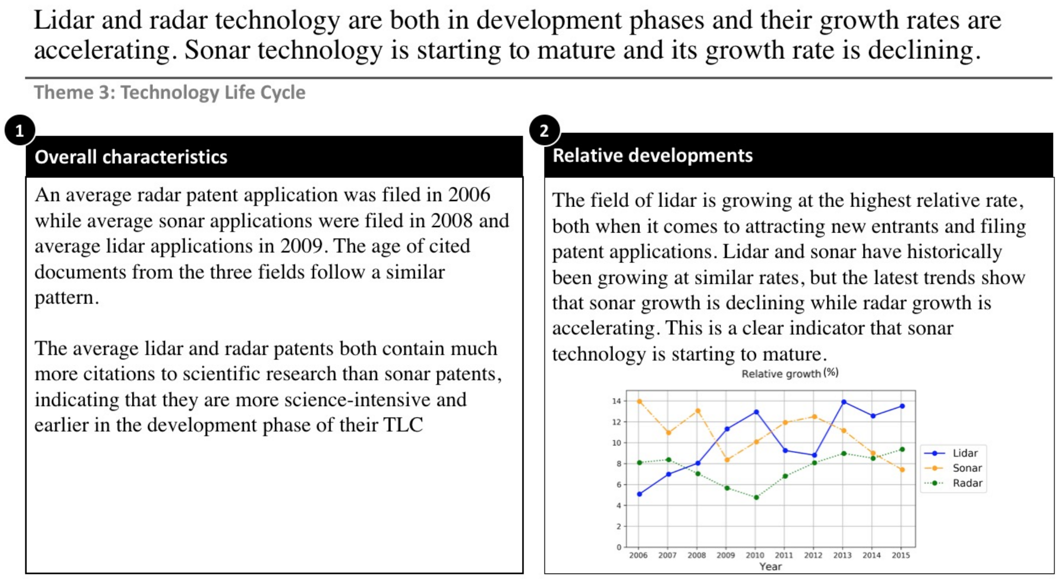

Figure 5.19: Discovered knowledge from theme 3

6
Conclusion

The overarching aim of this study was to find out if, and how, patent data can be used to guide technology selection decisions in rapidly changing technology-environments. This chapter revisits the research questions that helped reach the purpose and explains the conclusions that have been reached throughout the process, both regarding the suitability of patent data as an information source, how a structured process of analyzing it can be constructed, and what results this process yielded when applied to a case study of three technological fields. To recall, the main research question was: *How can a patinformatics framework be constructed to help technology managers gain valuable insights for selecting which technologies to invest in?*

To answer this, four sub research questions were answered (presented here in logical order):

SRQ1: What information can be found in patent data?

Information found in patents encompasses a wide number of areas, including technical, legal, business-relevant or public-policy. This includes information that is either written on the patent application itself, which we refer to primary patent data, or information that can be found in secondary patent data features that amasses over the lifespan of a patent. Patent databases are the world's largest repository of technological information, and patent data is well-structured and pre-classified into technological fields. Patent applications are a direct result of an inventive process and capture the commercial and proprietary aspects of technological development, and since they are granted by authorities, it is also an objective information source for measuring companies' inventive activities. Previous researchers have successfully used patent data to predict the value of companies, assess success levels of technological fields, and predict the rate of technological progress within different fields.

SRQ3: How can a patinformatics framework be constructed for use in technology selection?

We have identified three main tasks that are crucial for using patinformatics to gain technology selection insights.The first task, *pre-processing*, concerns the patent retrieval process. In particular, the aim of the task is to create patent sets that represent the technology being studied with high completeness, relevance, and replicability. Traditionally, this has been done using boolean searches, but newer methods include the hybrid-keyword-classification algorithm and search engines based on supervised machine learning. Overall, the supervised machine learning method seems most promising due to its ability to achieve high levels of completeness and recall while still being explainable and relatively easy to use. Once the researcher has retrieved good patent sets for the technology or technologies being studied, he or she can start the actual *patent analysis*. In this task, the researcher chooses which aggregated metrics should be computed and performs the calculations over a relevant period of time. In case of a comparative analysis where different technologies are analyzed, the results can be visualized next to each other in graphs or tables to show the relevant levels and their dynamic trajectory. If a single technology is analyzed, it might instead be fruitful to normalize the values using year-based cohorts and thus show how the technology is developing compared to other patented technologies as a whole. Finally, the *discovered knowledge* needs to be communicated in a clear and concise manner to technology selection decision-makers. This should include a purposeful revisitation of the results to extract the most valuable insights and scrap the metrics that proved uninformative. The informative metrics should then be visualized and described using available communication tools, such as slide decks or videos.

