Miguel Alves de Azevedo

Knowledge-based engineering supported by the Digital Twin: The case of the power transformer at Efacec

Dissertation carried out under the master's degree in Information Science, supervised
by Professor António Lucas Soares

Faculty of Engineering and Faculty of Arts

University of Porto

July 2020

Knowledge-based Engineering supported by the Digital Twin: the case of the power transformer at Efacec

Miguel Alves de Azevedo

Dissertation carried out within the scope of the master's in information science,
supervised by Professor Doctor António Lucas Soares.

Jury Members

President: Professor Doctor Olívia Manuela Marques Pestana
Faculdade de Engenharia da Universidade do Porto

External Examiner: Professor Doctor António João Pina da Costa Feliciano Abreu
Instituto Superior de Engenharia de Lisboa

Supervisor: Professor Doctor António Manuel Lucas Soares
Faculdade de Engenharia da Universidade do Porto

"You can't manage knowledge – nobody can. What you can do is to manage the environment in which knowledge can be created, discovered, captured, shared, distilled, validated, transferred, adopted, adapted and applied."

Chris Collison and Geoff Parcell

Acknowledgements

I would like to express gratitude to a group of people whose contribution and support made it possible for me to successfully conclude my dissertation and master's degree, as well as evolve as a person during my academic career.

First and foremost, I would like to thank Professor Lucas Soares for his guidance and insights regarding the work carried out in this dissertation. Additionally, I would like to acknowledge his role in sparking my interest in the Information Science area, which has grown ever since attending the course that he taught in my freshman year.

Furthermore, I am thankful for the support given by Efacec, especially Sérgio Tavares, for his role in coordinating my work with the organization, such as in scheduling interviews and providing information regarding how Efacec operates. This dissertation would not have been possible without him.

I am also grateful to all the friends I have made during the 5 years I attended university. They have left a lasting impression on me, and I am sure I am a more honest, humble, and mature person than I was, because I met them.

In a similar vein, I am glad to have been mentored by an excellent group of professors, both during my bachelor's degree, and my master's degree. They went out of their way to pass on their experience and skills to the students, and as such, I am certain that they have prepared us well to carry on and expand the work carried out by the giants of Information Science.

I would also like to note that I owe all my achievements to my family, who gave me the means to pursue a higher education. They endured tremendous hardship, and because of it, I will be able to live a better life. Moreover, I thank my mother for the affection she gave me my entire life, my dad for instilling a moral compass in me, my sister for teaching me to be cheerful and hardworking, and my brother, who, although I have a hard time admitting it, is my role model in many ways.

Finally, I dedicate this dissertation to my grandmother Conceição, who unfortunately passed away before she could see me graduate, and transition into adult life. "Miguel studies so far away, my poor boy", she used to say.

It was worth it, avó. It was worth it.

Abstract

Industry 4.0 has made it possible for emerging technologies to revolutionize how organizations operate. New applications, supported by the Internet of things, cyber physical systems, and cloud computing, take advantage of large data exchange networks that capture data from the real and virtual world, to generate valuable insights for product development. This, together with the growing digitalization of product lifecycle information, has made information the most valuable asset of an organization, as it can be applied to improve product design, reduce lead time and decrease monetary costs.

However, the growing volume, formats, and purposes of the information an organization captures, also brings challenges for information management, and consequently, appropriate IM and KM instruments and strategies must be adopted to successfully take advantage of organizational knowledge.

The adoption of Knowledge-based Engineering (KBE) can accomplish these goals. KBE refers to the knowledge management tasks of capturing, storing, modeling, coding, and sharing of organizational knowledge, both in explicit form, such as documents, and tacit form, present in the minds of employees. Ultimately, this results in systems that can automate design tasks.

Also in the context of technological advances, a new concept called Digital Twin has emerged, which employs bidirectional data transmission to mirror the lifecycle of a physical product, in the virtual realm. Proposed DT functionalities actively use organizational knowledge to improve and automate product design, and as such, this technology can be an adequate vessel for KBE.

This dissertation focuses on the implementation of the Digital Twin in power transformer development processes. Using the case of Efacec, a portuguese firm of the energy sector, the DT concept was developed, and this involved defining functionalities that are driven by organizational knowledge to automate, optimize, and streamline PT design tasks, thus accomplishing the goal of KBE. Some of the proposed DT features are the generation of design templates, the identification of design nonconformities, and the capture of engineer feedback.

Furthermore, the DT information architecture that is required for these functionalities to successfully be implemented, was envisioned, by defining all captured and generated information in each PT lifecycle phase. Finally, a classification scheme that classifies DT information and enables queries within the DT platform, was developed.

Keywords: Information management, knowledge management, Knowledge-based Engineering, Digital Twin, power transformer development

List of Figures

List of tables

List of abbreviations and acronyms

CAD - Computer-aided Design

CAE - Computer-aided Engineering

DT - Digital Twin

IM – Information Management

IoT - Internet of Things

KM - Knowledge Management

KBE - Knowledge-based Engineering

KBES - Knowledge-based Engineering Systems

PT - Power Transformer

R&D - Research and Development

Table of Contents

1. Introduction

1.1. Context and motivation

The high degree of technological development that has been happening since the end of the 20th century, allowed companies to implement a set of new technologies that revolutionized the way they operate. This paradigm shift, called Industry 4.0, is characterized by the integration of Information and Communication Technologies with physical objects and devices, resulting in the creation of large data exchange networks supported by cyber-physical systems and the Internet of Things.

The capture of operational data, as well as its analysis through algorithms and artificial intelligence, encourages the growing digitization of product development processes, making information the main asset for the creation of highly flexible, optimized, and customizable products.

However, crucial information for product design is not limited to operational data, as it also includes information generated by the product development processes, including requirements, product design models and data related to product testing. This information, captured over time, is an important knowledge base for the organization, assisting engineers in decision-making, which in turn improves product design, and reduces lead time and monetary costs. Therefore, knowledge management is pivotal to identify, capture, organize and make available relevant knowledge to product engineers, during design processes.

Information Management (IM) and Knowledge Management (KM) are especially critical in sectors of the industry that create complex products, the development of which involves multiple engineering areas, and where collaboration and coordination between employees with different skills is crucial.

This is the case in the development of power transformers, machines whose design involves the collaboration between different engineering teams and the generation of a large volume of design artifacts, and whose operation can generate valuable insights from sensor data. Accordingly, these processes benefit from strategies and methods that allow the capture of useful knowledge acquired throughout the lifecycle of each transformer, making it available to engineers at the right times to assist them in decision-making.

This dissertation focuses on knowledge management in transformer design processes at Efacec, a portuguese firm from the energy, engineering and mobility sectors, and its goal is to define GC strategies and instruments to meet the challenges and opportunities of Industry 4.0.

The developed strategies and mechanisms will be supported by the Digital Twin, an emerging concept that takes advantage of Industry 4.0 technologies, such as sensors, data analysis algorithms, simulation and artificial intelligence, to mirror the physical product lifecycle in virtual space, and that can also be a repository for artifacts originating from product development.

1.2. Problem description

The development of power transformers generates a large amount of information, which varies in format and utility. Among this information are the requirements, standards and restrictions that make up product specifications, the electrotechnical and mechanical models generated during product design, data originating from product testing, and data captured by sensors during the operation of the transformer.

