Valerius Yläjoki

IMPLEMENTING DIGITAL TWIN IN PRACTICE: PERSPECTIVES AND CHALLENGES

JYVÄSKYLÄN YLIOPISTO
INFORMAATIOTEKNOLOGIAN TIEDEKUNTA
2023

TIIVISTELMÄ

Yläjoki, Valerius
Digitaalisen kaksosen käytännön toteutuksen perspektiivejä ja haasteita
Jyväskylä: Jyväskylän yliopisto, 2023, 78 s.
Tietojärjestelmätiede, pro gradu -tutkielma
Ohjaaja(t): Abrahamsson, Pekka & Moisala, Jani

Tämän pro gradu tutkielman aiheena on digitaalisen kaksosen toteutuksen perspektiivejä ja haasteita. Kyber-fyysiset järjestelmät käsitteenä on ollut käytössä laajasti teollisuus 4.0:n yhteydessä, minkä alle digitaalinen kaksonenkin lukeutuu. Digitaalinen kaksonen voidaan nähdä osana kyber-fyysistä järjestelmää teollisuus 4.0 perspektiivistä tarkastellessa. Digitaalisen kaksosen hyötyjä on dokumentoitu laajasti akateemisessa kirjallisuudessa, mutta käytännön toteutuksien dokumentointi on vajavaista. Akateeminen kirjallisuus sekä harmaa kirjallisuus tarjoavat laajalti erilaisia viitekehyksiä digitaaliseen kaksoseen sekä sen kehittämiseen liittyen. Lean startup methodi on kevyt ja nopea iteratiivinen tuotekehityksen viitekehys, jota voidaan hyödyntää ymmärtämään markkinoiden ja asiakkaiden erilaisia tarpeita kehittäessä tuotetta pienestä ensimmäisestä versiosta lopulliseen tuotteeseen asiakkaan tarpeiden perusteella. Tämän tutkielman tarkoitus on tutkia käytännössä tapahtuvia toteutuksia digitaaliseen kaksoseen liittyen, niiden haasteita sekä perspektiivejä. Tutkimuksen suoritustapa on laadullinen empiirinen haastattelututututkimus, jossa haastateltavana olleet toimivat erilaisissa tuotekehityksen johtotehtävissä digitalisaatiossa yrityksissä. Tutkimuksen tuloksista selviää, että yrityksillä on erilaisia digitaalisen kaksosen ratkaisuja sekä komponentteja käytössä. Myös digitaalisen kaksosen haasteet sekä tuomat hyödyt ovat yrityksien mielestä selkeitä.

Asiasanat: kyber-fyysiset järjestelmät, digitaalinen kaksonen, lean startup methodi, teollisuus 4.0, minimum viable product

ABSTRACT

Yläjoki, Valerius
Implementing digital twin in practice: perspectives and challenges
Jyväskylä: University of Jyväskylä, 2023, 78 pp.
Information Systems Science, Master's thesis
Supervisor(s): Abrahamsson, Pekka & Moisala, Jani

The subject of this thesis is the perspectives and challenges of implementing a digital twin. The concept of cyber-physical systems has been widely used in the context of Industry 4.0. The digital twin can be seen as part of the cyber-physical system from an Industry 4.0 perspective. The benefits of the digital twin have been extensively documented in academic literature, but the documentation of practical implementations is incomplete. The academic and grey literature offers a wide range of frameworks for the digital twin and its development. The Lean startup method is a light and fast iterative product development framework that can be used to understand the different needs of the market and customers when developing a product from an initial minimum viable product to a final product based on customer needs. The purpose of this thesis is to explore the practical implementations of the digital twin, their challenges, and perspectives. The research method is a qualitative empirical interview study, in which the interviewees work in different product development management positions in digitalization. The results of the research show that companies have different digital twin solutions and components in use. The challenges and benefits of the digital twin are also clear to companies.

Keywords: cyber-physical systems, digital twin, lean startup method, industry 4.0, minimum viable product

FIGURES

TABLE

TABLE OF CONTENTS

1 INTRODUCTION

The ever-evolving world is moving rapidly towards the digitalization of societies. Digitalization has been evolving our society from the early 1990's all the way to this day. The phenomenon has been visible for the early adopters and industries across various domains from the infancy of digitalization. Digitalization has been revolutionizing various ways of how products are made, formed completely new industries, destroyed existing industries, and is still going forward. The motto for many of the succeeding people and companies in the continuously changing business landscape has been to stay ahead of the curve and learn to adapt to the change – or wither away with time.

Digitalization was the driving force of the third industrial revolution and now interconnectedness of the cyber-physical systems or computers, emergence of Internet of Things, the ever-growing connectiveness and networking of the machines and systems – fourth industrial revolution is on its way and currently happening across various industries. Industry 3.0 revolutionized the information flow on the planet, and this brought up new ways of trade and business. Supply chains became crucial parts of the new global trade network, which brought costs down in production by sourcing production to their most cost-effective places.

The concept of Industry 4.0, also known as the Fourth Industrial Revolution, refers to the current trend of automation and data exchange in manufacturing technologies, including the integration of artificial intelligence, the Internet of Things, and cloud computing. This trend is transforming traditional manufacturing and industry. This transformation has the potential to bring significant improvements in efficiency, productivity, and competitiveness. Jazdi (2014) lists the industry 4.0 characteristics as smart networking, mobility, flexibility, integration of customers, and new innovative business models. Seizing these opportunities and building new businesses on these is crucial part on staying on top of the wave and staying competitive for companies in their own domains. Companies need understanding and information regarding these business opportunities. Seeing the big picture and the direction of the future is needed as well. Daring innovation is bringing companies competitive advantages, as well as bringing knowledge from various teams into a collective pool of data, where companies can derive information to base their decisions on.

The adoption of Industry 4.0 technologies, including digital twins and cyber-physical systems, has the potential to bring numerous benefits to companies, such as improved production processes, enhanced supply chain management, and increased customer satisfaction. It may also facilitate the creation of new business models and revenue streams, as well as the development of innovative products and services. China has identified Industrial Internet, Industry 4.0 in their advanced manufacturing strategies (Tao et al., 2019). According to Capgemini Consulting (2017), Industry 4.0 will have the following value drivers: smart solutions, smart innovation, smart supply chains, and smart factory. Capgemini Consulting (2017) also lists the requirements for the digital transformation of the manufacturing organization under the following umbrella terms: digital infrastructure, governance & processes, people leadership & change, and the agile operating model.

Digital Twin has been flashed in many white papers alongside with cyber-physical systems in the context of industry 4.0 transition (Uhlemann et al., 2017). A digital twin is a virtual representation of a physical asset, process, or system, and it allows for real-time monitoring and optimization through the integration of data from sensors and other sources. Digital twin being part of the industry 4.0 transition, it is a concept with broad potential once understood to it’s fullest potential. Thus, it is important for companies to have understanding on the concept and how to possibly start implementing digital twins.

When dealing with new concepts such as digital twins it is important to choose the right method in developing such products. The lean startup method has gained significant attention in recent years as a framework for developing and launching new products. This approach emphasizes a focus on iteration and continuous learning, with a particular emphasis on the concept of "validated learning" through the collection of data and customer feedback. In the fast-paced and constantly evolving digital product development industry, the lean startup method offers a valuable approach for minimizing risk and increasing the chances of success for new ventures.

The core principle of the lean startup method is the development of a minimal viable product (MVP) to quickly bring a product to market and gather feedback from customers (Ries, 2011). This approach allows for rapid testing of assumptions and the ability to pivot, if necessary, rather than investing a large amount of time and resources in building a fully-featured product that may not meet the needs of the target market.

As such, the lean startup method has the potential to be a valuable tool for entrepreneurs and startups seeking to enter the digital product development space. Overall, the lean startup method can be an effective approach for product development because it helps to minimize risk, increase speed to market, facilitate continuous learning and iteration, and promote a customer-centric focus.

1.1 Motivation

The term "digital twin" has been around for two decades as a concept firstly coined by Grieves in a conference held in University of Michigan in 2003 about Product lifecycle Management (Grieves, 2014). Digital twin has been characterized as one of the driving forces of next industrial revolution, Industry 4.0. One cannot argue that digital twin has no benefits for taking the next step in terms of digitalization in the manufacturing, processing, and technological industries. With the even more applicable adaptations of IoT and IIoT solutions, sizing coming down and costs of adapting these solutions – digital twins seem more favorable options to upgrade into in the general plant environment.

However, there still are challenges within the digital twin and deployment of it in the plant environments and adoption in practice. Firstly, the concept digital twin covers a wide variety of concepts and the answer to the question "what is a digital twin" depends a lot on whom you are asking. This creates a problem in the industry and academia, where the digital twin has become this huge umbrella term for a variety of solutions which could be characterized as different kinds of digital twins. Secondly, the challenges regarding the digital twin's adaptation on the technology side has been researched and documented. These challenges are for example the need of the processing power required for the digital twin's and physical twin's data & information exchange, which has been criticized and questioned. The cost of processing power and calculations cannot outweigh the benefits of the digital twin's pros. By bringing value for the customer and especially for the customer's money, digital twin can become the next big thing via the next industrial revolution, industry 4.0.

Minimum viable product (MVP) has been firstly introduced by Eric Ries in 2011 in his book The Lean Startup. The general idea of minimum viable product is to produce the absolute minimum product for the customer to test the product, and then iterate the product based on the customers' feedback (Ries, 2011). This product development framework aims to cut the waste in the development process and focus on delivering value to the customer alongside the development process from start to finish.

By combining MVP framework for the development of the digital twin, this thesis aims to benefit companies in getting started with the implementation of digital twins. There are multiple moving elements ranging from the high costs of creating the actual digital twin, the maintenance, and data analysis of the twin's data, how, by who to whom, and where to utilize the digital twin – all these need to be addressed to get the requirements for the implementation of digital twin. By finding these crucial elements in the equation and answering the possible pitfalls of digital twin's development, this thesis tries to provide the keys for kickstarting the process by giving a picture to digital twins through academical literature and industry.

1.2 Research questions

This thesis seeks information on the implementation of digital twins, its perspectives, and challenges. The information this thesis seeks is crucial in finding new business opportunities and to better development of the engineering solutions the industry 4.0 transition is forcing companies to adapt to. The goal of these questions is to help find the different technological requirements for implementation of the digital twin, and its technologies. By focusing on these different aspects, the goal is to get a good grasp of different industries' standards, uses of digital twin technologies, the different requirements industries need and have currently.

This study seeks answers to the following primary research question:

- **How to implement a digital twin in practice?**

And to support the thesis' research process and primary research question, the following supporting research questions are proposed:

- **From what different components are digital twins built from?**

This research question will try to piece together different components on how the digital twin is built. By getting requirements for the components of digital twin, it should ease the implementation of digital twins.

- **What are the challenges of implementing digital twins?**

This research question will try to answer the different challenges related to the implementation of digital twins in practice and in theory. Digital twins being very complex concepts, the challenges can be complex too. By understanding these different challenges, it should ease the implementation of digital twins in practice.

- **How is the industry currently utilizing digital twin technologies?**

Lastly, this research question aims to get the different industries' perspectives on the topic of implementations of digital twins. By getting answers to this question, it should bridge the gap between academical literature and industry's practices, which aren't always aligned or are missing in one way or the other.

1.3 Structure of thesis

This thesis begins with the introduction chapter first, which includes the motivation for the research and why as a study topic it is relevant. Also, research questions are defined on the 2nd sub-chapter of the introduction chapter. This 3rd sub-chapter lays the foundation on the structural side of the thesis and how it flows thoroughly through the research process.

From the basis of the introduction chapter, the first data collection as a literature review is conducted on the academic material found on the study topic for the following concepts – digital product development, cyber-physical systems, and digital twin.

After the literature review has been concluded and key topics covered, the thesis will lay out the research framework for the empirical research, which is a case study conducted as a semi-structured thematic interview. This will be the 2nd data collection, which forms the empirical data for this thesis.

After the interviews are completed and the empirical data has been gathered, begins the analysis of the collected data. In the data analysis, the empirical findings of the interviews are analyzed through integrated approach. Discussions chapter will follow the empirical findings chapter, which will discuss the results from the empirical findings. Lastly, the conclusions chapter concludes this thesis and will provide answers to the research questions, limitations of study and further research angles.

1.4 Conducting the literature review

This thesis uses Google Scholar search engine and database, where source material will be located firstly. Due to the complexity and relatively new concepts, this thesis will utilize a broad selection of sources in the literature review. These sources include major scientific publication journals in the likes of SAGE, IEEE, Springer, or other similarly rated on the trustworthiness on the JUFO-rating (Julkaisufoorumi). Some references were backwards searched through the sources of peer-reviewed academic research articles, which were the first hit on the Google Scholar for used keywords. On top of the academical research seen relevant on the topics of this thesis, the scope will be broadened to relevant white papers and journals, which could be considered less viable as they aren't peer-reviewed academically. However, due to the major industry leaders having an edge on the practical side of cyber-physical systems, digital twins, and lean startup model, this thesis will utilize sources relevant to the topics outside the academic research databases. By including grey literature, it tackles the complexity of the interventions (Garousi et al., 2016).

Some of the sources utilized in this thesis couldn't be verified in their JUFO-rating, but these sources were seen relevant and filled information gaps from the academic research. These sources could be seen as grey literature. According to GreyNet International (2023):

> Grey literature is a field in library and Information science that deals with the production, distribution, and access to multiple document types produced on all levels of government, academics, business, and organization in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body (GreyNet International, 2023).

Grey literature can be helpful addition to complex research topics. According to Mahood et al. (2013), "including grey literature can broaden the scope to more relevant studies, thereby providing a more complete view of available evidence". Some of the source material this thesis utilizes can be seen being on the edge of academical research papers and grey literature. Since clear majority of the sources are peer-reviewed academic research, the literature review is done based on peer-reviewed academic research. Inclusion of the grey literature adds industry's take on the research topic, which broadens the scope to more practical side, which is beneficial due to the studied topic being relatively new.

The literature review will include two different themes – cyber-physical world and digital product development. These two themes house key concepts for the basis of this thesis. Firstly, cyber-physical world will open cyber-physical systems and digital twins, and second, digital product development will open lean startup method, the early-stage software startup development model, and continuous experimentation.

1.5 Disclaimers

In this thesis, OpenAI's ChatGPT has been utilized for the raw text generation in the introduction chapter, and it has also been utilized to better the text together with the writer of the thesis on the same chapter.

2 CYBER-PHYSICAL WORLD

In this chapter, first cyber-physical system as a concept is opened through academical literature. After opening cyber-physical systems first, this chapter then proceeds to give an overview on the concept of a digital twin. The concept of a digital twin is thought-out through academical literature and various industry white papers, to get the most coherent picture of the concept itself. Cyber-physical systems are crucial part of the fourth industrial revolution and can be seen as one of the enablers in the revolution just like the computers were in the third industrial revolution. The third industrial revolution digitalized the industry through ICT, and now through cyber-physical systems and IoT the fourth revolution is taking its place (Jazdi, 2014).

2.1 Cyber-Physical Systems

This chapter focuses on the cyber-physical systems. The chapter starts with a brief introduction to cyber-physical systems followed by the present and industry 4.0 transition of the cyber-physical systems. Lastly, it gives a brief overview of components from which the cyber-physical system is formed from. After reading this chapter, the reader of this thesis should understand the context of cyber-physical systems in the industry 4.0 transition and be able to form an understanding of the different components cyber-physical systems are built from, and the different challenges related to the cyber-physical systems.

2.1.1 Cyber-physical systems as a concept

First, we need to define a cyber-physical system and have the foundations for the concept. A brief introduction is laid out by Zhong et al. (2017) on which different components form a cyber-physical system:

> A CPS-enabled system, unlike a traditional embedded system, contains networked interactions that are designed and developed with physical input and output, along with

> their cyber-twined services such as control algorithms and computational capacities. Thus, a large number of sensors play important roles in a CPS. For example, multiple sensory devices are widely used in CPS to achieve different purposes, such as touch screens, light sensors, and force sensors. (Zhong et al., 2017.)

A definition on cyber-physical systems can also be done on a much broader scale without mentioning different technologies or naming components outright. Baheti and Gill (2011) define cyber-physical systems in the following fashion:

> A new generation of systems with integrated computational and physical capabilities that can interact with humans through many new modalities. (Baheti & Gill, 2011.)

This definition by Baheti and Gill (2011) has a more human-centric way of defining the cyber-physical systems as a concept, compared to many other research articles which focus on the more technological or architectural definitions.