Building on these insights, the HELD patinformatics framework for technology selection insights was created, and the main steps are shown in figure 6.1 below.

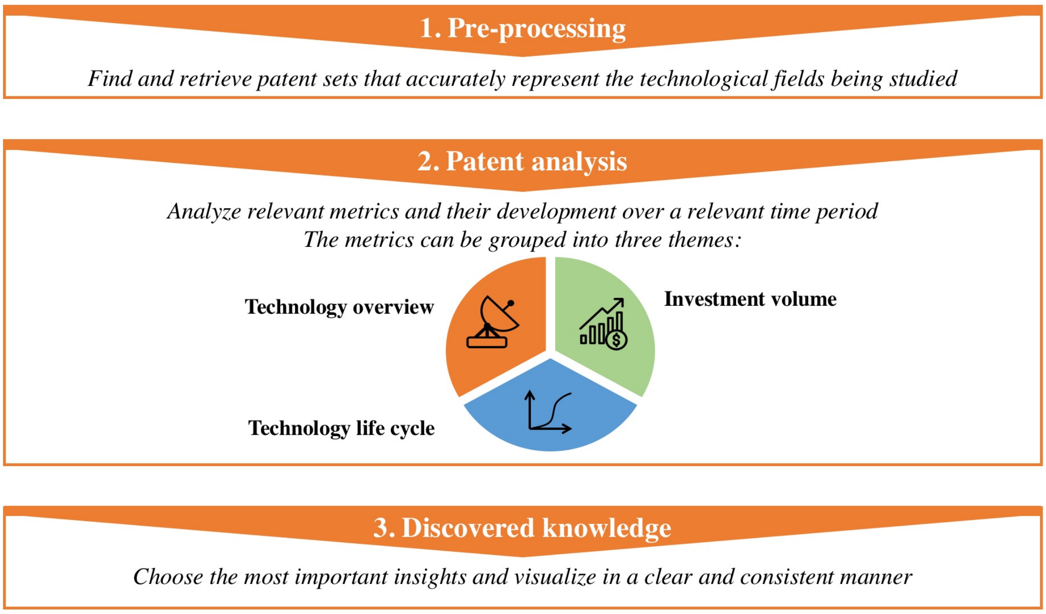

Figure 6.1: HELD patinformatics framework for technology selection insights

SRQ2: What patent data metrics can be used to reach insights about technologies?

The HELD framework groups patent metrics into three broad themes of insights. Theme one, technology overview, includes metrics that are helpful to give a first overview of the technological field. This includes the total effort spent on inventive activities, the technology improvement rate, and a number of indicators concerning the technical aspects of patents, such as their technological breadth, diffusion, and impact. Theme two, investment volume, includes metrics that measure how much resources are being spent on inventive activities within the field. This includes the yearly number of patent applications, the interest in the field from different actors, and indicators of the patents' economic value. Finally, theme three, technology life cycle, contains indicators about the current maturity level of the technology. This includes the patent growth relative to previous years, the number of new entrants, and the science intensity of the technological field.

SRQ 4: What insights can the constructed framework give about the technological fields of lidar, radar, and sonar?

The framework was applied in a case study and many insights were gained about the technological fields of lidar, radar, and sonar for autonomous vehicles. Lidar is the youngest technology, with by far the least historical effort spent on inventing activities. The field of lidar technology has, however, been growing at an increasing rate and a steady flow of new entrants are attracted yearly. Lidar patents are also the broadest, most generally diffused, and most technically impactful of the analyzed patent sets. Lidar thus seems to be at an early stage of its technology life cycle, and many companies seem convinced about its future importance. Radar is the largest field, with the most interested companies and the oldest patents. However, like lidar, radar is also experiencing a surge in the level of inventive activity, attracting increasingly larger numbers of yearly patent applications and assignees. Radar can thus be summarized as a large technology field still in the growth stage of its technology life cycle. Finally, sonar was the second largest field, covering three times as many patent applications as lidar and still growing at a faster yearly rate. However, metrics indicate that sonar patents cite almost no scientific literature and that interest and investments in the technology are starting to decline. Other indicators still indicate that companies are willing to invest resources in keeping their sonar patents alive. In total, this describes a patent field in the maturity stage of the technology life cycle, where the need for basic research and development activities are starting to decline, and companies are now increasingly focused on commercializing their existing assets.