This information is an important resource for the organization, allowing engineers to complement their knowledge and experience, with the organizational knowledge captured over time. However, the management of this information is often problematic, with gaps in the definition of what information should be captured or discarded, how it should be organized and how and when it should be made available.

Thus, the implementation of strategies and instruments that make the capture, description, organization, and retrieval of information generated during the lifecycle of the transformer, more efficient and effective, is a crucial factor to obtain benefits from organizational knowledge.

An emergent technology that can meet these requirements is the Digital Twin. The potential features of this platform, especially the capture and analysis of data originating from sensors, can play an important role in the codification of knowledge, continuously providing insights into the optimized characteristics of the transformer design. In addition, the fact that DT is present at all stages of the product's lifecycle means that it can also serve as a repository of all the information generated during the design of the transformer, providing engineers with an organizational knowledge base that assists them in decision making.

However, the Digital Twin is an emerging research and development topic and, consequently, its potential applications are mostly in the theoretical design phase, instead of being already a reality in industry and services. Particularly in product development, the work that focuses on the usefulness of Digital Twin is still scarce, so there is a need to conceptualize this tool in terms of functionalities, technologies, and information architecture.

Thus, this dissertation proposes a specification of the concept of Digital Twin from the perspective of information, knowledge, and collaboration management.

1.3. Objectives and expected results

The objectives defined for this dissertation can be divided into two categories, the first relating to the development of the Digital Twin concept and the second to the integration of knowledge management instruments in the transformer development processes at Efacec. These categories are interdependent, since the integration of knowledge management instruments in the PT development processes will be supported by the Digital Twin, and this concept can only be built considering existing processes.

The objectives related to the concept of Digital Twin include the development of the informational architecture of this platform, how to organize, classify and make available the artifacts created during the design of a power transformer, defining requirements for the management of captured data in previous development processes, and identifying features that the platform should include to foster collaboration between process and product engineers.

Based on the developed concept, the objective is to reformulate the process of development of transformers in the perspective that they have Digital Twin as a central tool for access, storage and availability of information, and also for this platform to be a source of continuous improvement for product design.

The achievement of these objectives is supported by the elaboration of the state of the art regarding the application of KBE and Digital Twin in product development. This task has the purpose of identifying the ideas and work carried out by academic and professional authors in these areas. Another essential basis for this work is the analysis of Efacec's transformer development processes, with the goal of attaining a general understanding of the IM, KM and collaboration management strategies and instruments put in place.

This work should constitute a conceptual basis for the implementation of Digital Twin in the development of power transformers at Efacec, as well as contribute to the research related to this emerging technology.

The objectives tree for the dissertation is displayed in the following figure:

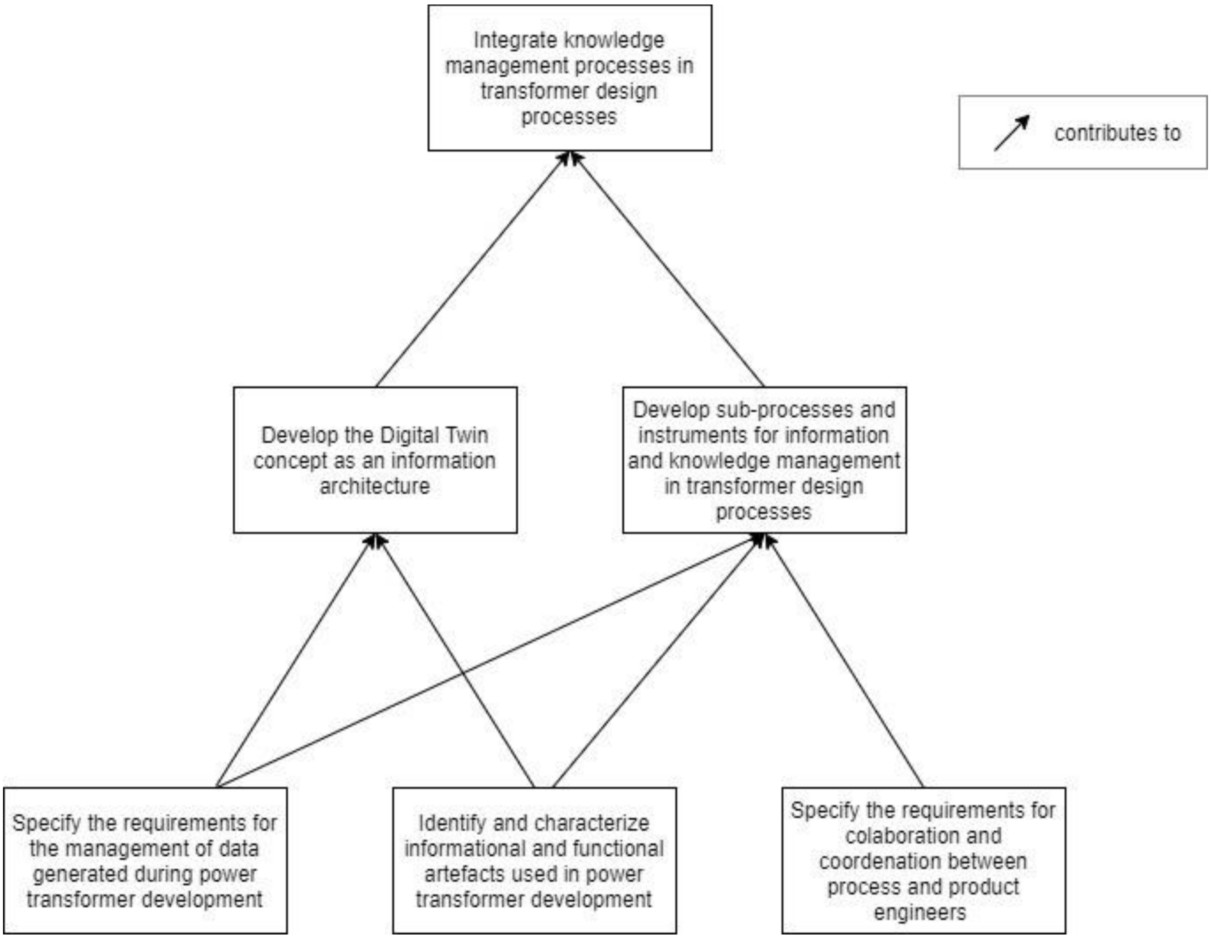


Figure 1- Objectives Tree of the dissertation

1.4. Dissertation structure

This dissertation is divided in eight chapters, starting with an introductory chapter that is followed by five chapters related to each theme and phase of the work developed, and a final chapter that details the conclusions and future work that can be carried out in the follow- up of this dissertation.

The introductory chapter addresses the context, motivation and objectives of the work carried out, and encompasses a description of each chapter and topic that makes up the structure of the dissertation.

This is followed by two state of the art chapters, which describe the most relevant work carried out in the academic and professional fields, in the areas of application of KBE and Digital Twin in product development, respectively. The first describes the concept of KBE, its application and utility in product development, common aspects with knowledge management, and implementation methodologies. The second addresses the concept of Digital Twin, including its initial applications, its potential utility in product development, how it relates with KBE and the technologies and methodologies necessary for its implementation.

The fourth chapter consists of the analysis of the power transformer development processes at Efacec, with a special focus on aspects of information, knowledge, and collaboration management. This includes a brief contextualization about the company and the power transformer, a general description of the tasks that these processes involve, the identification of the information accessed, generated and made available throughout PT development, and the characterization of the knowledge management strategies and mechanisms employed.