Cyber-physical systems can vary a lot in their application, since the cyber-physical systems as a term can cover a wide variety of domains and fields. Cyber-physical systems can be seen for example HVAC (heating, ventilation, and air conditioning) machines, aircrafts, asset management, power generation and distribution to different military systems (Lee, 2015). In FIGURE 1, there is a cyber-physical system's structure according to Lee (2015). This FIGURE 1 represents the basic information flow behind the cyber-physical system.

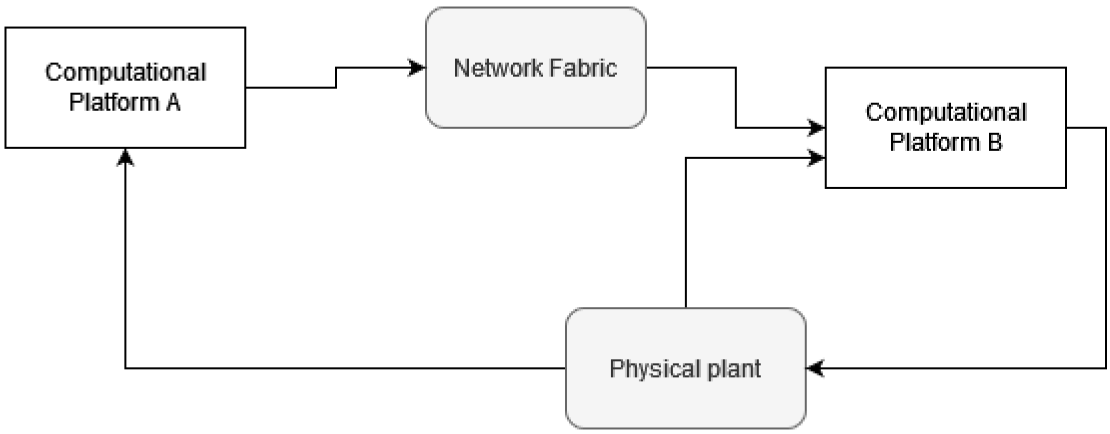

FIGURE 1 Cyber-physical system described in a figure (Lee, 2015)

Cyber-physical systems consist of many different systems, which form a new entity. According to Rajkumar, Lee, Sha and Stankovic (2010), "CPS represent a confluence of technologies in embedded systems, distributed systems, dependable systems, real-time systems with advances in energy-efficient networking, microcontrollers, sensors and actuators." Lastly, if we want to define the key characteristics for a cyber-physical system, Alur (2015) characterizes cyber-physical systems' key features being the following:

1. Reactive computation,
2. Concurrency,
3. Feedback Control of the physical world,
4. Real-Time Computation, and
5. Safety-Critical Applications

2.1.2 Challenges of cyber-physical systems

Cyber-physical systems challenges have been documented in various studies in the academic literature. Challenges and opportunities regarding continuous experimentation for cyber-physical systems have been identified as hardware constraints, feedback data, safety guarantees, involving more stakeholders and supportive instruments for continuous experimentation (Giaimo et al., 2016).

Cyber-physical systems require a lot of different components and specifications ranging from computing power to networking technologies and software-orientation. These requirements prove cyber-physical systems difficult to optimize to their fullest potential. Adding to the problem is the modern existing design processes, level of abstractions, and/or verifying the today's abstractions in the models, which are needed to be rebuilt for cyber-physical systems to gain their fullest potential (Lee, 2008). This adds the architectural challenges of the cyber-physical systems into the mix, which would need to be solved by innovating the processes used to develop these systems in the first place.

This pattern of same challenges continues in other studies done regarding the potential of cyber-physical systems. In a study conducted by Rajkumar et al., (2010), the researchers concluded "CPS technologies must be scalable across time and space, and must deal with multiple timescales, uncertainty, privacy concerns and security issues ". This would support the safety concerns as well as hardware constraints in the terms of hardware power being the limiting factor. Another survey study conducted by Yu et al. (2021) concluded that it is important to understand the ever-changing landscape with cyber-physical systems and their surfaces, which are vulnerable to different kinds of attacks ranging from signal injection to different data & information leakages.

Data breach concerns are one of the issues highlighted in many research articles. For the companies utilizing cyber-physical systems it would be harmful to their business if data or information regarding their core processes would leak or get systems breached. According to Wurm et al. (2016), the issue with cyber-physical systems' massive data collection is the breaches in different parts of the data collection process, which then can lead to sensitive or private information leakage in a large scale.

One of the issues of cyber-physical systems was named in the efforts for scaling the operations. This in term means proper utilization of hardware behind powering the cyber-physical system. According to researchers Mosterman and Zander (2016), it is important for the hardware resources to be shared in a proper way to ensure the system's proper utilization.

2.1.3 Cyber-physical systems in industry 4.0

Cyber-physical systems and digital twins have been used extensively in the industry 4.0 literature, both together and as individual concepts from one another. It is crucial that the difference between the concepts of Digital Twin and Cyber-physical systems is laid out. While digital twin emerged more from the engineering related problem-solving, cyber-physical systems emerged from the scientific side of problem-solving. Where digital twin's core elements consist of models and data, cyber-physical system's core elements consist of sensors and actuators. (Tao et al., 2019).

In industry 4.0 the cyber system transformation to cyber-physical system is highlighted in many research articles. Industry 4.0 development benefits from cyber-physical systems through direct system extension, system expansion by microcontroller board, and extension by smart actuators and sensors (Jazdi, 2014). Another article mentions the cyber-physical system being in the center of the industry 4.0 transition according to the German government for smart factory enabling (Jiang, 2018). Research done by Wang et al., (2016) highlights the importance of cyber-physical systems in the industry 4.0 transition in the following words:

> The proliferation of cyber-physical systems introduces the fourth stage of industrialization, commonly known as Industry 4.0. The vertical integration of various components inside a factory to implement a flexible and reconfigurable manufacturing system, i.e., smart factory, is one of the key features of Industry 4.0. (Wang et al., 2016)

In TABLE 1 below from Lee et al., (2015), there are key differences illustrated, how future factories differ in industry 4.0 transition to factories in currently. Main key takeaways are the self-governing aspects of the attributes, where they enable the operation of the factory to rely more on the technologies' prediction, configuration, and organization, rather than the operators operating and doing decision-making actively reacting to the feedback they get.

TABLE 1 Comparison between the today's factories and industry 4.0 factories (Lee et al., 2015)

Comparison of today's factory and an Industry 4.0 factory.					
Data source		Today's factory		Industry 4.0	
		Attributes	Technologies	Attributes	Technologies
Component	Sensor	Precision	Smart sensors and fault detection	Self-aware & self-predict	Degradation monitoring & remaining useful life prediction
Machine	Controller	Producibility & performance	Condition-based monitoring & diagnostics	Self-aware, self-predict & self-compare	Up time with predictive health monitoring
Production system	Networked system	Productivity & OEE	Lean operations: work and waste reduction	Self-configure, self-maintain & self-organize	Worry-free productivity

Another review done by Zheng et al. (2018) paint the technological stack of information and communication technologies as enablers of physical and virtual world fusion, which is done via the cyber-physical systems in the manufacturing industry.

2.1.4 Architectures of a cyber-physical system

In this chapter there will be several different architectures derived from the academical literature, different white papers, and grey literature. Cyber-physical systems are wide concepts, and thus one architectural way of approaching them is not sufficient to get the similarities between the different approaches from different authors. By comparing multiple different cyber-physical system architectures, this thesis can open different ways of seeing and implementing different cyber-physical systems, and in the end – digital twins.

In a literature review conducted by Pivoto et al. (2021), they reviewed main cyber-physical system reference architectural models for the industry 4.0 and listed the following three as main architectures:

- 5C Architecture
- Reference Architectural Model Industrie 4.0 (RAMI4.0)
- Industrial Internet Reference Architecture (IIRA)

First of the architectural configurations for cyber-physical systems is by authors Lee et al. (2015), the 5C archictecture, which is illustrated by in FIGURE 2. According to Lee et al. (2015), the functions are represented by the left side of the pyramid and right side is illustrating the attributes.

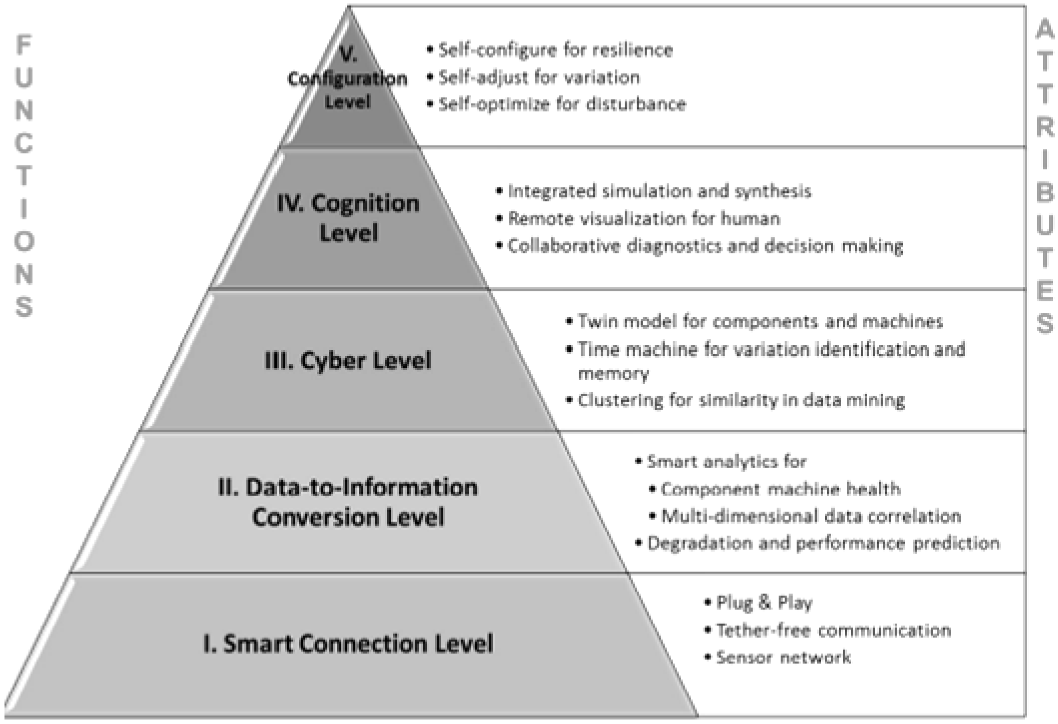

FIGURE 2 5C architecture for implementation of Cyber-Physical System (Lee et al., 2015)

In FIGURE 2 formed based on Lee et al. (2015): the pyramid starts from the 1st smart connection layer, which serves as the base layer of building a cyber-physical system. The 2nd layer going towards the top is data-to-information conversion layer, which serves as a connector between the smart connection layer and the cyber level. Cyber level is 3rd in the pyramid when going towards the top, and the cyber level acts as a bridge between the different machines operating in the network, it also controls the flow of information and analytics. 4th layer is cognition layer, which lays the processed and analyzed information and knowledge to the decision-making parties to better understand the current situation and status of the network's machines. 5th final level on the top is the configuration level, which is a feedback system acting as a resilience control system for both physical and cyber space in the whole cyber-physical system. TABLE 2 below adapted from Lee et al. (2015) represents different applications and techniques in the various levels illustrated in the FIGURE 2's 5C architecture model.

TABLE 2 Applications and techniques associated with each level of the 5C architecture adapted from Lee et al., (2015)

Level	**Applications & techniques**	**System**	**Goal**
Configure	Supervisory Control -> Required Actions	Resilient Control System (RCS)	Actions to Avoid
Cognition	Decision-making, optimatization & Analysis	Decision Support System (DSS)	Prioritize and Optimize Decisions
Cyber	Fleet of Machines, adaptive analysis, P2P monitoring & Time-Machine Snapshots	Cyber-Physical Systems (CPS)	Self-Compare
Conversion	Machines	Prognostics and Health Management (PHM)	Self-Aware
	Components		
Connection	Sensors & Effective Sensor Selection	Condition Based Monitoring (CBM)	Condition Monitoring

5C architecture was the first of three architectural models for cyber-physical systems in the industry 4.0 transition according to Pivoto et al. (2021). The second architecture, RAMI4.0 is illustrated in FIGURE 3 below. The RAMI4.0 by DIN SPEC 91345:2016-04 standard is illustrating different architecture model for cyber-physical systems: “The reference architecture model RAMI4.0 is a reference model for an Industrie 4.0 reference architecture and gives a structured description of fundamental ideas.”

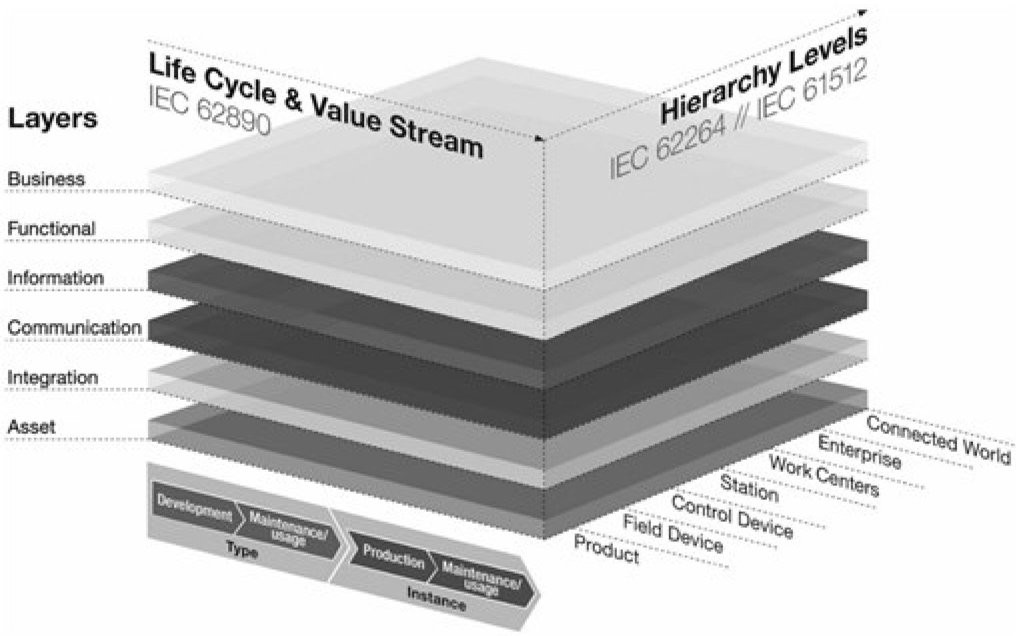

FIGURE 3 Reference architecture model Industrie 4.0 (RAMI4.0) (DIN SPEC 91345:2016-04)

RAMI4.0 according to DIN SPEC 91345:2016-04 is broken down to three different axes:

- "The architecture axis ("Layers"), with six layers to represent the information that is relevant to the role of the asset;
- The "Life cycle & value stream" axis to represent the lifetime of an asset and the value-added process, based on IEC 62890;
- The "Hierarchy levels" axis for assigning functional models to specific levels, based on the DIN EN 62264-1 and DIN EN 61512-1 standards"

Last of the architectural models named for cyber-physical systems for industry 4.0 was IIRA. According to Lin et al. (2017), "IIRA is a standards-based open architecture for IIoT systems. Its broad industry applicability maximizes its value. It provides an architecture framework, including methods and templates, to design industrial internet systems, without making specific recommendations for standards or technologies that comprise these systems. Core to IIRA are the different business and technical perspectives described as viewpoints for identifying and addressing architectural concerns." In IIRA these multiple viewpoints for different stakeholders are business, usage, functional and implementation viewpoints (The Industrial Internet Reference Architecture, 2022).

IIRA could be seen as a guideline for developers, as the model itself doesn't specify technologies, but rather gives outlines and guidelines for different use cases and viewpoints. According to The Industrial Internet Reference Architecture (2022), "the purpose of the IIRA is to provide guidance to system architects to assist the architects in building IIoT systems. The IIRA v2.0 has been designed to improve the user experience (and increase the value provided) by addressing the stakeholder concerns more clearly." Below in the FIGURE 4 there is an illustration adapted from iiconsortium.org to illustrate the general idea on IIRA architecture's model.