7 Discussion

This chapter reflects on the theoretical and practical implications of the study, addresses the limitations of the research and provides suggestions for future research.

This study was based on the practical problem of positioning companies' technology bases in increasingly changing technology-environments. Patent data is currently underutilized as a source of information to guide these decisions, much due to the thorough understanding of the field of patents needed to conduct analyses. Even for those that know and understand patents, there's a long way to go before recognizing what insights can be gained from studying different patent metrics. Our hope is that the findings from this study help change this through degunkifying and disambiguating the existing techniques for using patinformatics in technology selection and outlining existing metrics in a clear and understandable manner. We hope that the study can help emphasize the value of patent information and bring the topic of patinformatics into more companies, more management meetings, and more technology selection decisions.

This study was performed as a part of a master thesis project, and the research findings are limited by a number of factors. First, the constructed framework was based solely on patent data metrics. It is important not to think of the framework as a stand-alone method for technology selection but rather as a way to gain insights about technological developments. Secondly, the constructed framework was tested on three different sensor technologies for autonomous vehicles. Since no dominant design has yet emerged, it is impossible to say conclusively whether or not the insights gained from the analysis would have helped technology managers to make the right decisions. Simultaneously, the use of such technologies can be considered a strength of the study, since there was no way to fit the analysis results and insights to a known outcome. The choice of solely using patent data in the case study also limited the range of conclusions that could be made. Some metrics were hard to interpret, and it is quite possible that more information could have been gained by "connecting the dots" between the patent data and other data, such as known market conditions and investment figures. Furthermore, the three fields might have different propensity to patent, especially if some rely more on software than the others. An extensive technology mapping of the fields could have helped identify such problems but was outside the scope of the study. Finally, the patent sets used to represent the three technologies were based on Cipher Automotive's pre-performed patent grouping, and any flaws inherent in this grouping will have impacted the analysis. For instance, the patent sets include data from known automotive companies and their first-tier suppliers, which can be misleading if one of the technologies is unproportionally being developed by other companies, such as start-ups. The patent sets were originally created for both advanced driving assistance systems (ADAS) and autonomous driving (AD) technologies, which means that they represent a larger technology than just autonomous driving, and the inclusion of ADAS might have skewed the analysis. At the same time, ADAS and AD are technologically inseparable since one is just a more complete application of the technologies than the other. We mean that analyzing multi-purpose technologies for a subset of their use case is difficult and that this should be considered an inherent problem in patent analysis.

Throughout the course of this project, a number of interesting topics for further research have been uncovered. For one, the proposed framework could be applied to other technological fields. It would be interesting to test the metrics on technology battles with known outcome and see if there exist trends and similarities between the dynamic development in the technological fields which ultimately proved successful. If such similarities exist and it is possible to create formal definitions of what characterizes technology success, this data could be used as training examples to train a supervised machine learning model to predict successful technologies based on their patenting development. Another thing that could help extend the value of this study would be to integrate the patinformatics framework with other data to create a more complete framework for technology selection. By assessing not only the patent characteristics of the technologies but also their fit into the company's product, strategies, and capabilities, a more general process for making technology selection decisions could be defined.

Bibliography

[1] Altuntas, S., Dereli, T. and Kusiak, A. (2015). Forecasting technology success based on patent data. Technological Forecasting and Social Change, 96, pp.202-214.

[2] Archibugi, D. and Planta, M. (1996). Measuring technological change through patents and innovation surveys. Technovation, 16(9), pp.451-519.

[3] Aristodemou, A., Tietze, F., Athanassopoulou, N. and Minshall, T. (2017). Exploring the Future of Patent Analytics: A Technology Roadmapping approach. Centre for Technology Management working paper series, No. 5.

[4] Asche, G. (2017). "80% of technical information found only in patents" – Is there proof of this?. World Patent Information, 48, pp.16-28.