The fifth chapter involves the application of the Digital Twin concept to Power Transformer development, and this includes describing its core and KBE supporting features, along with a general explanation of how the DT will affect current Efacec processes.

Following this, the information architecture chapter includes the description of the information that the PT Digital Twin captures and generates, along with the thought process behind the selection of an information classification approach for DT information retrieval. Moreover, each category is described, and examples of query strategies are provided.

The dissertation is finalized with the conclusions and future perspectives of work in the field of DT applicability in product development.

2. State of the Art

2.1. State of the art methodology

The task of elaborating the state of the art consisted of 2 main stages. The first stage encompassed the tasks related to information retrieval, including the definition of search terms, the query of databases and the curation of information, while the second stage included the analysis of selected literature, the structuring of ideas, and the writing of the state of the art chapters.

These tasks were iterative, as the analysis of retrieved information eventually led to the identification of other useful search terms, which in turn, prompted new information retrieval tasks related to those terms.

Consequently, the scope of search terms was broad at first, and eventually became increasingly narrow as the analysis of information provided specific search terms. Some of these search terms are presented in the following table:

Search Terms
“Digital Twin”
“Knowledge-based Engineering”
“Digital Twin” AND “Knowledge-based Engineering”
“Digital Twin” AND “Knowledge Management”
“Digital Twin” AND (“product design” OR “product lifecycle”)
“Knowledge-based Engineering” AND (“product design” OR “product lifecycle”)
“Knowledge-based Engineering” AND “Knowledge Management”
“Digital Twin” AND “Information Management”
“Digital Twin” AND “product lifecycle management”
“Digital Twin implementation”
“Digital Twin technologies”

Table 1- State of the art search terms

The databases accessed to retrieve information were selected based on the quantity and quality of information that they offer, as well as the number of search result filters they provide. Ultimately, five databases were used, almost all of which being general databases that cover several subject fields in the computational, life, social and medical sciences. The exceptions are IEEE Xplore, which focuses on computational sciences and electronic engineering, and Library & Information Science Source, which is a specialized information science database.

The table below lists the databases accessed to support the work undertaken in this dissertation:

Databases
Google Scholar
IEEE Xplore
Scopus
ScienceDirect
Library & Information Science Source

Table 2- State of the art databases

The queries done in these databases involved the definition of a set of search result filters that ensured the relevance and quality of the literature. These filters were:

- English language;
- Peer-reviewed;
- Frequently cited;
- Published in Q1 or Q2 journals; *and*
- Published by a renowned author or organization.

While the language, peer-review status and citation count could generally be filtered by the database search engine itself, the journal rankings had to be checked manually. This was achieved by searching a given journal in the Scimago Journal & Country Rank website.

As for the professional literature used in the state of the art, it was determined that the credentials or notability of the author or organization that published it, was the most important factor in the reliability of the information, and thus, only professional literature that met this criteria, was selected.

After the relevant literature was retrieved, its analysis was done through an initial skim-read of the introduction, methodology and results sections, which was followed by more in-depth reading if the document was deemed to be useful. Additionally, portions of the text were highlighted, and notes were taken in the Mendeley software.

Finally, all this information was structured in topics, with the final state of the art focusing on the Digital Twin and Knowledge-based Engineering as they relate to information and knowledge management, with the goal of being a theoretical foundation for the work done in this dissertation.

2.2. The application of Knowledge-based engineering in product development

2.2.1. The KBE concept

Product development is a sequence of organizational, intellectual, and physical activities designed to conceive, design and market a product. Processes may differ between organizations, yet there are a set of strategies that most companies adhere to. These include the creation of prototypes that conform to gradually more detailed specifications, with the final output being a product that can be produced reliably by production systems. They also encompass information processing as the central aspect of development, as several design tasks rely on information such as organizational objectives, strategic opportunities, requirements, and available technologies, to be successfully completed.

Generally, product development processes are divided into 6 phases. The first consists of planning activities, and results in the definition of key project characteristics such as target market segment, assumptions, and restrictions. Then, the product concept is developed, including the selection of potential forms and functionalities of the product. Following this, steps related to product design are undertaken, both at the system and detail level, including the definition of the product architecture, its decomposition into components and the design of its geometric layout. Finally, the planning of materials to be purchased from suppliers is carried out, testing is done on product prototypes, and production processes are detailed (Ulrich and Eppinger 2012).

The phases that focus on engineering design are especially dependent on information and knowledge, as engineers use these resources to balance different design objectives related to aesthetics, manufacturability, and product functionality. This implies the integration and use of information in different formats and from different sources, as well as its refinement and communication between stakeholders until a desired solution is reached.

This workflow has an inherent problem at the design stage, and that is the difficulty in capturing and making available, the experiences, knowledge and information related to the other phases of the product lifecycle. If these information resources are not available to engineers, negative impacts can arise on later tasks such as production and transportation. Furthermore, since the initial design tasks include several moments of decision making, the lack of knowledge about the product, customer or processes can lead to changes in the work previously done, increasing the cost and lead time of the project (Chapman and Pinfold 1999).

In technological terms, current engineering design processes are usually supported by

CAD systems. Engineers make use of these computer systems to develop, analyze and modify a design, carrying out these tasks through inputs that change virtual geometric elements, such as size, location, and orientation of a component. Due to their functionalities, these tools increase the productivity of the product engineer, improve the quality of the design and facilitate the creation and storage of the documentation originating from these tasks (Narayan, Rao, and Sarcar 2008).

However, CAD systems are specialized in the geometric manipulation of product representations, with little ability to associate and manage information that goes beyond the parametric and functional design. Therefore, it is necessary to develop systems that combine the capabilities of both tools, offering geometric design features and the provision of knowledge regarding rules, engineering practices and manufacturing processes (Calkins, Egging, and Scholz 2000).

In this context, Knowledge based Engineering Systems were introduced to the market in the early 80's, being presented as a technological solution to capture organizational knowledge and encourage its reuse, as well as to automate routine tasks. KBES gradually merged with CAD tools, either integrating their capabilities or being themselves incorporated by those tools. As a result, it became possible to implement analytical procedures, communicate and collaborate with external tools, and formalize a domain of knowledge through a single platform (Reddy, Sridhar, and Rangadu 2015).

KBE became a popular research and development topic during the last decade of the 20th century and the beginning of the 21st century, and the literature of this period includes several definition proposals for this concept. These definitions focus on technological aspects, characterizing KBE as the programming of rules, good practices, and specialized knowledge, to automate parts of the engineering design.

This is evident in the work of Chapman and Pinfold (1999), who claim that KBE is an evolution of CAD, integrating these technologies with object-oriented programming and Artificial Intelligence. In this point of view, the purpose of KBE systems is the capture of information about the product and design processes, allowing the modeling of engineering design and the automation of some of its tasks. According to the authors, the primary objective is to capture the best product design practices and the experts' knowledge and experience in an organizational knowledge base (Chapman and Pinfold 1999).

A similar characterization was proposed by Cooper, Fan and Li (2001), with the authors describing KBE as a type of knowledge-based system supported by object-oriented programming and integrated with a geometric modeling tool. This software provides engineers with a structured approach to product design that automatically generates geometry through rules coded on the platform, reducing the time spent on these tasks (Cooper, Fan, and Li 2001).