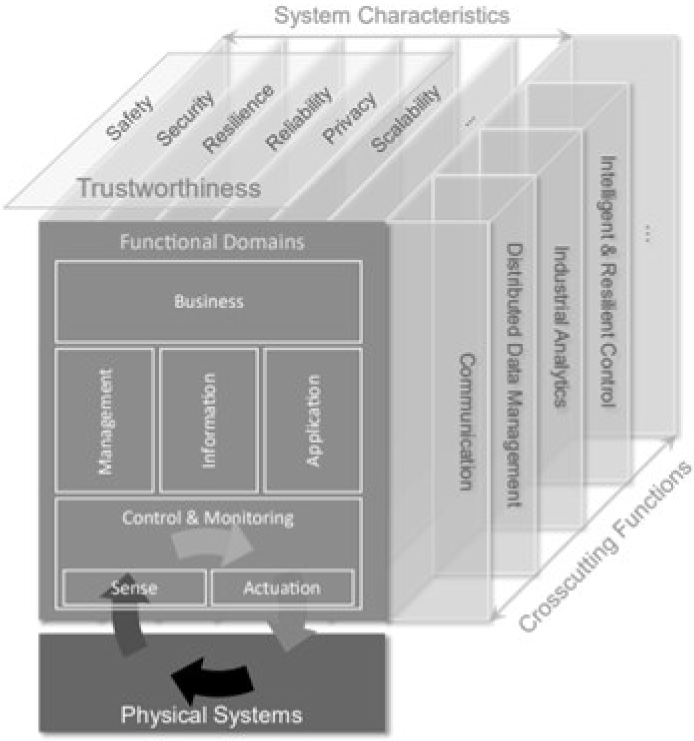

FIGURE 4 Functional domains, crosscutting functions and system characteristics (The Industrial Internet Reference Architecture, 2022)

After the introduction of these three different architectural models, it is time to wrap up them together. These three different architectural models can be mapped into a figure adopted from Pivoto et al. (2021) in the FIGURE 5 below:

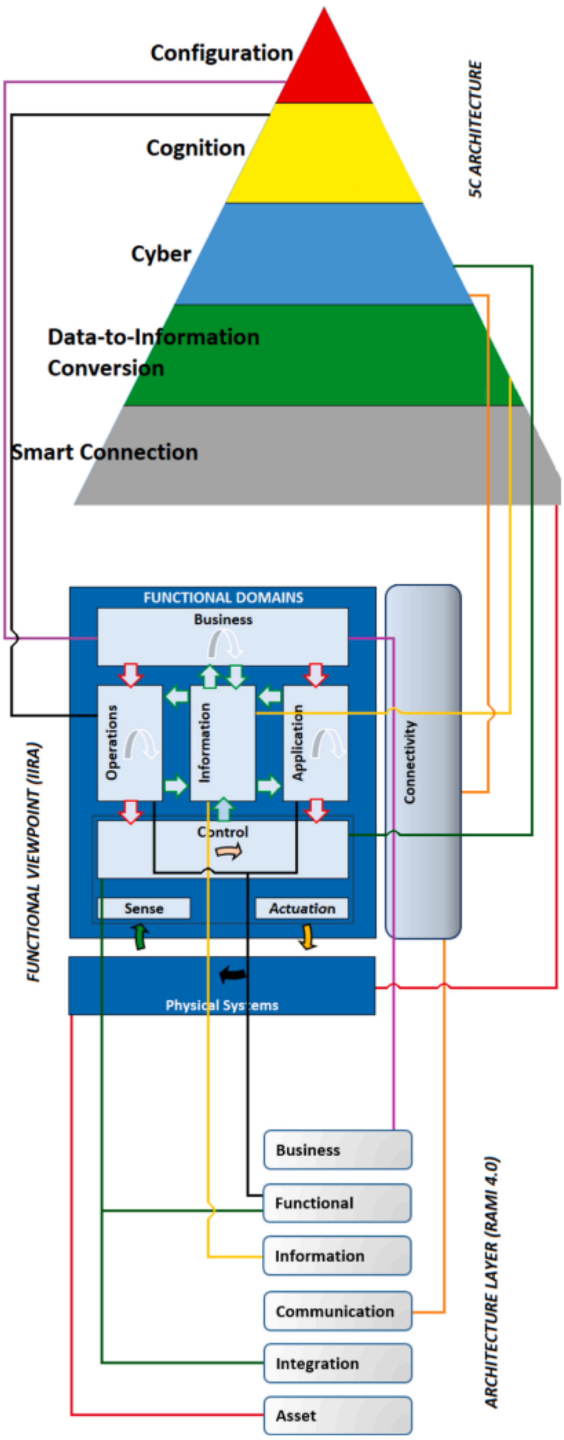

FIGURE 5 Functional mapping among 5C Architecture, IIRA and RAMI4.0 (Pivoto et al., 2021)

All the models 5C, IIRA and RAMI4.0 overlap in some sections and have similarities in them. According to the authors Pivoto et al. (2021), the differences between these models are in their different use cases:

- “The 5C Architecture is focused on assets data acquisition and processing, commonly used in embedded systems and small industrial environments. It is one of the first CPS architectures disseminated in the literature;
- Based on the Smart Grid Architecture Model (SGAM), RAMI 4.0 was created to adapt the CPS architecture in the I4.0 scenario. It defines how a manufacturing plant can operate, and it is centered in the manufacturing sector deeply through the product lifecycle, integrating the value chain of the company;
- With IIoT proposal as a highlight, IIRA is based on the ISO/IEC/IEEE 42010 and it defines how an IIoT system can be developed, centered in IIoT systems concerns in all sectors, such as products’ operation and maintenance, business and mainly in the interoperability among industries.”

Together these three architectures of 5C, RAMI4.0 and IIRA can be viewed as viewpoints or modelling tools for cyber-physical systems. Depending on the viewpoint of the user, these all have their own place in the cyber-physical systems development and architectural mapping.

2.2 Digital Twin

In this chapter, the concept of a digital twin is opened through the academical literature. This chapter also includes the challenges of a digital twin, architecture of a digital twin as well as the next generation of digital twin. After this chapter, the reader has a clear picture how digital twin is defined in this thesis’ context, what forms the digital twin and what the future might hold for the digital twin as well.

2.2.1 Definition of the Digital Twin

Digital twin can be considered as a wide range of definitions depending on the domain and industry you are in. Also, the application of digital twin has a big impact on how digital twin can be defined, or what can be considered as a digital twin. Digital Twin has been around different industries from the early 2000’s. It was first coined in 2003 by Michael Grieves at the University of Michigan during Executive Course on Product Lifecycle management (PLM) (Grieves, 2014). According to Grieves (2014):

> At the time this concept was introduced, digital representations of actual physical products were relatively new and immature. In addition, the information being collected about the physical product as it was being produced was limited, manually collected, and mostly paper-based (Grieves, 2014).

After Grieves' first introduction of Digital Twin, the concept has gained maturity in the industries of practice and in research. The industry and academic representatives have differentiating definitions regarding the concept, varying from simple representations to highly abstract and technical definitions. According to Grieves (2014):

> In the decade since this model was introduced, there have been tremendous increases in the amount, richness, and fidelity of information of both the physical and virtual products (Grieves, 2014).

Almost a decade later from Grieves' article, information regarding Digital Twin has been increasing due to new research efforts, industries and technology maturing enough for the adaption of digital twin.

Highly technical and complex definitions of Digital Twin have been existing in the earlier days of the concept. NASA's researchers Glaessgen and Stargel (2012) paint Digital Twin as an ultra-realistic representation of real-life counterpart, which mimics the physics, predicts the future events to the twin's physical counterpart and give critical information regarding the lifecycle of the machine and operating status.

Digital Twin can be viewed as an outcome of multiple research areas coming from Virtual Manufacturing systems, Model-based Predictive control (MPC), and Building Information Modelling (BIM) (Semeraro et al., 2021.) A theme in multiple different research articles and white papers was defining digital twin by conducting systematic literature reviews. According to the researchers Semeraro et al. (2021), the industry and academic world defines Digital Twin in a following simplification derived from their systematic literature review:

> A set of adaptive models that emulate the behavior of a physical system in a virtual system getting real time data to update itself along its life cycle. The digital twin replicates the physical system to predict failures and opportunities for changing, to prescribe real time actions for optimizing and/or mitigating unexpected events observing and evaluating the operating profile system (Semeraro et al., (2021).

Digital Twin can be broken down into smaller pieces, on what forms the digital twin itself. According to Jones et al. (2020), characterization of Digital Twin through different contexts of research fields would benefit the definition of Digital Twin as a concept. These fields are such as Information Modelling, Computer-Integrated Manufacturing, Virtual Manufacturing Systems, Model-Based Predictive Control, Advanced Control Systems, and Health Monitoring/Prognostics (Jones et al., 2020). Digital Twin has been also defined as the following according to Stark and Damerau (2019):

> A digital twin is a digital representation of an active unique product (real device, object, machine, service, or intangible asset) or unique product-service system (a system consisting of a product and a related service) that comprises its selected characteristics, properties, conditions, and behaviors by means of models, information, and data within a single or even across multiple life cycle phases (Stark and Damerau, 2019).

According to research conducted by Trauer et al. (2020), Digital Twin is characterized with the following attributes:

> 1. The Digital Twin is a virtual dynamic representation of a physical artefact or system, 2. Data is automatically and bidirectionally exchanged between the Digital Twin and the physical system, 3. The Twin entails data of all phases of the entire product lifecycle and is connected to all of them (Trauer et al., 2020).

Trauer et al. (2020) also identified the concept with the following words with their industry partner: "A Digital Twin is a virtual dynamic representation of a physical system, which is connected to it over the entire lifecycle for bidirectional data exchange."

In the literature, the Digital Twin is characterized by many definitions, and this has its challenges with defining the concept with a uniform definition. Boschert et al. (2018) argue, that the current literature has a quite broad definition of the concept "Digital Twin", and to better understand the concept it must be broken down into a set of models. According to Rosen et al. (2019), Digital Twin should be seen an open concept or umbrella term, where you can add pieces of information such as data and models along the whole lifecycle of the Twin.

2.2.2 Digital twin architecture

Due to the complexity of cyber-physical systems and thus the inevitable complexity of digital twins, digital twins' architecture can prove difficult to map depending on the domain and use case of the digital twin. In FIGURE 6, digital twin's architecture is conceptualized by Parrott and Warshaw (2017) below.

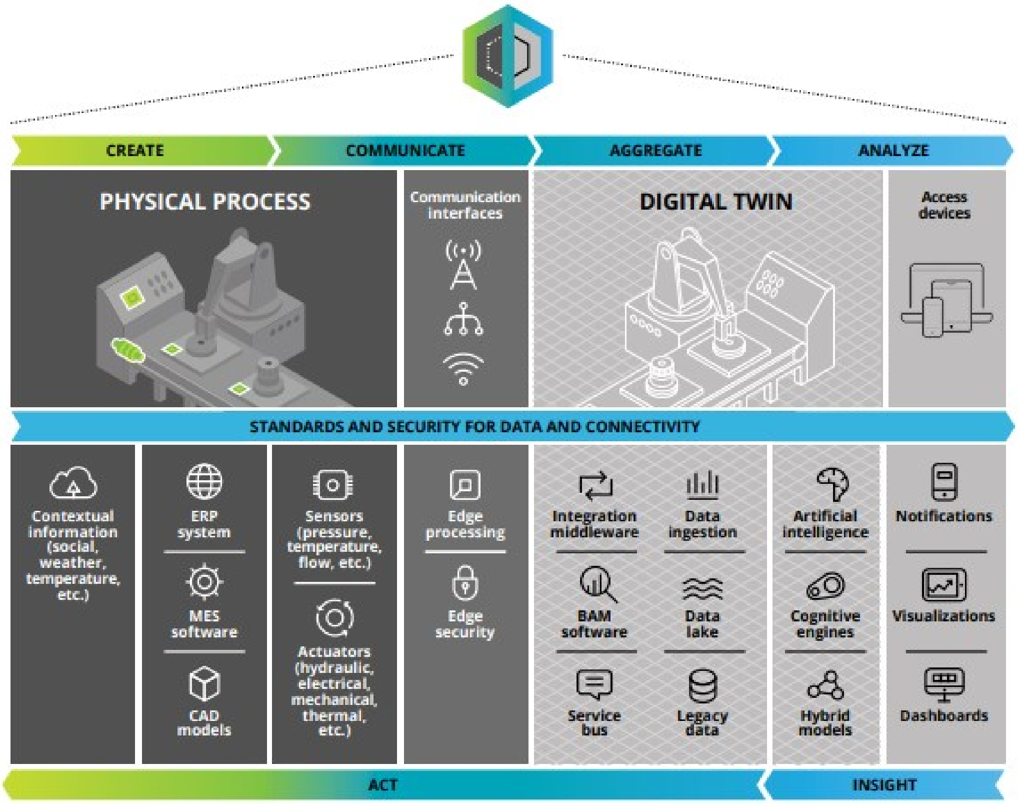

FIGURE 6 Digital twin conceptual architecture (Parrott and Warshaw, 2017)

To open the FIGURE 6, it is broken down into two main areas of the physical processes and the digital twin, which both are interconnected by the communication interfaces, edge processing and edge security. In FIGURE 6, conceptual architecture is laid out as a six-step process, which is broken down into the following steps:

1. Create – where the physical process is measured in critical inputs and putting in to operational and environmental measurements.
2. Communicate - This step of the process embodies the edge processing, communication interfaces and edge security. Using these the digital twin gains its network of connectivity and bidirectional interaction with the physical process – the physical twin.

3. Aggregate – this step is all about data being processed and refinement of data before analytics.
4. Analyze – visualization and analyzation of the data gained from the aggregation, which gives support to decision making through models, which are iterative.
5. Insight – improvement areas and information on the physical process are gained through the iterative models, which are visualized through various dashboards from the previous step.

6. Act – based on the insights and information provided through the whole five first steps, the process gains information which is fed through different actuators controlling the movement and control mechanisms.

Digital twin is used widely in correspondence with Industrial Internet-of-Things (IIoT). Harper et al. (2019) visioned digital twin's architecture in the IIoT to be deployable in various tiers of IIoT. Their expectations on the digital twin's architecture were evaluated on six different criteria: interoperability, information model, data exchange, administration, synchronization, and publish /subscribe. The expectations on digital twin's architecture's capabilities were the following according to Harper et al. (2019):

1. "App store deployment of configuration
2. Integrated information model
3. Flexible classification of types, properties, and instances
4. Encrypted data at rest and in transfer
5. Role-based access control configured for authenticated users
6. Data ingest configuration for each column store
7. CRUD (create, read, update, and delete) data exchange with cascading side effects based on role
8. Publish and subscribe notification of CRUD transactions
9. Filtered synchronization between tiers."

All the relations between the expectations and evaluation criteria applied in their interactions has been formed in TABLE 3 below, where the C's are from the expectations listed before according to Harper et al. (2019) and their relations to the six different interactions the same authors have proposed.

TABLE 3 Evaluation criteria applied to digital twin interactions (Harper et al., 2019)

	C1	C2	C3	C4	C5	C6	C7	C8	C9
Interoperability									
Information Model									
Data Exchange									
Administration									
Synchronization									
Publish/Subscribe									

2.2.3 Challenges of a Digital Twin

Even though the concept of the digital twin has been broadly studied and the emergence of the term itself dates to early 2000's giving it some maturity in the academia and industries, there remains challenges to this day for the digital twin. With digital twin tackling enormous amounts of data and bandwidth, relation to these comes its own problems in the likes of data analytics, IoT/IIoT, and digital twin's own challenges (Fuller et al., 2020). On FIGURE 7, which is based on Fuller et al. (2020), the challenges are broken down into smaller pieces to get a better understanding of the problems related to the digital twin. On FIGURE 7, Digital twin's challenges in data analytics and industrial IoT/IoT are shared in IT infrastructure, data, trust, privacy & security, and expectations - connectivity is only on IIoT/IoT side of the equation (Fuller et al., 2020).

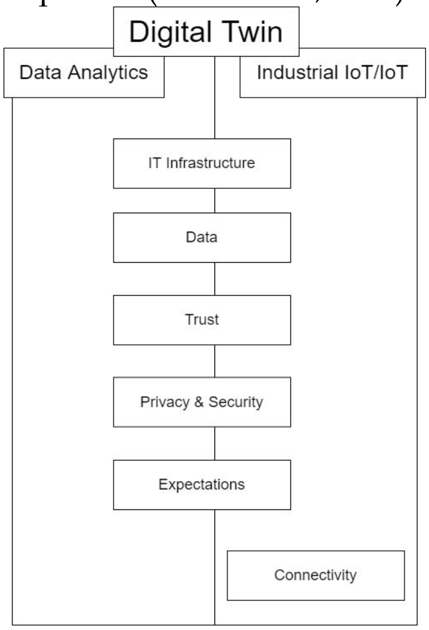

FIGURE 7 Digital twin's shared challenges on data analytics & IIoT/Iot based on Fuller et al. (2020)

Common challenges digital twins can have according to Rasheed et al. (2020), are the real-time connection between the twins and the computational resources needed to acquire this connection, the backwards compatibility between the twins while they both develop themselves, security & safety issues related to transparent decision-making, and lastly the ease of use for the end user or operator of the twins.