[5] Basberg, B. (1987). Patents and the measurement of technological change: A survey of the literature. Research Policy, 16(2-4), pp.131-141.

[6] Bass, S. and Kurgan, L. (2009). Discovery of factors influencing patent value based on machine learning in patents in the field of nanotechnology. Scientometrics, 82(2), pp.217-241.

[7] Benson, C. and Magee, C. (2012). A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field. Scientometrics, 96(1), pp.69-82.

[8] Benson, C. and Magee, C. (2014). On improvement rates for renewable energy technologies: Solar PV, wind turbines, capacitors, and batteries. Renewable Energy, 68, pp.745-751.

[9] Benson, C. and Magee, C. (2014). Technology structural implications from the extension of a patent search method. Scientometrics, 102(3), pp.1965-1985.

[10] Benson, C. and Magee, C. (2015). Quantitative Determination of Technological Improvement from Patent Data. PLOS ONE, 10(4), p.e0121635.

[11] Bonino, D., Ciaramella, A. and Corno, F. (2010). Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent

informatics. World Patent Information, 32(1), pp.30-38.

[12] Branstetter, L. (2005). Exploring the Link Between Academic Science and Industrial Innovation. Annales d'Économie et de Statistique, (79/80), p.119.

[13] Bryman, A. and Bell, E. (2011). Business Research methods. Oxford: Oxford University Press.

[14] Butler, F. and Martin, J. (2016). The auto industry: adapt to disruptive innovations or risk extinction. Strategic Direction, 32(11), pp.31-34.

[15] Bystrov, A., Hoare, E., Tran, T., Clarke, N., Gashinova, M. and Cherniakov, M. (2016). Road Surface Classification Using Automotive Ultrasonic Sensor. Procedia Engineering, 168, pp.19-22.

[16] Callaert, J., Van Looy, B., Verbeek, A., Debackere, K. and Thijs, B. (2006). Traces of Prior Art: An Analysis of Non-Patent References Found in Patent Documents. SSRN Electronic Journal.

[17] Carullo, A. and Parvis, M. (2001). An ultrasonic sensor for distance measurement in automotive applications. IEEE Sensors Journal, 1(2), p.143.

[18] CIPHER AUTOMOTIVE (2018). Harness AI to navigate complex technology landscapes. [online] Cipher. Available at: http://cipher.ai/solutions/cipher-automotive/ [Accessed 5 May 2018].

[19] Coffman K.G., Odlyzko A.M. (2002) Internet Growth: Is There a "Moore's Law" for Data Traffic?. In: Abello J., Pardalos P.M., Resende M.G.C. (eds) Handbook of Massive Data Sets. Massive Computing, vol 4. Springer, Boston, MA

[20] De Solla Price, D. (1965). The pattern of bibliographic references indicates the nature of the scientific research front. American Association for the Advancement of Science. Science, New Series, Vol. 149, No. 3683.

[21] Eisenhardt, K.M. (1989). Building Theories from Case Study Research. Academy of Management Review, 14: 532-50.

[22] Ernst, H. (2003). Patent information for strategic technology management. World Patent Information, 25(3), pp.233-242.

[23] European Patent Office (2018). Espacenet patent search. [online] Epo.org. Available at: https://www.epo.org/searching-for-patents/technical/espacenet.htmltab-1 [Accessed 14 May 2018].

[24] European Patent Office. (2018). F-IV, 3.8 Independent claims containing a reference to another claim or to features from a claim of another category - Guide-

lines for Examination. [online] Epo.org. Available at: https://www.epo.org/law-practice/legal-texts/html/guidelines/e/$\mathrm{f}_i v_{38}.htm[Accessed 18 Apr. 2018]$.

[25] Farrukh, C., Phaal, R. and Probert, D. (2003). Technology roadmapping: linking technology resources into business planning. International Journal of Technology Management, 26(1), p.2.

[26] Floyd, C. (1997). Managing Technology for Corporate Success. Gower Publishing Limited.

[27] Gao, L., Porter, A., Wang, J., Fang, S., Zhang, X., Ma, T., Wang, W. and Huang, L. (2013). Technology life cycle analysis method based on patent documents.Technological Forecasting and Social Change, 80(3), pp.398-407.