Thus, both previous views envision KBE as an integrated CAD software, assisting the product engineer specifically in engineering design tasks. Other authors, namely Bermell-García and Fan (2002), share this idea, however, their vision for the usefulness of KBE also extends to other processes before and after product design, namely manufacturing, cost planning and sales (Bermell-García and Fan 2002).

Starting in 2007, a newfound interest in KBE emerged, and the volume of production of scientific and professional literature in this topic suggests that its popularity remains until today. This recent wave of KBE R&D also brought new ways of thinking about the area, with the technological vision giving way to a vision that characterizes KBE as a knowledge management methodology applied to engineering.

Among the authors who share this idea are Ammar-Khodja et al. (2008), who describe KBE as an engineering methodology that stores knowledge about product design and production techniques, in models. These models represent relationships between geometric attributes, safeguarding information about physical aspects of the product, such as types of material and functional restrictions, information related to development processes, and data from external sources such as databases (Ammar-Khodja, Perry, and Bernard 2008).

On the other hand, Rocca (2012), defines KBE as a knowledge management approach during product design, allowing the identification of relevant knowledge in an organization and defining how to capture, formalize, represent and store it, in order to improve its access, maintenance and transfer (Rocca 2012).

Another perspective points to KBE as a solution that captures the knowledge created and used during product development, as well as the tacit knowledge of specialized engineers. Through these technologies, it is argued that the reuse of knowledge for the development of a new similar product is facilitated (“Knowledge Based Engineering (KBE) – CADVision Engineers” 2019)

These most recent definitions correspond largely to the purpose of knowledge management, especially the focus on capturing and transmitting knowledge as a way to improve organizational performance (Girard and Girard 2015). Thus, the tendency seems to be to associate KBE with knowledge management practices in the context of engineering, with a secondary focus on the development of knowledge-based software.

Summarizing common aspects among the perspectives described above, it appears that

KBE definitions revolve around engineering supported by knowledge management and the software by which it is performed. These computer programs, called Knowledge Based Engineering Systems, store information extracted from processes and products, as well as the tacit knowledge of specialist engineers. The result is the creation of models that enable the automation of processes and tasks, reducing the time and money spent on product development.

2.2.2. Applications and results of KBE projects

The validity and importance of KBE has already been tested in several projects, obtaining a high level of success, and revealing a set of benefits from its implementation. The reduction in lead time and the decrease in costs of product design processes appear to be the biggest benefits resulting from the implementation of KBE systems, and these advantages are achieved through the automation of processes where information and knowledge are used intensively.

An application that resulted in the reduction of lead time was carried out by Emberey et al. (2007), who applied KBE to the design processes of aircraft fuselage panels. Among the most benefited tasks was the subdivision of the layers of prepreg composite material, whose traditional processes comprise about 20% of the total production preparation time. There was a 75% reduction in time from the beginning of these processes to the end, resulting in the ability to complete 100 engineering design cycles using KBE, in the same time necessary to complete only 10 cycles with traditional processes (Emberey et al. 2007).

Another implementation, carried out by Lin, Chan and Wang (2008), also allowed a drastic reduction in the time needed to complete product design processes. In this case, the authors developed a knowledge-based engineering design system for wire drawing matrices. This was implemented in a pre-existing CAD software, converting information related to the matrices into design parameters, geometric operations and, finally, into models of their components. The result was a reduction in design time from 10 days to less than 1 hour, as well as drastic cost reductions (Lin, Chan, and Wang 2008).

In addition to reducing product development time, other benefits related to the quality and validity of the design have been identified in literature. This is due to the ability of KBE systems to autonomously validate the design according to product restrictions and requirements, and also the possibility of the engineer using his time to develop innovative and creative designs instead of wasting it on repetitive tasks. .

This type of benefit was achieved by Van Der Elst and Van Tooren (2008). The authors implemented KBE techniques in the design and production of cable bundles for aircraft, with the objective of disseminating knowledge throughout the organization and reduce the risk of its loss, in addition to reusing it for the automation of repetitive processes.

The implementation took place in 6 steps, encompassing the analysis of processes, the capture, structuring and application of knowledge, the integration of KBE modules, the maintenance of the system and the training of users. This resulted in the operationalization of a tool capable of generating new product models and analyzing the quality and conformity of the design with pre-defined requirements (Van Der Elst and Van Tooren 2008).

The British luxury car brand Jaguar also uses KBE in a similar way to the case previously described, namely for the creation of new headlight designs and for the automatic verification of design compliance with legislation. Furthermore, the company created a specific tool that can conform the size and characteristics of headlight components with the solutions provided by the suppliers.

The KBE workflow supported by this application begins with the import of the outer surface of the headlight from the CAD system. Then, the user selects the supplier, market where the car will be sold and the headlight requirements, information that allows the system to query the supplier's database and provide possible solutions to the engineer.

The benefits of this tool are the drastic reduction in the time spent on assessing the feasibility of the design, which has gone from weeks to minutes, and the greater freedom of creativity of the engineers, since they do not have to worry about delays arising from compatibility problems between the design and capabilities of suppliers (Cooper, Fan, and Li 2001).

Thus, some the benefits achieved through the implementation of KBE are lead time reduction, decreasing of costs, improving the quality of product design, facilitating design validation and greater freedom for product engineers to use their time on tasks where they can use their creativity.

2.2.3. KBE implementation

2.2.3.1. KBE from the perspective of knowledge management

The implementation of KBE requires a knowledge management strategy that considers the different organizational and technical aspects of the company, including its area of activity, human resources, technological resources, and organizational culture. Zack (1999) characterizes these strategies as an attempt to achieve competitive advantages by balancing an organization's knowledge resources and capabilities, and the knowledge necessary to sell products and services superior to the competition (Zack 1999).

According to Hansen, Nohria and Tierney (1999), strategies can focus on technology, specifically on the implementation of reliable information systems through the reuse of coded knowledge, or on people, when the priority is the provision of creative and analytically rigorous recommendations to solve problems through specialized individual knowledge.

In the first case, the knowledge management strategy must involve the development of digital document management systems that encode, store and allow the reuse of knowledge, while the second consists of creating channels and networks that connect people in order to foster tacit knowledge sharing (Hansen, Nohria, and Tierney 1999).

Whatever the strategy, it will have to be predicated by a set of steps that guarantee that it will be employed and accepted at a holistic level in the organization.

According to Shankar et al. (2003), these steps should include the definition of an organizational structure that facilitates the intensive use of knowledge throughout the various departments, and this involves the implementation of instruments such as repositories of organizational knowledge, the establishment of mechanisms for knowledge sharing, and a platform that captures customer requirements. Furthermore, organizational culture should be adapted to focus on knowledge and its sharing (Shankar et al. 2003).

As a consequence of the implementation of knowledge management, it is expected that there will be an increase in knowledge retention, that its access will be improved, and that communication and collaboration between engineers will be more frequent and effective, which translates into obtaining competitive advantages.

However, there are several barriers to the successful implementation of knowledge management strategies and tools. Among these are the lack of support from the leadership of the organization, the lack of understanding and motivation on the part of employees for the sharing of knowledge, and an organizational hierarchy that inhibits or slows down knowledge sharing (Riege 2005) (Santos, Soares, and Carvalho 2012).

In the case of KBE, implementation is generally guided by methodologies that guide the knowledge engineer from the identification of the problem that knowledge-based engineering can solve, right through to installing the software or tool by which KBE is supported.