Research article on an overview on digital twins for oil & gas industry by Wanasinghe et al. (2020) summarized several challenges for the adaptation of digital twins in the industry: scope and focus; lack of standardization; cyber security; data ownership and sharing; accuracy and validity; functionality; unlocking experience; business model, people and policies; data storage and analytics; maintenance; and incremental vs disruptive.

Researchers Semeraro et al. (2021) concluded that the lack of universal architectural models and real-life systems' complexity make modelling reality in the digital twins hard. In the previous chapter, where the different architectural models for developing digital twins were laid out in FIGURE 6 doesn't give many concrete principles or technological solutions to choose from when developing the digital twin. This would support the conclusion Semeraro et al. (2021) came to.

2.2.4 Future of Digital Twin

In the previous chapter, the Digital Twin is given many of definitions to better understand the concept as an umbrella concept. To better understand the little differences of the Digital Twin's definitions, it is good to get a glimpse of the concept's future of a next-generational Digital Twin. Boschert et al. (2018) introduce a next generation Digital Twin as the next paradigm after the digital twin. The paradigm suggests that the next generation Digital Twin links multiple elements from all lifecycle phases and systems, to improve the efficiency and performance of the whole operation of the system (Boschert et al., 2018).

Rosen et al. (2019) research suggests, that next generation Digital Twin (also formulated with nexDT abrevation) will be a value network with it's own ecosystem, where the ecosystem can be expanded or shrunk depending on the needs from the whole ecosystem. The next-generation digital twin is illustrated in FIGURE 8 according to Rosen et al. (2019).

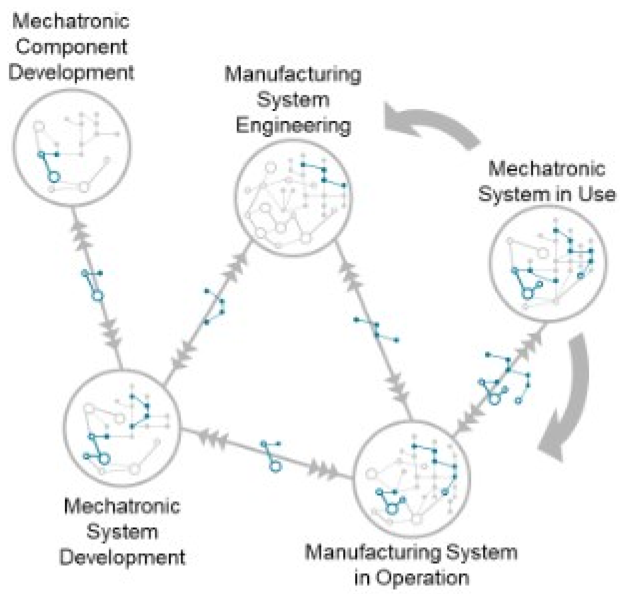

FIGURE 8 NexDT forms a value network and creates an own ecosystem (Rosen et al., 2019)

In FIGURE 8, the blue drawings are representing digital twins. This means, that a whole ecosystem, or value network can have multiple different digital twins in the system handling various tasks or responsibilities. The same authors Rosen et al. (2019) illustrate the architectural setting for the nexDT idea with the following schematics in FIGURE 9 below.

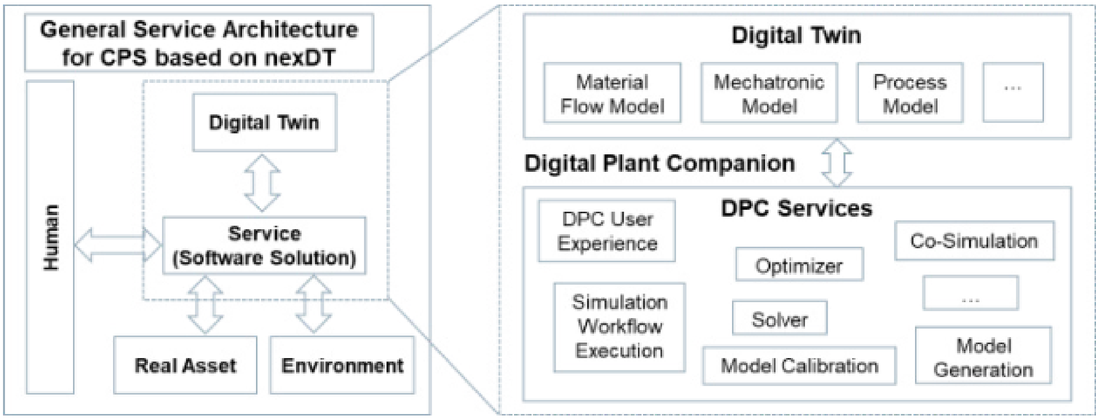

FIGURE 9 Schematic correlation between the DPC and Digital Twin (Rosen et al., 2019)

In the FIGURE 9, we can see, that the digital twin is intertwined with the digital plant companion, which illustrates the different benefits a digital twin is supposed to give based on the academical literature this thesis has cited before. According to Rosen et al. (2019), the digital plant companion is illustrating the benefits for the future's manufacturing factories or process-based plants, and this companion wouldn't be possible without the digital twin.

3 DIGITAL PRODUCT DEVELOPMENT

In this chapter the research's foundation on digital product development based on literature is laid out. Digital product development practices are focused on the MVP models and development frameworks, which support the MVP model product development framework.

3.1 Lean Startup Method

Lean startup method is one of the agile product development frameworks, which has the minimum viable product philosophy in it. Lean Startup Method is a framework for developing products and services, which is based on the feedback loop called build-measure-learn (Ries, 2011). This feedback loop is presented in the FIGURE 10. Ries' (2011) Lean Startup model encourages startups to experiment with the most barebones product concept to gain information on the product's market response. Minimum viable product can be seen as a proof-of-concept to test the market, is there any demand for the initial product. Ries' (2011) build-measure-learn feedback loop provides a framework for the development phase of product. This framework can be used to analyze the product's initial market response by analyzing and learning from the feedback the product receives from the market.

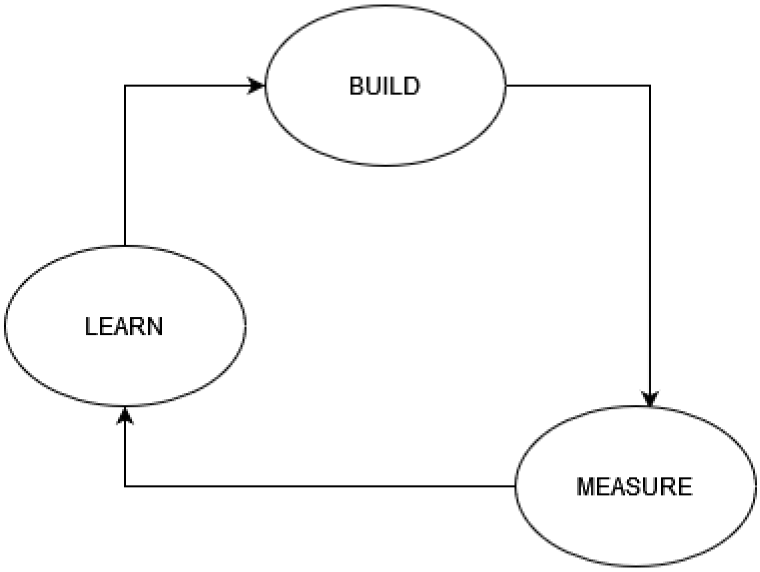

FIGURE 10 Build-measure-learn feedback loop based on Ries (2011)

Lean startup approach has been firstly used mainly for startups or non-matured companies developing their first products. This doesn't mean it is not useful for more mature and established companies. In Scheuenstuhl et al. (2020) study, the researchers found that lean startup approach performed better compared to the traditional innovation process in more established companies, when the performance metric was on financial output of the process.

Lean Startup Method has gained a lot of praise from the industry and its cut-throat approach to developing products, especially in its MVP framework. Lean Startup Method has its limitations regarding product development and innovation process, where the Lean Startup Model's process is not understood in the right context or is used wrongly (Yordanova, 2018). Many practitioners using the Lean Startup Model does not have fully grasped the limitations of the model either. According to Yordanova (2018), "The research proves that actually experts and companies that took part in the research and use the LSM are not fully aware of the method's disadvantages". This in turn means, that it is important to understand the limitations and advantages of Lean Startup Method to utilize it to its fullest potential.

Ries (2011) highlights that the Build-Measure-Learn loop needs to go full circle faster, rather than spend time within the loop. This means you should maximize the number of circles you go through with your product instead honing the product into the small details within the first iteration circle.

The size of the company and matureness in the industry also influences the successful utilization of lean startup method for the company. For example, lean startup method for a more mature company has been characterized as mercury business in a study conducted by Järvinen et al. (2014). The key differences between the two – lean startup method vs. mercury business is the following according to Järvinen et al. (2014):

> Mercury business is closely related to the Lean startup framework, and in many ways the Lean startup has been an inspiration for Mercury business. However, while Lean startup is about the creation of a new company and the definition of its products, Mercury business aims at transforming and extending already existing businesses, which requires a different approach (Järvinen et al., 2014).

One could also argue that the startup approach presented in the lean startup method can be thought of as an inner entrepreneurship formed in the companies. This can be rephrased as team members being the entrepreneurs within the company seeking new business opportunities and ways to create value to the customer. According to Ardichvili et al. (2003), the entrepreneur creates successful business from opportunities, which are evaluated multiple times to find the winning combination for the business.

3.1.1 Minimum Viable Product

In today's ever-evolving landscape of digital services, it is important to iterate your product rather than try to perfect your product during development. By introducing the minimum viable product framework into the product development, you start by creating some value for the customer in the minimum viable product and iterate your product based on the customer's feedback. According to Girgenti et al. (2016), market's future supply and demand are considered when new products and services are developed. In a study conducted by Lenarduzzi and Taibi (2016), they identified following minimum functionalities for the MVP's purpose:

- "To allow the product to be deployed
- To target market opportunities
- To create a viable product for the customer
- To test the fundamental business hypothesis
- To allow to test the product in the market
- To gather customer feedback
- To identify the most viable features by iteratively experimenting the market"

Startups and the lean startup model have their own challenges. Usually, the startups' MVP's lack resources to be developed to their fullest potential of product market fit (Dennehy et al., 2016). According to Freeman and Engel (2007), "in a one-on-one competition, the startup usually has less capital, fewer scientists and engineers, less legitimacy or brand presence, fewer strategic alliances, evolving organizational structures, and incomplete or even non-existent business processes". This further suggests that more mature and bigger companies have better starting positions to develop their MVPs.

Also, the bigger companies may have an edge on the supporting factors for the MVP development. According to a study conducted by Tripathi et al. (2019), "a constructive startup ecosystem around software startups can boost up the creation of an effective MVP to test product ideas and find a product-market fit". This would solve the problem around the lack of resources related to the startup's capital. Bigger and more established companies have more resources to seize the opportunities MVP's give.

3.2 The Early-Stage Software Startup Development Model

Early-Stage Software Startup Development Model is a combination of the best practices in the lean philosophy with a focus on giving the tools for the industries practitioners for product development (Bosch et al., 2013). This model is adapted in the FIGURE 11 based on Bosch et al. (2013), where the process of developing scalable products is laid out in the following way:

1. Idea generation
2. Backlog of product ideas
3. Funneling the ideas systematically through multiple build-measure-learn loops
4. Scaling the product/service through validation from the BML loop

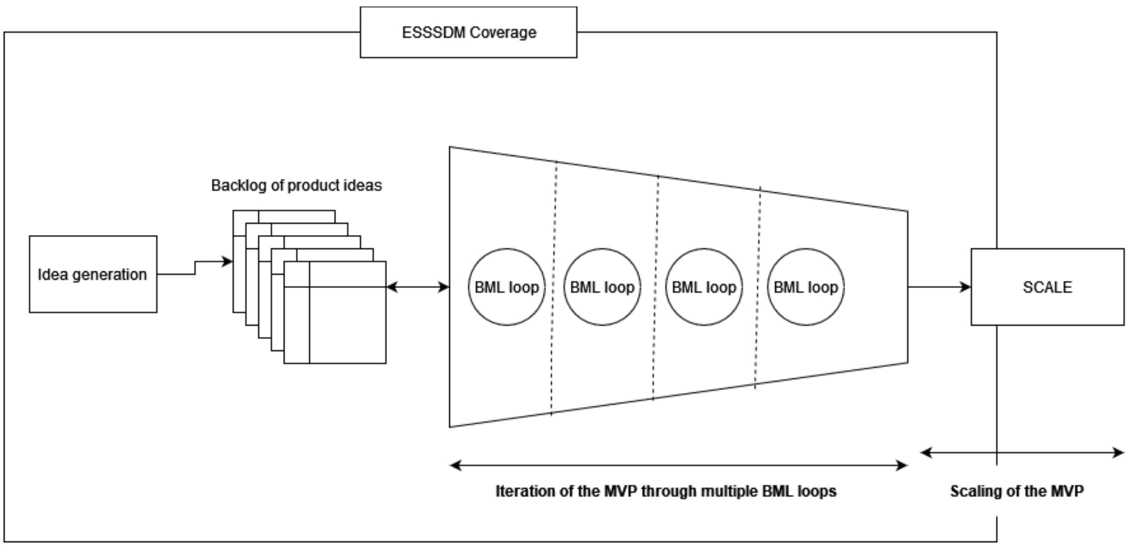

FIGURE 11 The Early Stage Software Startup Development Model (ESSSDM) adapted from Bosch et al. (2013)

This thesis is focusing on the product ideas and BML (build-measure-learn) loop's connection and how to derive the different requirements to get the product ideas from the backlog to the funnel and into validation via the BML loop. The ESSSDM model's foundation lies in the process of getting ideas to through the process of iterations until the final product is ready for scaling. In the FIGURE 11, the 3rd step of funneling the ideas is the key in finding the MVP's keys for success. By testing and iterating multiple rounds instead of perfecting one solution, it will provide better test for the market's response. The funnel in FIGURE 11 consists of four stages according to Bosch et al. (2013):

1. Validate problem
2. Validate solution

3. Validate Minimum Viable Product small-scale
4. Validate Minimum Viable Product large-scale

Overall, ESSSDM model provides a clear framework for the software development using the build-measure-learn feedback loop. ESSSDM model will provide a clear framework for the initial forming of the interviews and the framework for the empirical research.

3.3 Continuous experimentation

Continuous experimentation could be seen as another product development tool, which builds on top of Eric Ries' (2011) Build-Measure-Learn feedback loop illustrated in the FIGURE 10 earlier. According to Schermann et al. (2018a), "continuous experimentation is an up-and coming technique for requirements engineering and testing, particularly for Web-based systems".

Due to digital twin being a virtual reality entity, a software-based solution, it is important to view the development of one within the scope of digital product development. To find the technological requirements for the practical implementation of digital twin, continuous experimentation can be considered as one of the further development frameworks for the development of the digital twin later in the pipeline. Continuous experimentation can be seen as a parallel framework for the build-measure-learn loop, where the continuous improvement of the product is key for both. Techniques for further development and implementation of continuous experimentation are identified as code-level techniques and deployment-based techniques (Schermann et al., 2018a). This means that to start the implementation of the digital twin from the beginning, you need to know the possible end goals the digital twin will serve once in operation.

To keep continuous experimentation constant throughout the whole lifecycle of the digital twin, it is important to be able to see the digital twin's end form, which then gives the eventual founding technological requirements for the potential MVP as well. According to a study conducted by Schermann et al. (2018b), companies usually rely more on intuition rather than data and actual information, when using continuous experimentation as a development process. This would support, that the continuous experimentation needs to have clear goals regarding the digital twin's development and implementation.

Digital twins are usually part of a bigger project, where digital twin is one of the components of the value-creation chain. To receive feedback from the customer and keep improving the product, it is necessary to have a framework for the iteration in the future. This makes a solution more like a service, even though it can be first considered more as a physical product. Transition from a product-based business to more service-based model has been the trend during recent years (Bosch, 2012).

Depending on the scale of the digital twin, it is important to note criticism towards continuous experimentation and the challenges related to the framework as a development model. According to Fagerhol et al. (2017):

> In some cases, an experimental approach may not be suitable at all. For example, certain kinds of life-critical software or software that is used in environments where experimentation is prohibitively expensive, may preclude the use of experiments as a method of validation (Fagerhol et al., 2017).