[28] Gao, L., Porter, A., Wang, J., Fang, S., Zhang, X., Ma, T., Wang, W. and Huang, L. (2013). Technology life cycle analysis method based on patent documents. Technological Forecasting and Social Change, 80(3), pp.398-407. Grandviewresearch.com. (2018). Automotive Radar Market Worth $12.16 Billion By 2025 | CAGR: 20.8%. [online] Available at: https://www.grandviewresearch.com/press-release/global-automotive-radar-market [Accessed 11 May 2018].

[29] Gregory, M. (1995). Technology management : a process approach. Journal of Engineering Manufacture. Hall, B. (2005). A Note on the Bias in Herfindahl-Type Measures Based on Count Data. Revue d'économie industrielle, 110(1), pp.149-156.

[30] Hall, B.H., Jaffe A.B. and Trajtenberg, M. (2001). The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools. Centre for Economic Policy Research. Discussion Paper No. 3094

[31] Hall, B.J., Jaffe, A. and Trajtenberg, M. (2005). Market Value and Patent Citations. The RAND Journal of Economics, Vol. 36, No. 1 (Spring, 2005), pp. 16-38.

[32] Harhoff, D. and Wagner, S. (2009). The Duration of Patent Examination at the European Patent Office. Management Science, 55(12), pp.1969-1984.

[33] Harhoff, D., Narin, F., Scherer, F. and Vopel, K. (1999). Citation Frequency and the Value of Patented Inventions. Review of Economics and Statistics, 81(3), pp.511-515.

[34] Harhoff, D., Scherer, F. and Vopel, K. (2003). Citations, family size, opposition and the value of patent rights. Research Policy, 32(8), pp.1343-1363.

[35] Hasch, J., Topak, E., Schnabel, R., Zwick, T., Weigel, R. and Waldschmidt, C. (2012). Millimeter-Wave Technology for Automotive Radar Sensors in the 77 GHz Frequency Band. IEEE Transactions on Microwave Theory and Techniques, 60(3), pp.845-860.

[36] Hegde, D. and Sampat, B. (2009). Examiner citations, applicant citations, and the private value of patents. Economics Letters, 105(3), pp.287-289.

[37] Heineke, K., Möller, T., Padhi, A., Tschiesner, A. (2017). The automotive revolution is speeding up. Retrieved from https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/the-automotive-revolution-is-speeding-up :

[38] Hill, M. N. (1962). Physical Oceanography. Allan R. Robinson. Harvard University Press. p. 498

[39] Hummon, N. and Dereian, P. (1989). Connectivity in a citation network: The development of DNA theory. Social Networks, 11(1), pp.39-63.

[40] Jun, S., and Lee, S. (2012). Emerging Technology Forecasting Using New Patent Information Analysis. International Journal of Software Engineering and its Application, Vol 6, No.3.

[41] Kaplan, S. (2012). Identifying Breakthroughs: Using Topic Modeling to Distinguish the Cognitive from the Economic. Academy of Management Proceedings, 2012(1), pp.1-1.

[42] Kim, S-Y., Oh, S-Y, Kang J-K et al. (2005). Front and rear vehicle detection and tracking in the day and night times using vision and sonar sensor fusion. Intelligent Robots and Systems, IEEE/RSJ International Conference.

[43] Konsert Strategy & IP (2018). INTELLECTUAL ASSET MANAGEMENT FOR TECHNOLOGY- BASED COMPANIES: An introduction to Intellectual Asset Management methodology and its application for improving competitive advantage and profitability. [online] Available at: https://static1.squarespace.com/static/5808a01715d5db271afe301e/t/5908da86d482e90d12aff900/1493752609921/IAM+in+technology-based+companies.pdf [Accessed 14 May 2018].

[44] Kurzweilai (2001). The Law of Accelerating Returns | Kurzweil. [online] Available at: http://www.kurzweilai.net/the-law-of-accelerating-returns [Accessed 14 May 2018].

[45] Lachner, R. (2013). Industrialization of mmWave SiGe technologies: Status, future requirements and challenges. Silicon Monolithic Integrated Circuits in RF Systems (SiRF).

[46] Lai, Y. and Che, H. (2009). Modeling patent legal value by Extension Neural Network. Expert Systems with Applications, 36(7), pp.10520-10528.

[47] Lanjouw, J. and Schankerman, M. (2001). Characteristics of Patent Litigation: A Window on Competition. The RAND Journal of Economics, 32(1), p.129.