2.2.3.2. Methodologies

2.2.3.2.1. MOKA

The MOKA methodology is a proposed standard for the development of KBE systems, serving as a bridge between raw knowledge and software, through its framework for the capture and representation of knowledge (Reddy, Sridhar, and Rangadu 2015).

Oldham et al. (1998) describe the 6 steps that make up this methodology, which are listed below:

- Identify: Here it is determined whether the KBE system is necessary and, if so, what type should be implemented. The result is a conceptual specification for the KBE application.
- Justify: In this step, the project plan is created, and risk management is carried out. Both these tasks are undertaken in order to obtain approval from decision-making stakeholders in the upper levels of the organizational hierarchy.
- Capture: These tasks involve capturing raw knowledge and structuring it in an informal model composed of illustration, restriction, activity, rule, and entity forms.
- Formalize: In this phase, the formal model is created based on the informal model. This model is built using the MOKA Modeling Language, a UML variant.
- Implement: The KBE system is implemented.
- Activate: Following the previous step, the KBE system is installed, distributed and supported (Oldham et al. 1998).

Thus, the MOKA methodology comprises two types of frameworks, formal and informal. Informal frameworks propose an approach designed to transform human-understandable knowledge into formalized and validated knowledge in a model. In addition, they create a platform where knowledge engineers, specialists and software developers can share impressions and opinions.

The informal model consists of a set of ICARE forms, in which each letter of the acronym represents an aspect of the product or design process that must be modeled, specifically Illustration, Constraints, Activities, Rules and Entities. Illustration forms capture general knowledge, descriptions and comments, Restrictions forms reproduce interdependencies between entities, Activities forms express crisis resolution steps, Rules forms model organized knowledge and Entities forms define the structure, function and behavior of the product.

The formal model, on the other hand, encodes knowledge so that it can be represented and stored on a computer. This modeling is performed using MML, a UML style adapted to the MOKA methodology (Reddy, Sridhar, and Rangadu 2015).

Emberey et al (2007) applied the MOKA methodology for the development of a KBE System in the scope of the design of aircraft fuselage components. To this end, they adopted the vision of Stokes (2001), which divides this methodology into 4 types of knowledge management activities, namely identification, capture, storage, and reuse.

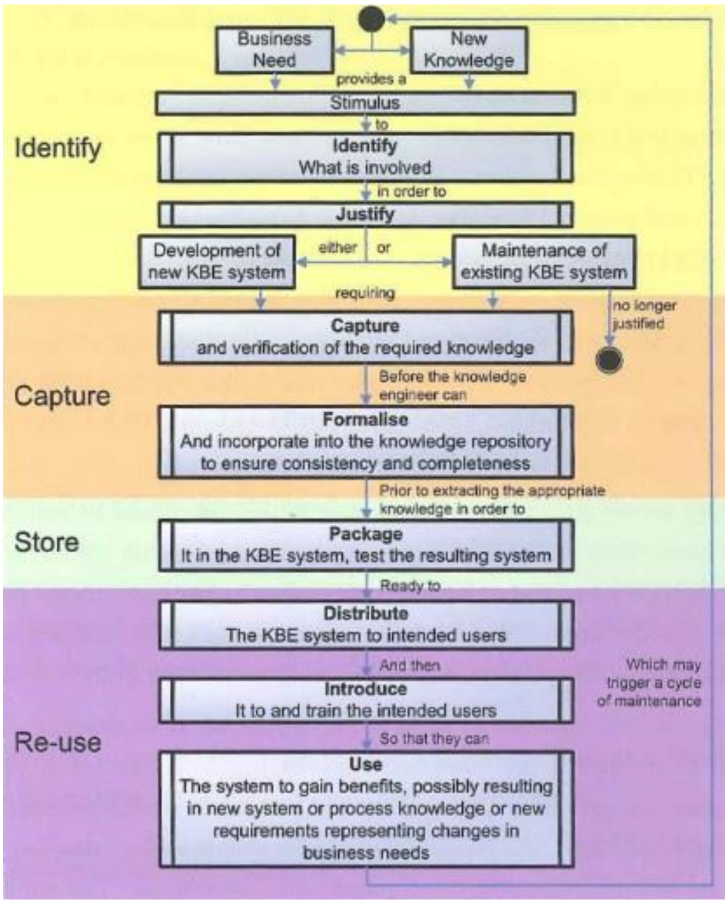


Figure 2- Phases of the MOKA methodology according to (Stokes 2001)

.

In the identification phase, the authors indicated the processes in which KBE could be applied and calculated the potential return that this would bring to the organization. To discern the processes that can be performed through KBE, a set of metrics were collected through the following guiding questions:

- Is the process routine?

- Is it based on formalized rules?
- Is it complex and iterative?
- How long does it take to complete?
- What will be the return on investment for the KBE application?

As a result of this phase, repetitive and rule-based processes were identified, such as the creation of a product definition dossier and the subdivision of prepreg into layers. The most significant benefit was a reduction in the time spent on these processes up to a maximum of 60%.

Then, the capture phase began, which was subdivided into the tasks of extracting, validating and structuring knowledge.

Knowledge extraction was carried out through interviews and meetings with specialists, subdivided into initial extraction and in-depth extraction. The initial structure of the knowledge base was created through an initial interview in which the knowledge engineer explained the purpose of KBE and the objectives of its implementation, followed by a second meeting in which the domain experts presented the processes, restrictions and production principles.

The deepening of the base took place through new meetings where the knowledge engineer and specialists segmented the processes into knowledge modules, with the latter solving a series of case studies through their explicit knowledge, allowing the engineer to identify this knowledge and define the rules that describe the problem solving strategy.

After the resolution of each case study, the knowledge engineer described the problem-solving strategy to the domain experts themselves, thus verifying the validity of the knowledge. Subsequently, the extraction of tacit knowledge from experts was attempted, through questions about used strategies and consequences of the rules applied.

The structuring of knowledge was carried out using the PCPACK software, a tool that makes ICARE forms available and interconnects them to create informal models consisting of activity diagrams and hierarchical diagrams.

Finally, the knowledge storage and reuse phases were supported by CATIA V5 software. This application uses informal models previously created in order to automate parts of traditional processes, producing different design solutions that the engineer can choose (Emberey et al. 2007).

2.2.3.2.2. CommonKADS

Another prominent methodology in the implementation of KBE systems is CommonKADS. Schreiber et al. (2000) describe its foundations, specifically the models and agents that it captures and builds.

The methodology base is built from 3 sets of questions designed to justify the implementation of KBE, to identify organizational knowledge, and to discern the characteristics of the software that will be installed. These questions are listed below:

- Justify the implementation.
- What use will the solution have?
- What problems will it solve?
- What impact will it have on the organization?
- Identify and characterize organizational knowledge.
- What is the type and structure of knowledge?
- How is it transmitted?
- Characterize the computer system.
- How will the knowledge be implemented in the system?
- What will be the system architecture?
- What features will the system have?

The answers to these questions are translated into 6 models that focus on the organization, tasks, actors, design of the solution, and knowledge and its communication (Schreiber et al. 2000).

Valentim (2019) details the functions and characteristics of these models, stating that the information used to complete them is collected through interviews, brainstorming and documentary research, and subsequently recorded in forms. The models, described through this author's view, are presented below:

- Organization model: encompasses the analysis of the company's characteristics, the identification of problems and opportunities and the assessment of the viability and impact of the solution.