At first glance, this could mean further development of the MVP of a digital twin with this continuous experimentation would prove hard due to the expensiveness of the project upfront. However, continuous experimentation could be seen as a framework later down the line in the product development, when the digital twin's MVP has been validated in the Ries' (2011) Build-Measure-Learn feedback loop illustrated in the FIGURE 10. According to Mattos et al. (2017):

> Experiments in the field are used in a problem-solving process to drive both innovation and optimization of post-deployed systems. Companies are moving towards to experiment-based development, where experiments support the decision-making process. Several challenges, such as resources, the experiment architecture and novel engineering approaches arise when running experiments in a large scale (Mattos et al., 2017).

Due to the topic of this thesis having to do with implementations of digital twin, it is important to take this limitation in to consideration. Still, the use of ESSSDM model by Bosch et al. (2013) together with continuous experimentation has benefits for product development later down the line, so it needs to be accounted for. ESSSDM is useful in uncertain situations, where there are many adjacent product development ideas floating around and being explored parallel at the same time (Bosch et al., 2013).

Continuous experimentation has been rising as a method for testing and identifying requirements for engineering, especially in web-based solutions used by the likes of Google and Facebook for example (Schermann et al., 2018a). Challenges and opportunities regarding continuous experimentation for cyber-physical systems have been identified as hardware constraints, feedback data, safety guarantees, involving more stakeholders and supportive instruments for continuous experimentation (Giaimo et al., 2016). This would support the findings in the challenges of cyber-physical systems chapter on the challenges regarding development of cyber-physical systems.

4 FORMING THE RESEARCH FRAMEWORK

In this chapter this thesis will focus on building the framework from the literature review's research and aim to bridge the gap between the different industries which would benefit from the implementation of a digital twin. The research framework will be a combination of multiple aspects including the goals company has on developing a digital twin's, what the academical literature has to say on the different topics, and how these all can be combined pooled into various findings for the different perspectives and challenges on implementation of the digital twin. The research framework will be targeting the idea generation phase of the ESSSDM model illustrated previously in FIGURE 11.

The goal of the thesis' empirical part is to understand the requirements of a digital twin and to better understand the possible bumps and roadblocks on the road for the development of the digital twin as a service and/or product, when starting from a clean table as a company. Product development can be tackled from various angles, and in this thesis the chosen product development framework is Ries' (2011) lean startup method's minimum viable product, which is further developed to an early-stage software startup development model based on the authors Bosch et al. (2013), which is illustrated in the FIGURE 11 previously. In the FIGURE 11 according to Bosch et al. (2013), the first step in the ESSSDM is idea generation, which could be done for example as exploratory interviews. Step two before the funnel is the backlog, which acts as an idea box, which is then prioritized based on the customers' most significant problems and how the startup will solve them most efficiently as possible (Bosch et al., 2013). Bosch et al. (2013) name exploratory interviews as one of the data gathering methods as well in the idea generation phase of software product development.

4.1 Getting started with the digital twin

Digital twins as a concept were opened in the chapter 2.2. previously. Due to the digital twin being a rather complex concept from a development perspective, and the chosen product development framework is the MVP framework, it is better to breakdown the development into different phases. This thesis will focus on seeking the requirements and specifications from the various customers from a wide selection of industries, and thus the research's focus will be on the ideation phase.

One of the potential frameworks for the actual development roadmap of the digital twin in the bigger picture can be based on the framework by Parrott and Warshaw (2017). In FIGURE 12, Parrott and Warshaw (2017) presents an overview of getting started with the digital twin. This model could be used as an example for the development of the digital twin's MVP.

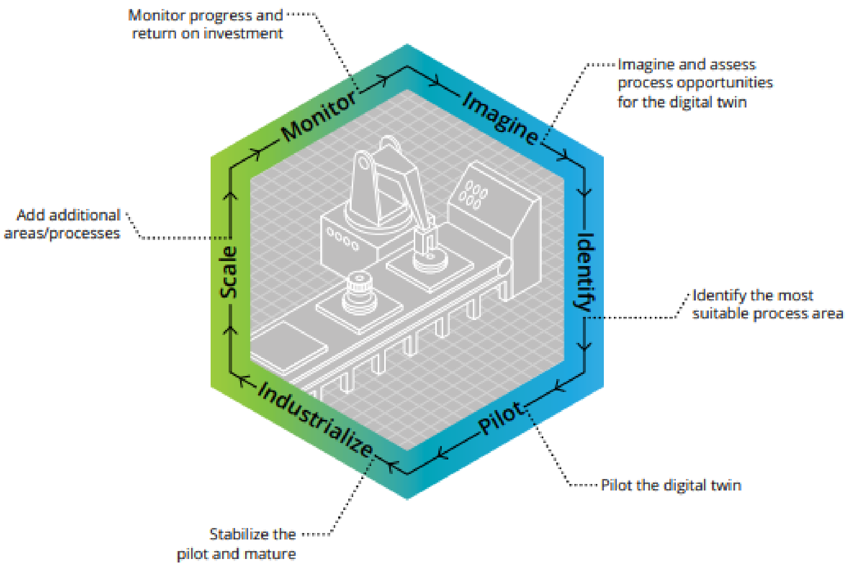

FIGURE 12 An overview of the getting started with the digital twin (Parrott & Warshaw, 2017)

In the FIGURE 12, Parrott and Warshaw (2017) lay out the ground of getting started with the digital twin. The key is not being too simple or too complex to get the ball rolling with the digital twin's development. The overview flows in the following order: imagine the possibilities, identify the process, pilot a program, industrialize the process, scale the twin & monitor and measure. The key in this six-step process is the constant iteration of the digital twin through feedback and information gained from the process itself (Parrott and Warshaw, 2017). This resembles the same logic of Build-measure-learn feedback loop by Ries (2011).

As this thesis focuses more on the requirements and specifications setting for the MVP model of a digital twin rather than the actual development of the digital twin itself, it is important to only focus on the relevant steps in the process.

4.2 Foundations for the empirical data gathering

In FIGURE 12 by Parrott and Warshaw (2017), the first and second steps in getting started with the digital twin are imagine the possibilities and identify the process. These are very similar to Bosch et al. (2013) FIGURE 11's first and second steps. Overall, the resemblance of the two is quite clear in terms of how processes flow from start to end. Combining these two and focusing on the early steps gives a good foundation for the next step in this research. These two steps will be the areas focused on the questionnaire. Through the questionnaire, this thesis will form the requirements and specifications for the starting of development of a digital twin's MVP model for company. In the next TABLE 4 the two different frameworks are put side by side and combined.

TABLE 4 Side-by-side comparison of two different approach development models from Parrott and Warshaw (2017) and Bosch et al. (2013)

Step of the process	Parrott and Warshaw (2017)	Bosch et al. (2013)
1	Imagine the possibilities	Idea generation
2	Identify the processes	Backlog of product ideas
3	Pilot a program	Funneling ideas systematically through multiple BLM loops
4	Industrialize the process	
5	Scale the twin	Scaling the product through validation from the BML loop
6	Monitor and measure	

The focus on this thesis is on the first two steps in the TABLE 4. According to Parrott and Warshaw (2017), the first step of imagine the possibilities consists of having various scenarios, where the digital twin would be beneficial for the organization and circumstance, where usually these two key aspects are fulfilled; the digital twin is beneficial financially to build & it brings value to the customer or organization by resolving outstanding issues. Second step is focusing on the identifying processes and areas which have ability to scale and enough broadness to deliver value to the whole organization (Parrott & Warshaw, 2017).

The process described by Bosch, Holmström, Björk & Ljungblad (2013) starts from the idea generation step, where usually the process gets its beginning from the company's desire to bring a new product or expand existing product portfolio. Bosch, Holmström, Björk & Ljungblad (2013) name one of the idea generation techniques in form of exploratory interviews where: "one way to extract problems from potential customers is to go out and talk with them… The purpose is to understand how potential customers run their businesses, and what problems they experience". Bosch et al. (2013) model's step two is the backlog, where potential products are stored and given priorities, where these different products can be compared based on these set priorities, while working on the multiple ideas in parallel.

According to Bosch et al. (2013), "ESSDM supports working on, or investigating, multiple product ideas in parallel, as part of an idea portfolio". This idea portfolio in this thesis is where the different product ideas are the ways various industries may utilize digital twin in their operations. Perspectives and challenges of implementing digital twins are thus gotten by interviewing different industries on their different pain points as Bosch et al. (2013) highlighted in their step 1 of ESSSDM model. This exploring multiple ideas in parallel is supported as well in the Parrott and Warshaw's (2017) article's step two, where it is key to identify the areas and configuration where the possible pivot would be most successful. Here the pivot could be seen as the MVP model, which is funneled through the BML loop as the TABLE 4 previously compared between the two different approaches.

4.3 Summary

This chapter will sum up the research framework. The goal of this thesis is to find different perspectives and challenges on the implementation of digital twins. The literature review provides answers to some of the research questions, but some will need to gather empirical data from the practical field – various industries – on how they are utilizing different digital twin technologies. The research framework opened in the chapter 4 and it's sub-chapters are the reasoning for the questionnaire and the different themes it will explore. Through the questionnaire this thesis aims to find answers to the research questions and findings of the possible requirements and specifications for the digital twin's MVP model, which would begin development based on the different findings in this thesis. The background for the research framework on empirical data is based on the combination of ESSSDM model by Bosch et al. (2013) and Parrott and Warshaw's (2017) approach to starting with the digital twin.

5 RESEARCH DESIGN

This chapter focuses on the research design and lays the foundations for the thesis' questionnaire, the data collection, and data analysis.

5.1 Goals of the empirical research

Goal of this research is to bridge the gap between academical literature and industry on the topic of implementing digital twins, and it's perspectives & challenges. By combining the information gathered from academic sources and forming the foundation of the questionnaire based on the literature review, the companies being interview can support the topic of this thesis through empirical research. The primary research question this study aims to answer was the following:

- **How to implement a digital twin in practice?**

And the supporting research questions were:

- **From what different components are digital twins built from?**
- **What are the challenges of implementing digital twins?**
- **How is the industry currently utilizing digital twin technologies?**

By getting the answers for these questions from practice and industry's professionals, we get insight the academic literature in its current form lacks. While the topic is kind of broad in a sense, the goal is to understand multiple different industries take on the digital twin. This way a broad topic on this thesis supports the gathering of various implementations, challenges, and perspectives regarding digital twins in industry. By staying broad enough, it gives relative directions where to head based on the potential customers' needs. The value comes from the knowledge and knowhow of industry's professionals, and these professionals have the knowhow of what their current challenges are and where they need possible help with. Due to digital twin being relatively broad concept and different depending on who you are talking to in the academia or industry, the literature review gives a good background on how to form the questions for the initial thematic interview.

5.2 Forming the questionnaire for interviews

After conducting the literature review, this thesis is starting on the step 1: Idea Generation of the ESSSDM model by Bosch et al. (2013), which is illustrated in FIGURE 11. According to Bosch et al. (2013), exploratory interviews can be one way to generate and form ideas before creating the new product:

> One way to extract problems from potential customers is to go out and talk with them. It is recommended to investigate one customer segment at a time, so that the team stays focused and dig deep within each segment. The purpose is to understand how potential customers run their businesses, and what problems they experience (Bosch et al., 2013)

The background for the exploratory interview questions was firstly brainstormed based on the academic literature, where the definitions for following concepts where searched – minimum viable product, cyber-physical systems, and digital twins. After conducting the literature review on the listed concepts, the questionnaire was thought out together with the thesis' instructors. The key areas (dimensions) of the interview are opened in the subchapters.

The order of the questionnaire was laid out on how to get the most efficient and coherent answers to the research questions to be able to go from step 1: idea generation in the ESSSDM model to step 2: The Backlog, where the initial product ideas will be stored (Bosch et al., 2013). Before starting the creation of digital twin's MVP, it was crucial to understand the customer's problems in their corresponding businesses. Questions were formed based on the emphasis, that we could understand better the customer's problems, and find the possible pain points to address with the initial MVP.

Firstly, in the interview, the person's background and field of expertise is thought out. After that follows the figuring out how the person deals with the critical information related to the machine and how do they collect the data. From collecting the data, we move on to the modelling and handling of the data, to get the requirements for the whole process how the data gets processed to information. Monitoring and decision-making follow the collection of the data as themes. This then moves on how the processed data as information is used in the monitoring of the machine and decision-making as well. Lastly, the overview and digital image (digital twin) is discussed and how it has been relevant to the company in their business, and how it has given them possible business value.

It should be noted that the ESSSDM in FIGURE 11 adapted Bosch et al. (2013) together with Parrott and Warshaw (2017) starting with the digital twin, were the founding footings for building the interview's questionnare. Before the interviews, this thesis is on the Backlog phase of the ESSSDM model. By proceeding to the interviews before initializing the build-measure-learn loop, this thesis aims to get the core problems customers are battling with currently in their respected business and industries. To get the first MVP built, it is necessary to understand the pain points and current situations the customers are in with their machines and digital twins.

Lastly, these questions were then vetted through the company's various business unit's leaders. This vetting process gave feedback on the questions and the leaders also pointed out possible candidates, who to interview regarding the thematic interview. Next up in the sub-chapters are the different dimensions in the questionnaire's questions, which will be briefly introduced.

5.2.1 1st dimension of the questions

Person's field of expertise

First the funnel of the questionnaire aims to find the interviewee's background and position related to the machine/product the company manufactures. These questions also address the understanding of critical information, and how this person is dealing with critical information in their job's responsibilities. This dimension aims to understand the person and their definitions of critical information, machines, and products.

5.2.2 2nd dimension of the questions

Machines' critical information

2nd dimension of the questions starts to map the technological requirements for the MVP of the digital twin. Even-though the questions are not laid out in the form of "what are the technical requirements…", these questions provide actual information that can be utilized for the development of the MVP, since these answers provide broad requirements themselves. It is also important to understand, what possible information the case companies would like to gather and have, but are not currently able to get from the machines. This further elaborates the pain points of potential customers and gives information for the development proof-of-concept or MVP.

Data & models

Also, one of the requirements comes from the data and model's requirements. Since digital twin will be needing vast amounts of data to be able to simulate the real physical twin counterparts, it is important to know how the companies utilize their current gathering of data and analysis. This will give a footing on the digital twin's architectural layer. How the potential customers utilize their different models for the data layer is important to understand as well.

5.2.3 3rd dimension of the questions

Monitoring & decision-making

On the 3rd dimension of the questions, we delve deeper into to the operational side. For the customer to get the value from the digital twin, we need to find out what problems they aim to solve on decision-making level. Before the decision-making process, the monitoring of the machine's critical information happens, and this will lay the foundation for the actual decision-making.

Overview & digital image (digital twin)

The overview & digital image (digital twin) aims to get the answers on how the company utilizes their already existing solutions for the machine's critical information overview in a possible digital image or digital twin. By getting the answers for the customers problems they are facing with their digital images & twins, we can provide the possible solutions in the first iteration of the MVP's model. Of course, we must remember this thesis only aims to get the initial requirements for the technological implementation of the digital twin.

5.3 Data collection

Data collection was done utilizing semi-structured thematic interviews. The formation of the questionnaire for the interviews and structure was opened previously in the chapter 5.2. This questionnaire was sent to the interviewees before the interview took place to have enough preparation for the upcoming interview.

The interviews took place in Microsoft Teams, and the reserved time for the interview setting was an hour. This hour consisted of the possible questions risen before the interview, introduction of both parties in the interview, the recorded interview, and possible further questions.

5.3.1 Selection of companies

Companies were selected based on their size and potential to have adaptation of digital twin in their businesses. Ideal interviewees for the study were thought out to be different R&D directors, directors working with digitalization, digital twin product owners, or people who had the most knowledge related to digital twin's utilization within the organization. Targeted people were ones, who could elaborate the whole architectural layer from the bottom to the most upper levels. It was crucial that the interviewees being contacted had the experience and knowledge of their responsible product/machine and understood the concept of the digital twin as well.

As previously stated on chapter 5.2 how the questionnaire for the data collection was formed, the company's same business unit leaders gave potential leads, who they had contact with and connected the dots between the right people from the lead companies. Also, Google was utilized in the searching for major big Finnish corporations across multiple different industries, with the key words for the Google search being "digital twin [company name]". This usually gave a blog article for the searched company with the key personnel in adaptation of digital twin having their contact details in the blog article.