[48] Lanjouw., J.O. (1998). STYLIZED FACTS OF PATENT LITIGATION: VALUE, SCOPE AND OWNERSHIP. No. El/20 Leddartech. (2018). LeddarTech | Mastering LiDAR Sensor Technology. [online] Available at: https://leddartech.com [Accessed 28 April 2018].

[49] Lerner, J. (1994). The Importance of Patent Scope: An Empirical Analysis. The RAND Journal of Economics, 25(2), p.319.

[50] Luettel, T., Himmelsbach, M. and Wuensche, H. (2012). Autonomous Ground Vehicles—Concepts and a Path to the Future. Proceedings of the IEEE, 100 (Special Centennial Issue), pp.1831-1839.

[51] Luettel, T., Himmelsbach, M., Müller, A. and Wuensche H-J. (2008). LIDAR-based 3D object perception. Conference Paper: Proceedings of 1st International Workshop on Cognition for Technical Systems, At Munich, Germany. Luminartech.com. (2018). Luminar. [online] Available at: https://www.luminartech.com [Accessed 6 May 2018].

[52] Magee, C. and Devezas, T. (2011). How many singularities are near and how will they disrupt human history?. Technological Forecasting and Social Change, 78(8), pp.1365-1378.

[53] Militaryradar.iqpc.co.uk. (2018). Military Radar. [online] Available at: https://militaryradar.iqpc.co.uk [Accessed 10 May 2018].

[54] Modis, T. (2001). Forecasting the growth of complexity and change. Technological Forecasting and Social Change, 69(4), pp.377-404.

[55] Moebius, B., Pfennigbauer, M. and Pereira do Carmo, J. (2017). Imaging lidar technology: development of a 3D-lidar elegant breadboard for rendezvous and docking, test results, and prospect to future sensor application. International Conference on Space Optics, Vol. 10565.

[56] Moehrle, M., Walter, L., Bergmann, I., Bobe, S. and Skrzipale, S. (2010). Patinformatics as a business process: A guideline through patent research tasks and tools. World Patent Information, 32(4), pp.291-299.

[57] Moore, G. (1998). Cramming More Components Onto Integrated Circuits. Proceedings of the IEEE, 86(1), pp.82-85.

[58] Nagy, B., Farmer, J., Bui, Q. and Trancik, J. (2013). Statistical Basis for Predicting Technological Progress. PLoS ONE, 8(2), p.e52669.

[59] Nordhaus, W. (2014). The Perils of the Learning Model for Modeling Endogenous Technological Change. The Energy Journal, 35(1).

[60] Okada, Y., Naito, Y. and Nagaoka, S. (2016). Claim Length as a Value Predictor of a Patent. Institute of Innovation Research Hitotsubashi University, Tokyo.

[61] Pendharkar, P. (2009). Genetic algorithm based neural network approaches for predicting churn in cellular wireless network services. Expert Systems with Applications, 36(3), pp.6714-6720.

[62] Petrusson, U. (2015). Research and Utilization. Tre Böcker Förlag AB. Göteborg.

[63] Ragoese, E., Scuderi, A., Giammello, V. et al. (2009). A fully integrated 24 GHz UWB radar sensor for automotive applications. Solid-State Circuits Conference - Digest of Technical Papers.

[64] Roland Berger Strategy Consultants (2014). Autonomous driving Disruptive innovation that promises to change the automotive industry as we know it — it's time for every player to think:act!. München: Roland Berger Strategy Consultants GmbH.

[65] Roselli, L. and Alimenti, F. (2005). A cost driven 24GHz Doppler radar sensor development for automotive applications. Microwave Conference, European, Vol 3.

[66] Russell, M., Crain, A., Curran, A., Campbell, R., Drubin, C. and Miccioli, W. (1997). Millimeter-wave radar sensor for automotive intelligent cruise control (ICC). IEEE Transactions on Microwave Theory and Techniques, 45(12), pp.2444-2453.

[67] Sahal, D. (1979). A Theory of Progress Functions. A I I E Transactions, 11(1), pp.23-29. Schmookler, J. (1966). Invention and Economic Growth. Harvard University Press.

[68] Schwarz, B. (2010). Mapping the world in 3D. Nature Photonics, 4(7), pp.429-430.