- Task model: contains the identification of tasks that involve knowledge, which is done through the analysis of their logistics, including inputs and outputs, resources, and skills.
- Actors' model: describes the roles and characteristics of the humans or systems involved, including the skills and permissions they have to carry out their respective tasks.
- Knowledge model: describes the type and structure of knowledge that is applied in the execution of a task.
- Communication model: models communication between stakeholders.
- Design template: lists the specifications and technical requirements of the knowledge-based system (Valentine, 2019).

The CommonKADS methodology also includes the assignments that each stakeholder in the KBE implementation process must perform. A diagram that reflects these responsibilities can be found in the following figure.

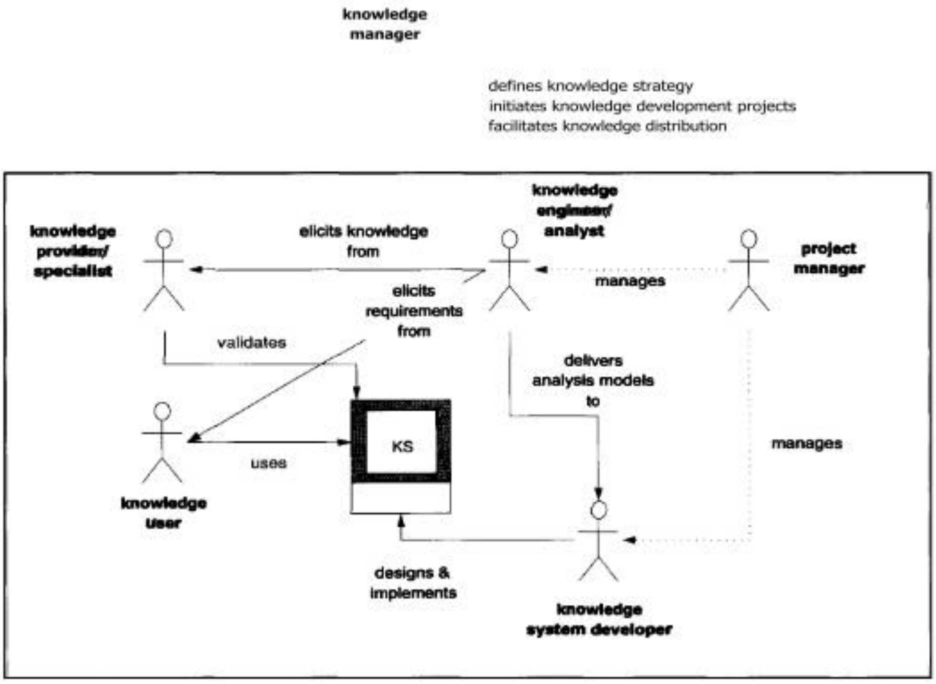


Figure 3- Role of the actors in the implementation of a KBE system according to (Schreiber et al. 2000)

The application of the KBE system includes 2 aspects, namely software development and knowledge analysis and modeling. The former is the responsibility of a programmer, while the latter is entrusted to a knowledge engineer. In both cases, this work is managed by the project manager.

The knowledge engineer extracts knowledge from specialists and raises requirements from users, transmitting the information collected to the programmer so that he can implement it in the software. After the work of the programmer is finished, the validation of the developed system is done by specialists (Schreiber et al. 2000).

2.2.4. Final considerations

KBE refers to the capture, storage, modeling and sharing of organizational knowledge, present in the minds of employees, or in rules, restrictions, rules of thumb or procedure manuals.

Through KBE and knowledge management, it is possible to automate some routine and time-consuming tasks of product design, reducing monetary costs and allowing engineers to devote more time to developing creative solutions.

The first KBE computer application that was available off-the-shelf was the Intelligent Computer Aided Design (ICAD) in the early 1980s, a system that is still used today by large companies like Boeing and that includes two interfaces, one dedicated to geometric design and another dedicated to the programming rules related to geometry (Sandberg 2003).

The tendency to add the engineering design capacity with features that automate and validate parts of it is still a reality today, with commercial KBE systems being generally integrated with CAD systems, complementing their functionalities. Among the solutions available on the market are GDL from Genworks, AML from TechnoSoft and Knowledge Fusion from Siemens (Sobieszczanski-Sobieski, Morris, and van Tooren 2015).

To implement these systems, several methodologies have been identified in the literature, with MOKA and CommonKADS being the most widely used. Although they differ from each other, both methodologies include steps to justify the implementation of KBE, identify, capture, and model relevant knowledge, and outline how it will be transmitted at the organizational level.

Other frequently implemented methodologies are KNOMAD, which focuses on knowledge from the perspective of its capture, standardization, quality control, organization through ontologies, modeling through Multi-Model Generator, analysis and availability (Curran et al. 2010), as well as KOMPRESSA, which includes guidelines, instructions, formats, techniques and tips for the analysis, design, modeling, implementation and documentation of complex knowledge (Reddy, Sridhar, and Rangadu 2015).

There are several approaches in KBE that aim to make knowledge management more efficient, providing tools that allow knowledge to be captured, modeled, and transmitted collaboratively by specialists in the organization. Among these are the application of online platforms. This approach was adopted by Kulon, Broomhead and Mynors (2006), who developed a KBE solution that uses a web platform and an integrated database for the design of hot forging processes.

Another KBE approach takes advantage of ontologies, sets of concepts and their relationships within a domain, for the capture and reuse of knowledge. Baumeister, Reutelshoefer and Puppe (2011) adopt this approach in the implementation of a semantic wiki as a collaborative tool to help decision making. In this case, ontologies are used to classify each wiki article and relate concepts within them. Through this strategy, a knowledge base is created on this wiki platform, helping employees to share or find solutions to a given problem (Baumeister, Reutelshoefer, and Puppe 2011).

Finally, case-based reasoning is often used to solve a problem through experience, information and knowledge acquired in a previous case (Riesbeck and Schank 1989). This approach begins with the recovery of similar cases, followed by the reuse of information and knowledge acquired in that case, the revision of the current proposal and, finally, the retention of experience that may be useful in the future (Aamodt and Plaza 1996).

2.3. The application of the Digital Twin in product development

2.3.1. The Digital Twin concept

The new industrial revolution currently being witnessed, a phenomenon also known as Industry 4.0, is the result of the development of new technologies that allow the improvement and automation of industrial processes, implying the development and continuous application of a set of intelligent systems and tools supported by Internet of Things, cyber-physical systems and cloud computing (Uhlemann, Lehmann, and Steinhilper 2017).

In the context of industrial change supported by new technologies, the concept of Digital Twin emerged, the first mention of which dates to 2002 by Michael Grieves. This author presented a conceptual model of product lifecycle management that includes a physical system and a virtual system that contains all the information related to the first. Because of this, all the information generated since the creation, manufacture, operation and disposal of the physical product also exists in its virtual copy (Grieves & Vickers, 2016).

Subsequent definitions of Digital Twin were revised and deepened to conform with technological advances and growing research in this area. Literature reveals two currents of thought, one that builds on Grieves' initial vision, focusing on Digital Twin as the junction of virtual and physical space through a connection where data flows, and another where Digital Twin is seen as a simulation or set of simulations supported by real-time data.