Interviewed companies were selected from a wide range of different industries and manufacturing areas. This was a conscious choice, to gather a wide variety of answers, to cover different implementations of digital twins and needs for different industries. By covering this wide range of industries, it is expected before the empirical findings, that with more diverse industrial coverage of the companies and different domains, it will yield a better understanding of the adaptation of digital twin in the respected industries, and what is common with all those industries' digital twin adaptations. The same logic is expected to happen with the different pain points potential customers have in their current operations, that the different industries will give a broader picture of their respected problems. TABLE 5 below illustrates the key performance indexes of interviewed companies.

TABLE 5 Scope of the inteviewed companies based on their performance in 2021

ORGANIZATION	REVENUE (2021, MILLION EUROS)	NO. OF EMPLOYEES (2021)	OPERATING RANGE BY COUNTRIES
CASE ORG. 1	550-600	900-999	10
CASE ORG. 2	9500-9999	16000-18000	12
CASE ORG. 3	1500-2000	15000-17000	50
CASE ORG. 4	90-100	100-150	9
CASE ORG 5.	25999-29999	100000-150000	100+

These companies presented in the TABLE 5 were the ones who agreed to do the interview, from the contacted companies, which were in the tens. The timing of the interviews was quite late in the year, mid-November to early January, which meant that the interviewees were hard to get ahold of and find a slot in both calendars due to Christmas holidays.

5.4 Data analysis

The interviews form a base for the empirical findings in this study. This thesis will utilize thematic synthesis in the data analysis. According to Cruzes et al. (2015) the strengths for thematic analysis are flexibility for the reviewer, supports divergent sets of evidence, and theory can be drawn from the analysis, in utilizing a thematic analysis as the data analysis method in a case study. As previously highlighted in the company selection chapter 5.3.1., the data set is purposefully taken from multiple angles, and this would support the utilization of thematic synthesis as the data analysis method for this thesis.

Thematic synthesis can be done in software engineering utilizing three different approaches to coding the data: deductive, inductive or combination of both, integrated approach (Cruzes and Dybå, 2011). However, this data analysis method doesn't come without flaws. According to Cruzes and Dybå (2011), these topics need to be assessed by the researcher(s): coding can be too general, implying own thoughts on the data rather than relying on the data itself, out of context coding.

This master's thesis' data analysis will utilize integrated approach. As digital twin is a vast concept, the coding of the data first started on the conceptual architecture of digital twin based on Parrott and Warshaw (2017) paper. Due to the digital twin's research lack of real-world applications and documentations, inductive coding approach allows the data to broaden the analysis to relevant topics. As the empirical data analysis aims to answer the industry's various implementations, perspectives and challenges, the data's coding will begin by setting the deductive coding through themes structured in the questionnaire's dimensions. The basis for the deductive coding is the following themes: challenges, opportunities, technological capabilities, and technological limitations. The inductive coding was drawn from the interviews to expand the knowledge of the companies currently utilizing digital twins or digital twin's components in their operations.

6 EMPIRICAL FINDINGS

In this chapter, this thesis will conduct a thematic synthesis of the empirical findings from the interviews of multiple companies from various domains of industries. The case study had a total of 5 organizations participating in the interviews. This was sufficient of a set of data, as the answers started to have similar themes in them. Even though same themes started to get repeated in the answers, the answers still had some variety in them due to the different industries and domains the interviewed organizations operated in.

The interviews gave a lot of text and video material to process. The number of pages was higher first, when pages consisted of the transcribed text from the interviews, which were firstly transcribed through Microsoft Teams' own transcriptive tools. After the text was refined to usable form, the total number of pages was 35 pages of material from the interviews.

6.1 Overview

In the analysis, the goal was to better understand the current problems potential customer companies are facing in their daily operations who are utilizing some sort of digital image or digital twin in the solving of these said problems. Also, the current infrastructure around their digital solutions was one of the topics of interest. The goal of the research framework was opened in chapter 4, where the goal of the framework is to find validation for digital twin's MVP requirements from multiple different industries through their various pain points and opportunities digital twins are supposed to solve. These pain points & opportunities were drawn from the interviews through various themes: challenges, opportunities, technological capabilities, and technological limitations. TABLE 6 represents assigned codes and their occurrences below.

TABLE 6 Assigned coding and their occurences in the data

Deductive code	Inductive code	Occurrences
Challenges	Status of the machine	2
Challenges	Lack of required information	2
Challenges	Real-time status of the machine	3
Opportunities	Improved safety	3
Opportunities	Product development	4
Opportunities	Improving servicing	5
Opportunities	Cost-optimization	5
Opportunities	Optimization of use	2
Opportunities	Process-optimization	4
Opportunities	Improved troubleshooting	5
Opportunities	Optimization of components	1
Technological capabilities	Digital twin	2
Technological capabilities	Virtual Reality	3
Technological capabilities	Automation systems	5
Technological capabilities	Visualization	5
Technological capabilities	Cloud-technologies	3
Technological capabilities	Internet of Things	2
Technological capabilities	Simulation	5
Technological limitations	Cloud-technologies	1
Technological limitations	Too high investment cost	2
Technological limitations	Lack of knowhow	3

The thematic analysis was a success in terms of finding different technological capabilities. It revealed a lot of the current landscape in the different industries how they are solving different problems and how these current technological applications could be utilized in terms of building blocks for digital twin. Opportunities compliment the technological capabilities, and this would support the digital twin's MVP building blocks being already present in the different industries. However, the challenges and technological limitations cannot be ignored here. Challenges mostly consisted of current ones faced in the operations, not as challenges on implementing digital twin, which are more on the technological limitations side. Two of the interviewees reported, that organization had the digital twin already in use or being built. Thus, we can form the first two primary empirical conclusions based on the analysis.

PEC1: Some of the interviewed companies reported having technological capabilities to form digital twins.

PEC2: All the interviewed reported opportunities on the following areas on utilization of digital twin or like solutions: cost-optimization, improved servicing, and improved troubleshooting.

Another empirical conclusion can be drawn from the analysis as well on the utilization of cloud technologies.

EC1: Cloud-technologies are an enabler and a limiting factor for forming a digital twin.

6.2 Results the questionnaire's answers

In this chapter we will draw a thematic analysis. Questionnaire got a wide array of answers. After the interviews, in multiple occasions the interviewee gave positive feedback on the study subject being relevant in their field of expertise, and to the future to come as well.

While conducting the interviews, it was at times apparent that the expertise and deeper domain knowledge of the interviewee had a big impact in the answers to the questions. If the interviewee had only surface-level information regarding the interviewed topic, it was hard to get them to answer the deeper questions in the questionnaire and if the domain-knowledge was deep the answers followed suit. This lead to the following empirical contribution:

EC2: Person's role is highly correlated on the understanding of digital twin.

6.2.1 Person's field of expertise

In the first-dimension questions, most of the interviewees had varying answers, where most of the differentiation came from their job positions and the machines the company was involved in or manufactured. The machines were ranging from mining machinery, maritime industry, to forestry machines, and manufacturing plants were represented with different end products as well. The different operating areas of the machines were also quite broad ranging from outdoor extreme weather conditions to stationary factory machinery. This gave the study a wide range of industries, machines, and user environments to get information on in the answers.

The first dimension of the questionnaire aimed to get the interviewee's domain information, and via this give background for the possible discussion within the questions to get the deeper knowledge from the next dimension questions. Important note from the stakeholder question was that many of the interviewees had multiple inner and outer stakeholders, whom they were acting within their daily jobs. These stakeholders they were interacting with varied a lot based on the job positions, and the responsibilities they had in their daily jobs. The answers had similar patterns, which can be seen from the answers to the question 2:

> "I work with inner and outer stakeholders and between different departments in the company. I work with different engineering, designer, measurement teams. Also, with and within software development teams related to simulation, which extends to our suppliers and subcontractors. I co-operate with marketing and maintenance departments as well. Customers can't be forgotten either." – Org 1

Another interviewee stated the following:

> "We have multiple stakeholders, where the most important ones are my colleagues in the mechanical department. We have project managers leading their own disciplines within the project, which I co-operate with. Also, different suppliers, subcontractors and teams depending on the state of the factory project. For example, my range of responsibilities is demonstrated by the number of contracts (20) I am responsible for regarding the suppliers for the project." – Org 2

Based on these we can form the following empirical conclusion:

> EC3: Depending on the job position, people interact with multiple different inner and outer stakeholders on their job.

The variety in finding people with different job positions and responsibilities gave good support on getting as large sweep through multiple industries. The companies, which participated for the interviews had multiple different business areas and sectors the companies operated with in. Due to this, the interviewees' answers regarding their job positions title and area of machines or operation are listed in the TABLE 7 below to shed light on the domains these interviewees are experts on. The titles of their job positions are broader than the correct positions to keep anonymity.

TABLE 7 Interviewees' job positions and domains/machines they operate within their positions

Case Organisation	**Job Title**	**Domain or machine(s) they work with**
Org 1	Engineering Manager	Forestry machinery
Org 2	Project Manager	Biorefinery factory
Org 3	Director	Mining machinery
Org 4	Data Scientist	Wire, cable, pipe & tube manufacturing machinery
Org 5	R&D Product manager & project manager	Maritime industry

As we can see from the TABLE 7, many of the interviewees were high position decision makers in their respective areas. Since the job titles remain quite broad to keep anonymity, a first glance can give you a false representation of the responsibility the interviewees had in their position. During the interview it was apparent, that the interviewees were very knowledgeable in their field, but the knowledge and know-how of digital twins was hard to get – even though the persons were pointed out to be the most ideal persons for the interview according to the organizations contacted. The domains and environments machines and factory machinery operated in varied depending on the use-case of the machine and was it static or mobile. Most of the interviewees' reported that their machinery was operating outside, and the rest was more in a factory setting, which can be seen as well in the TABLE 7 previously. These form the following empirical contributions:

EC4: Digital twin remains as a hard to grasp umbrella concept in the industry.

EC5: Organizations have knowledgeable persons on the topic of digital twin, even though organization might not be utilizing digital twin in their operations.

6.2.2 Machines' critical information

When asked about the monitoring of the machine, answers were broad depending on the machines' domains. Similarities could be drawn from the answers between the different sensor technologies the companies utilized, even though the domains were quite different. The same principles were repeated for the machines:

> "We utilize angle, pressure, suction, and acceleration sensors. The scale of the sensors is very broad. We also utilize LiDAR to identify surrounding elements." – ORG 1

Another interviewee mentioned same sensors in their machines:

> "Typical things we monitor in the machine's operation are closed site setting, power, feeder speed, chamber's surface area and RPM. In our mobile machines there is a default instrumentation, which on top the customer can select various options. In the static facilities the instrumentation is on the customer's responsibility." – ORG 3

When we compare these to the answers on the other two interviews, which were more on the factory setting environment for the machines, the answers differ a little. One of the interviewee's said the following:

> "We monitor the RPM, torque, power input – most of these are already built-in to the motors. We also utilize X-ray, temperature, and vibration sensors to measure different parts of the end-product through the manufacturing process. Ultrasound and laser scanners are also utilized. Lastly, we have different flow and pressure sensors to measure liquids and different gases." – ORG 4

Another interviewee from the factory environment gave a more top-level answer:

> "We have two different types of monitoring: process monitoring and maintenance monitoring. Process monitoring is done via automation system. The maintenance monitoring is mainly for machines which have moving parts, so it will alarm if there is a change in temperatures or bearings within the machines. It will also give predictive information regarding the states of the machine if it needs service." – ORG 2

These answers lead to the following empirical contribution:

> EC6: Monitoring of the machine requires a lot of different sensors depending on the machine.

Questions regarding the machine's critical information had a wide variety of answers depending on the domain the interviewee worked in. A good consensus on how the interviewees' determinate the critical information regarding the machine was from who's viewpoint you look at it. If you'd look it through the company producing or manufacturing the machines, it will be from their perspective what is critical for their operation, but customers could view critical information very differently. According to one of the interviewee's:

> "… Critical information can be in different levels depending on the person who you ask in the process" – ORG 2

Another interviewee's answer to the same question:

> "…Critical information can be client dependent. For example, our company doesn't need all the information our customers can get with our products. Critical information can be all the information the customer has gathered with our machine in the likes of different tasks, results, and business the customer has utilized the machine in." – ORG 1

As the interview's answers suggest, the empirical conclusion on critical information is the following:

> EC7: Critical information can vary a lot depending on the use case or user's perspective

All the interviewees had a different take on how the company itself sees critical information from the machine's perspective. This was probably due to the different machines and domains of operations for the companies. When asked on question 6 & 7, the relation was quite clear on all the interviewees' job positions and the information they believed was critical in their everyday work. Case organization 3 gave a good summary in their answer to the question 7 on what kind of machine's critical information they are working with:

> "In my role I develop digital solutions and automation to inner and outer users. It is essential that I understand, what the end users need and why, and on the other hand what are the different possibilities with different technologies. And how we can utilize these different technologies to carry on information to the end users." – ORG 3

Other interviewees answered the question 7 repeating many of the same aspects they already deemed critical information from the machine in the question 6 on what information regarding the machine is critical and why. However, when asked the question 8 on what kind of critical information they would want if they would be able to collect it from the machine, the answers varied a lot. For example, one of the interviewees stated the following:

> "In our current setting, we are measuring all the inputs and outputs in the factory. At this moment I have no information on inputs or outputs we are not measuring, but these can also appear in the future once the plant starts its operations. But due to the scale of the project, we have tried to be proactive in tapping/measuring all the information sources before the plant commissioning. This proactiveness is due to us not having the 100% info on what we should measure beforehand." – ORG 2

Another interviewee said the following, which was more focused on their machine's lack of specifications:

> "If we could get particle-size distribution straight from the feeder in real time, that would be valuable information. However, this is hard to do with current technologies, and we haven't figured out a way for setting this up." – ORG 3

Organization 4 continued on the very specific needs category:

> "If we could be able to see the tiniest of microscopic dirt particles from the molten plastic mass in the surfaces of the machine, we would strike gold. Due to this dirt particles the end-product will get compromised overtime. By seeing if these impurities are forming to the molten mass during the process, we are knowing during the process already that the end-product won't cut it." – ORG 4

These findings form the following empirical contribution:

> EC8: Companies see machine's critical information in various ways depending on the context

6.2.3 Data & models

In the data & models -section of questions, the answers to the interview were quite similar in their answers. Most of the answers to the question on where the storage of information & critical information related to the machine happens were in the automation system of the machine. One of the interviewees with slightly different answer said the following:

> "Different measurement data is collected to different network drives, specialists' hard drives & computers, which are used in product development. Critical information is mainly kept for specialists, servicing, and product development's use. Only people who need the critical information have access to it." – ORG 1

The same interviewee didn't give that much information on where in the machine the information is collected or stored. In all the other interviews the automation system was the machine's information storage, where the information then is extracted to different places and platforms. This forms the following empirical conclusion:

> EC9: Companies utilize multiple different storage solutions for their data

The modelling related questions had quite different answers depending on the interviewee's background and job positions. Due to others having a data science background and knowledge, and others having a more managerial role the answers were differentiated probably due to this. Answers varied to different flowcharts visualizing processes, KPI's, simulation models & system models for different parts of the machine, to prediction models alongside hardware-in-loop simulations. One of the interviewee's said the following:

> "We have different Simulink models we use, and Excel is one possibility if the data set is in usable form for Excel. I cannot specify other specific models we are using, rather platforms only we use." – ORG 5

The trend in the answers to these questions was that the information the question sought was usually critical business information the company wouldn't like to share publicly. This forms the following empirical conclusion:

> EC10: Companies are not eager to share information regarding their current developments for digital twins

However, one of the interviewees stated the following, when asked about the modelling of the critical data and/or information:

> "We utilize a digital twin, which is used to model the whole process from the start to the end. The process has all our processes within it. This digital twin is built from multiple different simulators, which all are interconnected with each other and the automation system." – ORG 2

Organization 2 & 5 were the only ones during all the interviews, where an actual digital twin was mentioned as a solution the company utilized. Rest of the interviewees had solutions in their utilization which can be seen as foundational parts of digital twin, but not the digital twin itself due to it being a rather complex concept compared to the solutions companies utilized. This leads to the following empirical contributions:

> EC11: Various companies already utilize digital twin in their operations

> EC12: Companies have abilities and required building blocks to build digital twins

6.2.4 Monitoring & decision-making

On the more managerial and top-level questions, the answers again depended on the interviewee's role and machine their position interacted with in their job. We must remember that all the interviewees were quite high in their organization and thus the answers can be considered quite relevant to the decision-making process of the company as well. Questions in this section were aimed at more of higher-level decision-making, not the everyday decisions, but rather decisions regarding the product's development or future needs for example.