[69] Squicciarini, M., Dernis, H. and Criscuolo C. (2013). Measuring Patent Quality: Indicators of Technological and Economic Value. OECD Science, Technology and Industry Working Papers.

[70] Steinbaeck, J., Steger, C., Holweg G. and Druml N. (2017). Next Generation Radar Sensors In Automotive Sensor Fusion Systems. Infineon Technologies Austria AG, Graz, Austria.

[71] Svensson, R. (2012). Commercialization, renewal, and quality of patents. Economics of Innovation and New Technology, 21(2), pp.175-201.

[72] Tesla (2018). Autopilot: Full Self-Driving Hardware on All Cars. [online] Tesla.com. Available at: https://www.tesla.com/sv$_SE/autopilot?redirect = no[Accessed 11 May 2018]$.

[73] Trajtenberg, M., Henderson, R. and Jaffe, A. (1997). University Versus Corporate Patents: A Window On The Basicness Of Invention. Economics of Innovation and New Technology, 5(1), pp.19-50.

[74] Trappey, A., Trappey, C., Wu, C. and Lin, C. (2012). A patent quality analysis for innovative technology and product development. Advanced Engineering Informatics, 26(1), pp.26-34.

[75] Trippe, A. (2002). Patinformatics: Identifying haystacks from space. ProQuest Central. Pg. 28. Trippe, A. (2003). Patinformatics: Tasks to tools. World Patent Information, 25(3), pp.211-221.

[76] Trippe, A. (2003). Patinformatics: Tasks to tools. World Patent Information, 25(3), pp.211-221.

[77] Triulzi, G., Alstott, J. and Magee, C. (2018). Estimating Technology Performance Improvement Rates by Mining Patent Data. SSRN Electronic Journal.

[78] Troilo, L. (2017). Dependent Claims National Patent Drafting Course. [online] Wipo.int. Available at: http://www.wipo.int/edocs/mdocs/mdocs/en/wipo$_{i}p_{c}nx_{1}7/wipo_{i}p_{c}nx_{1}7_{6}.pdf$ $[Accessed 14 Apr. 2018]$.

[79] University of Kent (2018). LibGuides: SPSS Tutorials: One-Way ANOVA. [online] Libguides.library.kent.edu. Available at: https://libguides.library.kent.edu/SPSS/OneWayANOVA [Accessed 27 Jun. 2018].

[80] Velodynelidar.com. (2018). Velodyne LiDAR. [online] Available at: http://velodynelidar.com [Accessed 1 May 2018].

[81] Wang, L., Zhang, Y. and Wang, J. (2017). Map-Based Localization Method for Autonomous Vehicles Using 3D-LIDAR * *This work is supported in part by the National Natural Science Foundation of China under Grant No. 61473209. IFAC-PapersOnLine, 50(1), pp.276-281.

[82] WIPO (1994). Uruguay Round Agreement: TRIPS Trade-Related Aspects of Intellectual Property Rights. [online] Wipo.int. Available at: http://www.wipo.int/treaties/en/text.jsp?file$_{i}d = 305907part2.5[Accessed 14 May 2018]$.

[83] Wipo.int. (2018). [online] Available at: http://www.wipo.int/edocs/mdocs/africa/en/ wipo$_{p}at_{h}re_{1}5/wipo_{p}at_{h}re_{1}5_{t1}0.pdf[Accessed 14 Apr. 2018]$.

[84] Wolcott, R. and Eustice, R. (2017). Robust LIDAR localization using multiresolution Gaussian mixture maps for autonomous driving. The International Journal of

Robotics Research, 36(3), pp.292-319.

[85] Yin, R. (1984). Case study research. Beverly Hills, Calif.: Sage Publications.

[86] Young, S. (2011). LIDAR in the Driver's Seat: New devices based on a concept similar to that of RADAR could revolutionize your daily commute. Light detection and ranging (LIDAR) technologies are providing the vision for a new generation of driverless vehicles. Optics Photonics Focus.

[87] Yvonne Feilzer, M. (2009). Doing Mixed Methods Research Pragmatically: Implications for the Rediscovery of Pragmatism as a Research Paradigm. Journal of Mixed Methods Research, 4(1), pp.6-16.

[88] Zhao, S. (2010). Automatic Underwater Multiple Objects Detection and Tracking Using Sonar Imaging. School of Mechanical Engineering The University of Adelaide, Adelaide, South Australia.