Gabor et al. (2016) argue that the increase in computational power allowed complex simulations to be carried out, integrating different models that focus on each aspect of the structural design in order to simulate the behavior of the entire system. The authors attribute this simulation functionality to the Digital Twin, noting that this technology is characterized by its ability to reliably simulate events at various scales of space and time, using specialized knowledge and data captured by physical systems (Gabor et al. 2016).

Weyer et al. (2016) state that Digital Twin is the next evolution of simulation, making it a central feature of production and operational systems throughout the product's lifecycle. According to them, all physical elements, from the product to the production facilities, will have a virtual twin that permanently represents their current state, and predicts their performance in the real world. In addition, the authors refer that all information created during these processes will be captured and stored, and then made available to stakeholders in a structured way through semantic description and metadata (Weyer et al. 2016).

The two perspectives previously described have in common the fact that they point to simulation as the central functionality of Digital Twin, as well as mentioning data as the main component of its operation.

Other authors base their definition of Digital Twin in the central idea that it is the virtualization of physical entities and the materialization of virtual models. This is the case for Tao et al. (2019), who see Digital Twin as the mapping of all components of the product lifecycle using physical and virtual data, and the information resulting from the interaction between the two (Tao et al. 2019).

In the academic and professional literature, the definition most often cited was proposed by Glaessgen and Stargel (2012), who defined it as an “integrated multiphysical, multiscale and probabilistic simulation of a physical system, using models, sensors and historical data to mirror the cycle life in a virtual environment ” (Glaessgen and Stargel 2012). Thus, this view seems to include both the idea of simulation and the virtualization of something physical through data and information.

Regardless of the line of thought where the authors are positioned, it appears that there are a set of characteristics and functionalities transversally accepted in literature as being inherent to the Digital Twin. Among these is its association with a physical element, so it is possible to conclude that it is not possible to have a Digital Twin without it being linked to one or more physical components. This connection is referred by most authors as being done through the transmission of data and information. In addition, simulation is seen as a feature invariably associated with Digital Twin.

2.3.2. First applications of Digital Twin

One of the pioneering organizations in the study of the possible incorporation of Digital Twin in organizational activities was NASA, which outlined the potential architecture and functionalities of this technology, as well as its applications in the context of aerospace engineering and, specifically, in the construction and operation of spaceships. A Digital Twin was built, and its informational architecture included the virtual representation of critical components of the vehicles, such as the avionics and propulsion systems, sensor data, maintenance history and other information related to the fleet. All of this allowed the tool to predict the probability of a mission's success, calculate the remaining life of a ship and recommend changes to a mission in order to improve these two metrics (Shafto et al. 2010).

A similar application was proposed by the United States Air Force, giving rise to a specific term called Airframe Digital Twin. As the name implies, this tool represented and modelled the mechanical structure of an aircraft and was developed to reduce maintenance costs and ensure the airworthiness of the plane. Among the information captured by this Digital Twin are the dimensions of the structure, the state of the materials that compose it throughout the lifecycle, the repairs and modifications made during service and other data such as the temperature and forces applied to the aircraft. . From this data set, calculations are made to predict the aircraft's ability to meet the requirements of a particular mission (Tuegel 2012).

Consequently, the Digital Twin is seen as a potential instrument of change for the aeronautical and aerospace sector, being especially useful for failure prediction and maintenance scheduling. However, the characteristics of this tool mean that its implementation in the Industry can fundamentally change business processes, not only in the tasks previously mentioned, but also in the design of the product (Tao et al. 2018).

2.3.3. Applicability of Digital Twin in product design

2.3.3.1. Potential Features

Studies on the application of Digital Twin in product design are still in an embryonic stage, although there are already several authors whose work describes its potential usefulness at this stage of the product's lifecycle.

Tao et al. (2018) describe the current design processes and the problems that are intrinsic to them. According to the authors, product design refers to all the steps taken from the conception of the product until it is validated for production, and two different trends that govern the way product development is carried out exist. In the traditional trend, professionals carry out various tests to guarantee the validity and usability of the design of a given product, making judgments based on their experience and knowledge, while the current trend is to place customers at the center of these processes, with their preferences, needs and feedback being central to the final product design.

Because of this, modern design processes benefit from virtualization, visualization, and networking to use a large amount of information to achieve an optimized design. However, the authors argue that these processes based on Big Data place emphasis on the analysis of physical data, with an absence of convergence between the physical product and the virtual space. In addition, the collection of feedback is usually done by crowdsourcing of users with professional knowledge, losing the suggestions and experiences provided by other non-specialized users (Tao et al. 2018).

As a way of solving these problems, Tao et al. (2019) propose the division of product development processes into three segments supported by Digital Twin, namely the task clarification, conceptual design, and virtual verification phases. Figure 4 presents the information and data applied in each phase of product design supported by the Digital Twin.

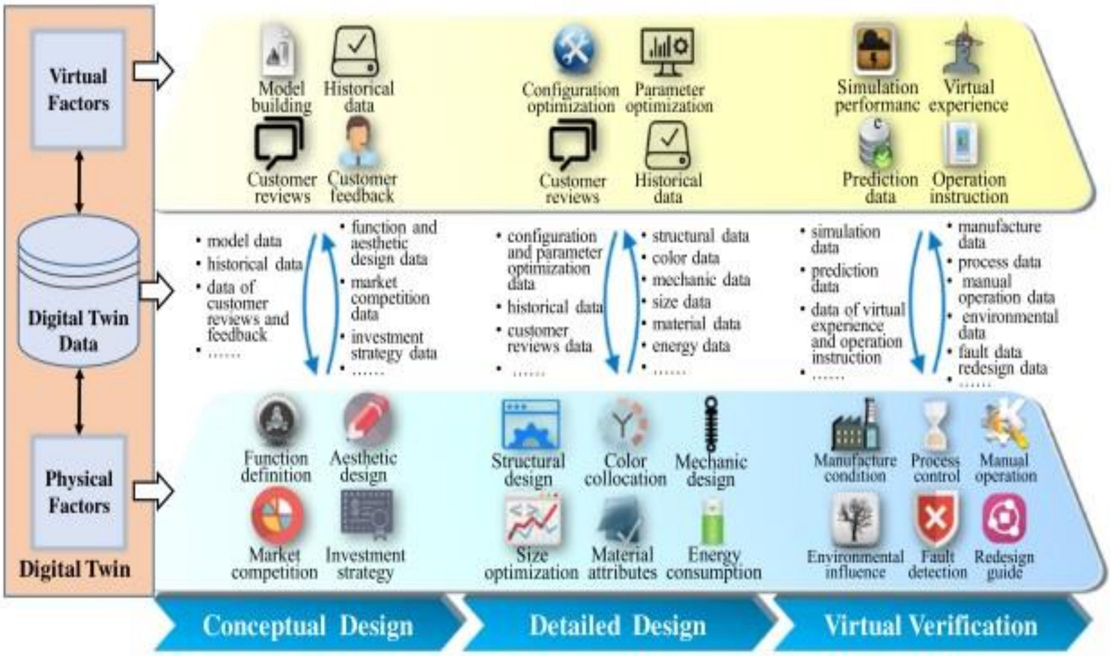


Figure 4- Data captured by the Digital Twin at each stage of product design (Tao et al. 2018)

In the initial design phase, customers' needs are translated into functional requirements governed by design restrictions. Feedback is obtained through online platforms, such as e-commerce websites and social networks, and other information, such as the activities and habits of users, is retrieved directly from devices based on the Internet of Things, such as smartwatches and smartphones. This Big Data is necessary to increase the customization and adaptability of the product, and so, its processing is crucial.