Question regarding the monitoring of the critical information from the machine had similar answers again than the previous ones in data & models -section. Most of the answers had everyday aspect of the information in the likes of what does the operator see from the machine's information, rather than the business decision-making based on the critical information gained from the machine's operation. Mostly the answers consisted of the company's way of monitoring the critical data in their processes. The other answers were more from a customers' point of view, where the answer informed the way customers or users use the machine in the daily operations. One of the interviewees said the following, when asked about the monitoring of critical information:

> "Critical information comes from the customer feedback and field. Safety is priority number one, which is monitored and the challenges it comes with are solved in the fastest reaction as possible. Mainly these are customer's problems, which are reacted to swiftly." – ORG 1

This answer demonstrates that even during the interview, critical information can be seen through various points of view depending on the question. This can be seen from the answer another interviewee gave on the same question:

> "From the machine's automation screen on-site remotely from the control room or from the machine itself. On top of this, we have an IoT solution which can be used to remotely view the machine anywhere given that there is an internet connection. With these IoT solutions, you can for example do remote monitoring, analyzing with specialists, and help troubleshooting problems with the machines." – ORG 3

The same pattern, with organization 3's answer had repeated on the other answers as well. These form the following empirical conclusions:

> EC13: Companies utilize various ways to monitor their machine's critical information

When asked about the information or data, the organizations don't have at their disposal now, most of them wanted to have better information to better their products in development and drive the performance of their machines upwards in the likes of cost-efficiency, more end-production or resource-efficiency. Two answers rose clearly differing from the other themes stated previously:

> "If we could replicate the real-world and do complex simulation models with real-world settings, this would be the information we want. Of course, we would need the measurement data to validate the models as well." – ORG 1

And from other perspective on the same theme:

> "We have tried to get every data point in measurement beforehand from the factory setting we have been able to think of. This enables us all the information and data from the plant – getting all the data, rather than leaving some of them out consciously. This enables us to simulate a lot more of the processes and settings in the plant compared to doing real-world testing in the plant with pipes and chambers." – ORG 2

When asked about the decision-making process related to the machine's critical information, the answers leaned more on the cost-effectiveness optimization of their processes and machines, to cut down costs and produce more of the end-product. Other benefits mentioned were market related data and servicing decisions. One of the interviewees said the following:

> “Even though our plant is not yet operating, we have been able to get real-world benefits through our dynamic simulator, which was built beforehand. We did some test runs in the simulator based on the real plant setting and found out problems, which we can fix more easily before the plant starts its operations. After operation these fixes would be harder compared to before. Once we get the real data from the running plant, the simulator will be utilizing the data in its’ simulations, and these will be beneficial to the plant as well.” – ORG 2

These form the following empirical conclusions:

> EC14: Simulation of real-world operations and processes is seen valuable to organizations

> PEC3: Organizations utilize digital twins or components of digital twins to monitor their machines

6.2.5 Overview & digital image (digital twin)

In the machine’s overview and digital image (or digital twin) related questions the answers were once again separated on the topic. Common themes raised from the answers were the utilization of virtual reality spaces and the simulation aspects of the machines and processes. One of the interviewees said the following:

> “We have status models, which are almost like flowcharts, but they represent the status of the machine. When we combine these status models with all the monitoring and sensoring from the actual production line, we can feed the process data to the process model and with this model we can establish normal state. Changes to this normal state can then be viewed and interpreted accordingly in the simulation.” – ORG 4

Getting the overview of the whole machine’s status and processes, seems to come from the multiple little things that add up to the total overview. The same interviewee continued answering on the same question:

> “We also have a virtual reality environment, where we can interact with virtual reality glasses with the machine’s virtual model and see the overview of a single production line or machine. This virtual reality environment hasn’t been connected to the real-world data points, so this wouldn’t be classified as a digital twin though.” – ORG 4

The technologies themselves were hard to get as answers, but some opened a few technologies they were using to interpret the overview from the machine:

> "We utilize different web technologies such as javascript and different cloud solutions such as Platform-as-a-Service." – ORG 3

This answer didn't specify technologies other than the JavaScript, but this was a trend in the answers of getting vague or generalized answers. The problem here seemed to be that the companies themselves didn't want to give out specific information regarding their current solutions, due to the solutions were quite business critical for them. This forms the following empirical contribution:

> EC15: Organizations are not eager to share information if it contains business critical information

Even though rest of the interviewees didn't mention digital twin exclusively in their operations, one of the interviewees mentioned specific programs and simulators they are utilizing in their digital twin:

> "We utilize Aspen's dynamic simulator for chemical quantities, SIMIT-simulator for physical quantities, and automation operation is operated on top of these different simulators." – ORG 2

Here the interviewee gave out specific programs or solutions they are utilizing in their machine's overview or digital image (digital twin). The same interviewee continued giving out information on the next question as well, when asked about the information gotten from the overview and digital image:

> "In principle, we will get the same process data that we are getting from the real-world as well. It won't differ at all. It is all based on the simulator. The end goal is that eventually our operators won't know if they are operating a simulator or a real-world counterpart." – ORG 2

These form the following empirical contribution:

> EC16: One of the goals to utilize digital twin is to simulate real-world operations of the machine

Other interviewees highlighted the troubleshooting capabilities and product development aspects the digital image gives to the users. Simulation rose as a one key aspect of execution through the digital image. The same theme continued the next question on decision-making based on the digital image or digital twin. Organization 5 summarized other interviewees answers quite tightly in the following answer:

> "The typical information we want is from the operation of the machines; how the machines behave in certain situations like normal use, malfunctions etc. and how can the machine's operator affect this situation by their actions" – ORG 5

When concluding the interviews, the business value gained from the digital twin differentiated in some aspects, but the main takeaway was that there is business value gained from the digital image or digital twin of the machine. Many of the interviewees highlighted one aspect of the process, which is bettered by having the machine's digital image in the process mix. The key is optimization in the areas important to the process or how organization sees fit. One of the interviewees concluded their interview in the following words:

> "Biggest use of digital twin and image together with real-world data and models is optimization of everything. This optimization gives the client a peace of mind to do their work as efficiently and reliably. The clear benefit from digital twins is to find the anomalies in the behavior of the products or machines, and then we can use these anomalies in servicing, which gives benefits for the customers. Minimizing quality costs is one the benefits of a real-time digital twin which we see, but currently we are not able to see or say all the malfunctions." – ORG 1

Another interviewee reported the following continuing the theme from organization 1's answer:

> "We have been able to cut down quality costs by utilizing digital images from our products. We have been able to test them before they are built, and by this we save time from testing with the actual physical ship." – ORG 5

Other interviewees highlighted value in the already risen themes of better servicing and product development. These can be seen as a loop as well, that the digital twin enables seeing a flaw in a product, which then gets serviced, and this flaw is considered in the product development as well, which leads to better functioning products and machines. One of the interviewees said the following as well regarding the business value of the digital twin:

> "Digital twin can be used as a virtual commissioning tool in plant startups. We used virtual commissioning for two weeks and found out a host of problems, which gave us an understanding of the flaws in future plant's operation. We could fix these issues before the plant powers up. This in terms saves us time and cuts costs since we can fine tune the processes beforehand and before the real plant commissioning. "– ORG 2

This concludes to the following empirical contributions:

> EC17: Companies have the prerequisites for utilization of digital twin in their operations

> PEC5: Companies utilize digital twin's components in their operations without the digital twin as an end goal

6.3 Summary

The chapters 6.1 and 6.2 included empirical data from the interviews. This data was the basis for the empirical and primary empirical conclusions, which were drawn with their corresponding connections to the interviews previously. The goal of the empirical data was to highlight different problems industry is tackling with in their operations, and how these companies are utilizing digital twins or likewise solutions to tackle these problems. Empirical conclusions were drawn to form the basis for the primary empirical conclusions. These empirical conclusions from the previous chapters are formed in the TABLE 8 below.

TABLE 8 Empirical conclusions from the data

Identifier	Empirical conclusion
EC1	Cloud-technologies are an enabler and a limiting factor for forming a digital twin.
EC2	Person's role is highly correlated on the understanding of digital twin.
EC3	Depending on the job position, people interact with multiple different inner and outer stakeholders on their job
EC4	Digital twin remains as a hard to grasp umbrella concept in the industry
EC5	Organizations have knowledgeable persons on the topic of digital twin, even though organization might not be utilizing digital twin in their operations
EC6	Monitoring of the machine requires a lot of different sensors depending on the machine
EC7	Critical information can vary a lot depending on the use case or user's perspective
EC8	Companies see machine's critical information in various ways depending on the context
EC9	Companies utilize multiple different data storage solutions for their data
EC10	Companies are not eager to share information regarding their current developments for digital twins

EC11	Various companies already utilize digital twin in their operations
EC12	Companies have abilities and required building blocks to build digital twins
EC13	Companies utilize various sensors to monitor their machine's critical information
EC14	Simulation of real-world operations and processes is seen valuable to organizations
EC15	Organizations are not eager to share information if it contains business critical information
EC16	One of the goals to utilize digital twin is to simulate real-world operations of the machine
EC17	Companies have the prerequisites for utilization of digital twin in their operations

All in all, five different primary empirical contributions were drawn from the empirical conclusions. These primary empirical conclusions are illustrated in the TABLE 9 below

TABLE 9 Primary empirical conclusions from the data

Identifier	**Empirical conclusion**
PEC1	Some of the interviewed companies reported having technological capabilities to form digital twins.
PEC2	All the interviewed reported opportunities on the following areas for the utilization of digital twin or -like solutions: cost-optimization, improved servicing, and improved troubleshooting.
PEC3	Organizations utilize digital twins or components of digital twins to monitor their machines
PEC4	Organization's goal is to utilize digital twin to simulate real-world operations
PEC5	Companies utilize digital twin's components in their operations without the digital twin as an end goal

These primary empirical conclusions are further enriched to give them context. The context enriched PECs are illustrated below on TABLE 10.

TABLE 10 Primary empirical conclusions from the data, which are enriched to their context

Identifier	Context-enriched conclusion
PEC1	Various industries have technological capabilities to form digital twins.
PEC2	Various industries have opportunities on the following areas to utilize digital twins or -like solutions: cost-optimization, improved servicing, and improved troubleshooting.
PEC3	Organizations utilize digital twins or components of digital twins to monitor their machines
PEC4	Organization's goal is to utilize digital twin to simulate real-world operations such as machines or processes
PEC5	Companies utilize digital twin's components in their operations without the digital twin as an end goal

7 DISCUSSION

In this chapter, the empirical findings will be analyzed through the theoretical background formed in the literature review, and implications will be discussed based on the results from the interviews.

7.1 Practical implications

This thesis was conducted on the practical implementation of digital twin, it's perspectives and challenges. Where the literature review focused more on the theory building for the development & implementation of digital twins, it's definitions and different frameworks & architectures, the empirical research focused more on the current states digital twin stands on the various industries interviewed. The findings were expected to be different problems and prerequisites for the implementation of digital twins in practice.

As PEC1 states, some of the interviewed companies reported having technological capabilities to form digital twins. During interviews many of the interviewed persons had the knowledge of the concept digital twin. These persons were also thought of the most qualified persons to be interviewed from the different organizations contacted for the interviews. Even though some of the organizations reported having the technological capabilities, it was apparent during the interviews, that there are still barriers for the actual development of digital twins.

As PEC2 gives information, all the interviewed reported opportunities on the following areas for the utilization of digital twin or -like solutions: cost-optimization, improved servicing, and improved troubleshooting. This was a clear conclusion based on the multiple different opportunities interviewees reported, when asked on how they have made decision based on the machine's critical data and the utilization of digital image or digital twin in the operations. It is important to highlight, that most of the interviewees didn't have a digital twin in utilization, but rather components of it, which were used to seize these described opportunities. Organizations who had digital twin in use or as a goal reported same benefits gained in terms of cost-optimization, improved servicing, and improved troubleshooting.

As PEC3 states, organizations utilize digital twins or components of digital twins to monitor their machines. Two of the interviewed persons reported having digital twins in use in the organization, which were utilized in different contexts. The rest of the interviewees reported having components of digital twins in use, with or without the end-goal of possessing digital twin in the future. The monitoring was happening on multiple different levels, and this would mean it is important to take multiple different levels or layers into account, when developing digital twins or components of them for machines depending on their use case.

PEC4 states, organization's goal is to utilize digital twin to simulate real-world operations. This was continuum to the PEC3, where the end goal of organizations having components of digital twin, or digital twin itself in use – their use case for the digital twin was to simulate the real-world to the best of their abilities. During the interviews multiple different use cases were mentioned for the possible digital twins' utilization, one of the interviewees reported the operators training being 1:1 with the real-world's counterparts, where the goal would be for the operators not being able to differentiate the training simulator from the real operating environment.

PEC5 concludes, that companies utilize digital twin's components in their operations without the digital twin as an end goal. This empirical conclusion was apparent from multiple interviews, where the interviewee reported the digital twin not having value, the machine not supporting, or use cases were difficult to see digital twin in. These pitfalls of digital twin were represented in other terms in the literature in the digital twin's adoption's challenges by Wanasinghe et al. (2020). However, the industries represented in the study by Wanasinghe et al. (2020) were oil & gas industries, which none of them was represented in the collected data on this study.

Below is a TABLE 11, which represents practical implications of primary empirical conclusions.

TABLE 11 Practical implications of primary empirical conclusions

Identifier	**Implication for practice**
PEC1	Capabilities for developing digital twins is present in various industries
PEC2	Opportunities gained by utilization of digital twin can be found on the following areas: cost-optimization, improved servicing, and improved troubleshooting.
PEC3	Digital twins or components of digital twins can be used to monitor machines
PEC3 & PEC4	Organizations can simulate real-world operations by utilizing digital twins and/or components of digital twins.

PEC5	Companies can utilize digital twin's components in their operations without the digital twin as an end goal

7.2 Implications for the MVP

In this chapter, the primary empirical contributions are used to reflect on starting product development of digital twin's MVP model. The framework for the development was opened in chapter 4. Based on the TABLE 11's practical implications of primary empirical conclusions the following implications are drawn for the requirements and guidelines for the MVP.

Due to the capabilities being present in various companies already, it is important to start the architectural mapping for the digital twin. Also, the digital twin's MVP need to be capitalizing on one or many of the various opportunities presented by organizations such as improved servicing, cost-optimization, and improved troubleshooting.

The simulation and monitoring aspects of digital twin rose in the implications for practice as well. This suggests that the digital twin's MVP needs to be able to monitor and simulate some aspects of the machine. Especially the simulation of real-world operations needs be present in the digital twin's MVP.

Different architectures for cyber-physical systems and digital twins were presented in the literature review. Based on these different architectures and the gotten parameters based on the interviews of the various sensors and components companies utilize in the use of digital twin and or like solutions, there needs to be some sort of coherent concept, which answers to all these previous requirements.

Also, the fact that companies utilize components of digital twins without the end goal of a digital twin, it suggests, that MVP cannot be done together with just any company. To utilize the ESSSDM model by Bosch et al. (2013) in chapter 3.2., continuous iteration and the enablement of build-measure-learn feedback loop is important to meet the expectations of the customer.

Lastly, these implications for the MVP are just the surface-level, and the actual development of the MVP requires a lot more work and research on the different requirements, settings, and potential use case.

8 CONCLUSION

This chapter will conclude the thesis by answering the research questions, providing limitations of study, and giving out possible future research areas.

8.1 Answers to the research questions

This thesis focused on the research on implementations and practice of digital twins through literature review and an empirical study. To grasp the implementation of digital twin in practice, this was approached through the primary research question:

- **How to implement a digital twin in practice?**

The question was answered through the literature review and research framework, where the definitions and architectural guidelines for digital twin were presented through various studies, but most concisely in Parrott and Warshaw (2017) article on industry 4.0 and the digital twin. However, it was apparent in the literature, that the actual technological part of implementation wasn't represented in clearly in the literature, and same applied to the actual implementation in practice, when collecting and analysing the empirical data. This can be due to the vastly different use environments for the digital twin and the different requirements, which are dependent on multiple variables.

The first secondary research question searched answer for the different componential factors:

- **From what components are digital twins built from?**

The research question was also answered in the literature review, where different models and architectures were presented for cyber-physical systems and digital twins through academical literature and industry standards. These components were laid out in umbrella terms and as potential technological components, but not as concrete pre-selected technologies. The digital twins are part of a larger mechanism, cyber-physical systems, where digital twin can be seen as one component or layer for the whole system as illustrated in the 5C Architecture, Reference Architectural Model Industrie 4.0 (RAMI4.0), and Industrial Internet Reference Architecture (IIRA).

The 2nd secondary research question tried to understand the different challenges on implementing digital twins:

- **What are the challenges of implementing digital twins?**

The research question was answered through the literature review and some pain points were touched on the empirical data as well. The challenges for digital twins were documented widely through literature, and the empirical data reflected on some of the practical issues such hardware constraints, lack of knowhow, and technical difficulty.

The empirical data collection was mainly utilized for the following secondary research question together with literature review on industry's practices:

- **How is the industry currently utilizing digital twin technologies?**

In the literature review, there was presented the different architectural models and frameworks for cyber-physical systems, which are the system a digital twin is a part of. On top of this, the empirical data suggested, that industries have various interests on utilization of digital twins. Some interviewees reported the company having only components of a digital twin in use for different machines, as others reported having complete digital twins in their operational use. The ones having a digital twin in use were on the factory setting processing industry and maritime industry, which were opened in the empirical findings chapter previously.

8.2 Limitations of study

Trying to bridge the gap on academia and industry on a relatively new research area such as digital twins proved hard. One of the challenges and clear limitations for the study was apparent during the interviews, due to the interviewees having a broad background and roles. The interviewees were one of the most knowledgeable or even the most knowledgeable in the organizations, which were selected to be in the interview, but still the interviews lacked certain information, which may be due to the information required being business critical information. Role and background of interviewees was broad as well, but it proved quite essential in the information and knowledge the interviewees had on the other topics. This was also one of the shortfalls in the interviews, since information gained was relatively general on the depth, but it still proved valuable.

In the literature review the limitations were certainly on the relatively new literature, and this literature lacking practical applications or industry's applications. This proved it hard to get the actual implementations and components through the literature, but the empirical data supported the lack of data from the literature, this provided some perspective and data for the study.

Also, digital twin proved to be a very broad concept to try deriving requirements and components for the MVP model for multiple industries for a universal digital twin MVP. Through the literature review the components and architectural requirements were derived on a broad conceptual level, but the information remained quite broad without actual technical components.

Challenges in the interview stage of this thesis was that the even though it started as a try to get many problems and technological requirements from different industries, it quickly came apparent that even within one industry there are multiple different factors in the implementation of digital twins to take into consideration. One interviewee per industry per company wasn't clearly sufficient to get the required information for digital twin MVP's technological requirements. The focus should've been tighter on one industrial domain, and the interviewed companies the ones with actual digital twins in their use already. It proved problematic to get the requirements for digital twin's implementation from companies, who didn't have them in implemented, but rather wanted to have them in the future.

Lastly, due to the study's data collection method being semi-structured thematic interviews, the answers varied greatly in depth and relativeness to the interview's structures. Interviews are as strong as the interviewees' answers to the provided questions. Having an even tighter scope on the study would have maybe provided better empirical data through concise answers. The same theme carried out in the literature review as well.

8.3 Future research

As laid out in the limitations to the study, the interviewees were on a quite broad level in terms of industries represented. Also, the provided questionnaire was a broad one and didn't provide very detailed answers to the actual research questions. In the future studies on digital twins, research should focus on more on one industry's applications on how these selected industries are implementing digital twins for their operations. Tackle one industry in one study at a time rather than multiple ones.

In the literature review, there are plenty of different studies providing challenges to the cyber-physical systems, and digital twins. Them being the main drivers together, and digital twins not existing without the cyber-physical system, future research should focus on answering these challenges and providing solutions to them.

Also, the development of digital twin's MVP model in terms of setting the bare minimum requirements for the digital twin on a component level should be studied as well. The problem with trying to find digital twin's MVP requirements or components was apparent in this study due to this study focusing on a broader level rather than a tighter scope. A tighter scope on a selected use case or even industry would give a better angle for the study to focus on and thus yield better concise results.

SOURCES

Alur, R. (2015). Principles of cyber-physical systems. MIT press.

Ardichvili, A., Cardozo, R., & Ray, S. (2003). A theory of entrepreneurial opportunity identification and development. Journal of Business venturing, 18(1), 105-123.

Baheti, R., & Gill, H. (2011). Cyber-physical systems. The impact of control technology, 12(1), 161-166.

Bosch, J. (2012, June). Building products as innovation experiment systems. In International Conference of Software Business (pp. 27-39). Springer, Berlin, Heidelberg.

Bosch, J., Holmström Olsson, H., Björk, J., & Ljungblad, J. (2013, December). The early stage software startup development model: a framework for operationalizing lean principles in software startups. In *International Conference on Lean Enterprise Software and Systems* (pp. 1-15). Springer, Berlin, Heidelberg.

Boschert, S., Heinrich, C., & Rosen, R. (2018, May). Next generation digital twin. In Proc. tmce (Vol. 2018, pp. 7-11). Las Palmas de Gran Canaria, Spain.

Capgemini Consulting. (2017). Industry 4.0 – The Capgemini Consulting View: Sharpening the Picture beyond the Hype [White paper]. Accessed https://www.capgemini.com/consulting/wp-content/uploads/sites/30/2017/07/capgemini-consulting-industrie-4.0_0_0.pdf

Cruzes, D. S., & Dyba, T. (2011, September). Recommended steps for thematic synthesis in software engineering. In 2011 international symposium on empirical software engineering and measurement (pp. 275-284). IEEE.

Cruzes, D. S., Dybå, T., Runeson, P., & Höst, M. (2015). Case studies synthesis: a thematic, cross-case, and narrative synthesis worked example. Empirical Software Engineering, 20(6), 1634-1665.

Dennehy, D., Kasraian, L., O'Raghallaigh, P., & Conboy, K. (2016). Product market fit frameworks for lean product development.

DIN SPEC 91345:2016-04 - Referenzarchitekturmodell Industrie 4.0 (RAMI4.0). (n.d.). https://www.plattform-i40.de/IP/Redaktion/DE/Downloads/Publikation/din-spec-rami40.html

Fagerholm, F., Guinea, A. S., Mäenpää, H., & Münch, J. (2017). The RIGHT model for continuous experimentation. Journal of Systems and Software, 123, 292-305.

Freeman, J., & Engel, J. S. (2007). Models of innovation: Startups and mature corporations. *California Management Review*, *50*(1), 94-119.

Fuller, A., Fan, Z., Day, C., & Barlow, C. (2020). Digital twin: Enabling technologies, challenges and open research. IEEE access, 8, 108952-108971.

Garousi, V., Felderer, M., & Mäntylä, M. V. (2016, June). The need for multivocal literature reviews in software engineering: complementing systematic literature reviews with grey literature. In Proceedings of the 20th international conference on evaluation and assessment in software engineering (pp. 1-6).

Giaimo, F., Yin, H., Berger, C., & Crnkovic, I. (2016, May). Continuous experimentation on cyber-physical systems: challenges and opportunities. In Proceedings of the scientific workshop proceedings of XP2016 (pp. 1-2).

Girgenti, A., Pacifici, B., Ciappi, A., & Giorgetti, A. (2016). An Axiomatic Design Approach for Customer Satisfaction through a Lean Start-up Framework. *Procedia CIRP*, *53*, 151–157. https://doi.org/10.1016/j.procir.2016.06.101

Glaessgen, E., & Stargel, D. (2012, April). The digital twin paradigm for future NASA and US Air Force vehicles. In 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA (p. 1818).

GreyNet International, Grey Literature Network Service. (2023). https://www.greynet.org/

Grieves, M. (2014). Digital twin: manufacturing excellence through virtual factory replication. White paper, 1(2014), 1-7.

Harper, K. E., Ganz, C., & Malakuti, S. (2019). Digital twin architecture and standards. IIC Journal of Innovation, 12, 72-83.

Järvinen, J., Huomo, T., Mikkonen, T., & Tyrväinen, P. (2014). From Agile Software Development to Mercury Business. In C. Lassenius, & K. Smolander (Eds.), Software Business. Towards Continuous Value Delivery : 5th International Conference, ICSOB 2014, Paphos, Cyprus, June 16-18, 2014. Proceedings (pp. 58-71). Springer. Lecture Notes in Business Information Processing, 182. https://doi.org/10.1007/978-3-319- 08738-2_5

Jazdi, N. (2014, May). Cyber-physical systems in the context of Industry 4.0. In *2014 IEEE international conference on automation, quality and testing, robotics* (pp. 1-4). IEEE.

Jiang, J. R. (2018). An improved cyber-physical systems architecture for Industry 4.0 smart factories. Advances in Mechanical Engineering, 10(6), 1687814018784192.

Jones, D., Snider, C., Nassehi, A., Yon, J., & Hicks, B. (2020). Characterising the Digital Twin: A systematic literature review. CIRP Journal of Manufacturing Science and Technology, 29, 36-52.

Lee, E. A. (2008, May). Cyber-physical systems: Design challenges. In 2008 11th IEEE international symposium on object and component-oriented real-time distributed computing (ISORC) (pp. 363-369). IEEE.

Lee, E. A. (2015). The past, present and future of cyber-physical systems: A focus on models. *Sensors*, *15*(3), 4837-4869.

Lee, J., Bagheri, B., & Kao, H. A. (2015). A cyber-physical systems architecture for industry 4.0-based manufacturing systems. Manufacturing letters, 3, 18-23.

Lenarduzzi, V., & Taibi, D. (2016). MVP Explained: A Systematic Mapping Study on the Definitions of Minimal Viable Product. *2016 42th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*. https://doi.org/10.1109/seaa.2016.56

Lin, S.-W., Mellor, S., Munz, H., & Barnstedt, E. (2017). Architecture Alignment and Interoperability An Industrial Internet Consortium and Plattform Industrie 4.0 Joint Whitepaper. Retrieved from https://www.iiconsortium.org/pdf/JTG2_Whitepaper_final_20171205.pdf

Mahood, Q., Van Eerd, D., & Irvin, E. (2013). Searching for grey literature for systematic reviews: challenges and benefits. Research Synthesis Methods, 5(3), 221–234. doi:10.1002/jrsm.1106

Mattos, D. I., Bosch, J., & Olsson, H. H. (2017, August). Your system gets better every day you use it: towards automated continuous experimentation. In 2017 43rd Euromicro Conference on Software Engineering and Advanced Applications (SEAA) (pp. 256-265). IEEE.

Mosterman, P. J., & Zander, J. (2016). Cyber-physical systems challenges: a needs analysis for collaborating embedded software systems.

Parrott, A., & Warshaw, L. (2017). Industry 4.0 and the digital twin. Deloitte Insights.

Pivoto, D. G., de Almeida, L. F., da Rosa Righi, R., Rodrigues, J. J., Lugli, A. B., & Alberti, A. M. (2021). Cyber-physical systems architectures for industrial internet of things applications in Industry 4.0: A literature review. Journal of Manufacturing Systems, 58, 176-192.

Rajkumar, R., Lee, I., Sha, L., & Stankovic, J. (2010, June). Cyber-physical systems: the next computing revolution. In Design automation conference (pp. 731-736). IEEE.

Rasheed, A., San, O., & Kvamsdal, T. (2020). Digital twin: Values, challenges and enablers from a modeling perspective. Ieee Access, 8, 21980-22012.

Ries, E. (2011). The lean startup. *New York: Crown Business* , *27* , 2016-2020.

Rosen, R., Fischer, J., & Boschert, S. (2019). Next generation digital twin: An ecosystem for mechatronic systems?. IFAC-PapersOnLine, 52(15), 265-270.

Schermann, G., Cito, J., & Leitner, P. (2018a). Continuous experimentation: challenges, implementation techniques, and current research. Ieee Software, 35(2), 26-31.

Schermann, G., Cito, J., Leitner, P., Zdun, U., & Gall, H. C. (2018b). We're doing it live: A multi-method empirical study on continuous experimentation. Information and Software Technology, 99, 41-57.

SCHEUENSTUHL, F., BICAN, P. M., & BREM, A. (2020). HOW CAN THE LEAN STARTUP APPROACH IMPROVE THE INNOVATION PROCESS OF ESTABLISHED COMPANIES? AN EXPERIMENTAL APPROACH. International Journal of Innovation Management, 2150029. https://doi.org/10.1142/s1363919621500298

Semeraro, C., Lezoche, M., Panetto, H., & Dassisti, M. (2021). Digital twin paradigm: A systematic literature review. Computers in Industry, 130, 103469.

Stark, R., & Damerau, T. (2019). Digital Twin. CIRP Encyclopedia of Production Engineering, 1–8. doi:10.1007/978-3-642-35950-7_16870-1

Tao, F., Qi, Q., Wang, L., & Nee, A. Y. C. (2019). Digital twins and cyber–physical systems toward smart manufacturing and industry 4.0: Correlation and comparison. Engineering, 5(4), 653-661.

The Industrial Internet Reference Architecture. (2022). Retrieved from https://www.iiconsortium.org/wp-content/uploads/sites/2/2022/11/IIRA-v1.10.pdf

Trauer, J., Schweigert-Recksiek, S., Engel, C., Spreitzer, K., & Zimmermann, M. (2020, May). What is a digital twin?–definitions and insights from an industrial case study in technical product development. In Proceedings of the Design Society: DESIGN Conference (Vol. 1, pp. 757-766). Cambridge University Press.

Tripathi, N., Oivo, M., Liukkunen, K., & Markkula, J. (2019). Startup ecosystem effect on minimum viable product development in software startups. *Information and Software Technology*, *114*, 77–91. https://doi.org/10.1016/j.infsof.2019.06.008

Uhlemann, T. H. J., Lehmann, C., & Steinhilper, R. (2017). The digital twin: Realizing the cyber-physical production system for industry 4.0. Procedia Cirp, 61, 335-340.

Wanasinghe, T. R., Wroblewski, L., Petersen, B. K., Gosine, R. G., James, L. A., De Silva, O., ... & Warrian, P. J. (2020). Digital twin for the oil and gas industry: Overview, research trends, opportunities, and challenges. IEEE access, 8, 104175-104197.

Wang, S., Wan, J., Zhang, D., Li, D., & Zhang, C. (2016). Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. Computer networks, 101, 158-168.

Wurm, J., Jin, Y., Liu, Y., Hu, S., Heffner, K., Rahman, F., & Tehranipoor, M. (2016). Introduction to cyber-physical system security: A cross-layer perspective. IEEE Transactions on Multi-Scale Computing Systems, 3(3), 215-227.

Yordanova, Z. B. (2018). Lean Startup Method Hampers Breakthrough Innovations and Company's Innovativeness. *International Journal of Innovation and Technology Management, Vol. 15, No. 2, March 2018*. https://doi.org/10.1142/9789811247729_0001

Yu, Z., Kaplan, Z., Yan, Q., & Zhang, N. (2021). Security and privacy in the emerging cyber-physical world: A survey. IEEE Communications Surveys & Tutorials, 23(3), 1879-1919.

Zheng, P., Sang, Z., Zhong, R. Y., Liu, Y., Liu, C., Mubarok, K., ... & Xu, X. (2018). Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. Frontiers of Mechanical Engineering, 13(2), 137-150.

Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of industry 4.0: a review. Engineering, 3(5), 616-630.

APPENDIX 1 : THEME INTERVIEW QUESTIONS FORM

Interviewee's domain

Question 1: What is your role in your company?

Question 2: What kind of stakeholders are you co-operating in your position?

Question 3: What are the machines, your company produces?

Question 4: In which domain the machines operate in your company manufactures?

Machine's critical information

Question 5: How do you monitor the machine's operation? What kind of sensors and probes do you utilize?

Question 6: What information regarding the machine is critical and why?

Question 7: In your profession, what kind of machine's critical information are you working with?

Question 8: What critical information would you like to get from the machine's state, which information you are not able to collect currently?

Data and models

Question 9: Where do you collect and store data regarding machine's data and critical information?

Question 10: What kind of models do you use to model the machine's critical data?

Question 11: What kind of models do you use to visualize machine's critical data?

Monitoring & decision-making

Question 12: How do you monitor machine's critical information?

Question 13: What kind of/level decision-making you are making based on the collected and analyzed machine's critical information?

Question 14: If you don't have data/information, where to base your decision-making, what kind of data you would require, need or want?

Question 15: What kind of decisions you have been able to make based on the machine's critical information?

Overview & digital image (digital twin)

Question 16: What kind of technologies do you use to illustrate machine's overview and digital image (digital twin)?

Question 17: What information do you get from the machine's overview and digital image (digital twin)?

Question 18: What decision-making you base on the machine's overview and digital image (digital twin)?

Question 19: What kind of business benefits you have gained through the machine's digital image (digital twin)?