In this step, it is proposed that the Digital Twin incorporates a feature recommendation system that analyzes the opinions of the customers and infers desirable characteristics for a product. Furthermore, it is argued that Digital Twin can contribute to the definition of functional requirements by making available critical data such as the number of times a feature is accessed and for how long it is used, and that it can synthesize design restrictions such as weight and height. Finally, it is suggested that Digital Twin may allow the analysis of user interaction with a product through technologies such as Virtual Reality, eliminating the need for costly ethnographic observations.

After the task clarification phase, conceptual design is carried out, and this includes tasks that guarantee that the design is logically viable, functionally simple, and physically correct.

During this stage, it is suggested that the Digital Twin can perform analysis of data captured by sensors to identify and assess the severity of design contradictions, that it can capture the uncertainties of the real world and simulate them in the virtual world, and that it may use real-time data to identify, diagnose and correct problems resulting from poor product design decisions. Crucially, the authors claim that Digital Twin can help create the conceptual design itself, by allowing engineers to use contextual usage data, such as time and temperature, to compare virtual and real contexts and understand the ideal conditions for using a product.

The last phase of the design is the virtual verification stage. At this time, the evaluations aimed at reducing inconsistencies between the expected and actual behavior of a product are carried out. The Digital Twin is described as a way to verify and simulate all aspects of the design, using historical and real-time data to progressively and iteratively improve virtual models. As a result, the production of a small amount of product for testing is no longer necessary, reducing money and time costs (Tao et al. 2019).

Other authors present a similar perspective on the applicability of Digital Twin in product design. Canedo (2016) presents it as a mechanism to manage IoT systems and devices, arguing that as the number of objects with a Digital Twin increases, the value of the information originated by them decreases, while data related to the context where they operate, specifically their spatio-temporal relations, increases. Therefore, the author argues that the feedback provided by Digital Twin on how products are used by users and how they deteriorate in different environments can dictate the design specifications of new products (Canedo 2016).

In this same aspect of the design, Schleich et al. (2017) conclude that the Digital Twin is a viable solution to link engineers' knowledge with customer preferences, enabling the verification of conformity between product specifications and project requirements (Schleich et al. 2017).

Boschert and Rosen (2016) discuss the possible application of Digital Twin in mechatronics. The first version of the design of the components of a device is usually done through the main requirements and the experience of the engineers, and in all the following steps, these have to be constantly validated by the requirements and premises of the project. To carry out this validation, models structured with the basic requirements and the functional decomposition of the design of each component, together with their interactions and dependencies between them, are created. With the gradual increase in the complexity of the components and the subsequent increase in people from different areas involved, the need for consistency between models increases, In this design aspect, the authors state that the ability to store, link and make data and information available allows Digital Twin to be a unique repository for all digital artifacts created during the development of the device, thus streamlining the design process and interconnecting various fields of knowledge (Boschert and Rosen 2016).

Tao et al. (2018) also share the idea of Digital Twin as a single source of information during product design, however, they also include the data related to the customer, such as their satisfaction and feedback, and about the company, like the number of sales of each product, in the information that the Digital Twin should store. This information has the purpose of helping the engineer understand where and how to improve the product. Furthermore, they argue that the Digital Twin, specifically its functionality for viewing 3D models integrated with real-time data, allows clients to have a greater understanding of the design, which results in greater input from them, about it (Tao et al. 2018).

Rosen et al. (2015) propose the storage and compilation of these conceptual models created during product design in Digital Twin, not only as a way to validate the product's specifications, but also to help create the procedures and operations to be carried out during the manufacturing phase of the product (Rosen et al. 2015).

An additional application for Digital Twin was identified in the design of production lines. Zhang et al. (2017) conceptualize the simulations supported by Digital Twin as an intermediate task between the input and output of the design of a hollow glass production line, allowing the optimization of systems, equipment, parameters and variables of the line (Zhang et al. 2017).

Linking these perspectives, there seems to be agreement among authors that the Digital Twin can aid in the definition of characteristics and functionalities of the design, taking advantage of the data collection in the physical space and its processing in the virtual space, to carry out simulations and suggest optimizations, ensuring the compatibility and validity of the product with project and user requirements.

2.3.3.2. Complementarity with KBE

The complementarity between KBE and Digital Twin has not yet been analyzed in the literature, however, taking into account the features proposed for the Digital Twin in the product design phase, especially the ability to act as a recommendation system, storing and providing information which helps the engineer define functional requirements and conform them to the project's objectives, there appears to be a possible complementarity between KBE and Digital Twin, with the latter serving as a support platform for knowledge-based processes.

In this perspective, Digital Twin hosts information, knowledge and data originating from various sources, carrying out its processing so that it can have a beneficial impact on engineering design. This includes the data captured by sensors installed in the physical twin, the tacit and explicit knowledge of specialists, information related to the needs and preferences of the clients, rules and restrictions of the design, and historical information about previous projects.

On the other hand, Digital Twin is not restricted to capturing and analyzing data to automate tasks or make recommendations. This technology can also foster the sharing of knowledge through the storage and availability of design artifacts throughout the product lifecycle and encourage collaboration through platforms where the different teams can communicate with each other and coordinate tasks.

Thus, through KBE supported by the Digital Twin, problems inherent to product design, such as its interdisciplinary nature and the need to make critical decisions at different stages of the design, are addressed due to the availability of structured and standardized information to engineers, and the possibility for them to communicate and collaborate with different teams at the organizational level. As a result, this approach takes into account not only the technological aspect related to the management of organizational knowledge, but also the human aspect, promoting not only the capture and sharing of information between the person and the digital system, but also between people through the same system.

2.3.3.3. Digital Twin from the point of view of Information Management

Given the definitions and functionalities attributed to the Digital Twin by professional and academic authors, is it possible to assert that the DT captures and structures all information related to a physical product, becoming a digital informational construct that is a separate entity of the product, yet is linked to it through its entire lifecycle.

Consequently, this information varies in format and size, and describes the product in its entirety, ranging from its geometrical specifications to its behavior under various circumstances.

Nevertheless, the beginning of the Digital Twin information lifecycle precedes the existence of the physical product, as the DT starts being developed in the product design phase. Specifically in this phase, the Digital Twin makes use of product requirements, 3D models, material specifications, design task lists and service records to virtually model the physical product that will later be produced. This virtual model is then tested by means of simulation, generating simulation models that enable the detection of design faults and potential unwanted behaviors.

Following the production and operationalization stages, the Digital Twin becomes a true reflection of the physical product, mimicking every mechanical and functional aspect of it in real time, as well as storing historical data originating from its operation. This includes information that describes the geometry of the product, a list of current and past components, and records of operations and services performed on the product.

Moreover, the information flow between the digital twin and physical twin becomes bidirectional in this stage, with sensors installed on the PT transmitting data to the DT, which then uses data analysis algorithms to generate meaningful insights regarding the product's performance. Ultimately, these insights are used for operation prediction and optimization purposes.

Finally, as the physical twin reaches the end of its lifecycle and is discarded, the information captured by its Digital Twin remains stored and available to design engineers, allowing them to take advantage of the acquired knowledge to create feasible, flexible and optimized designs in future projects (M. Grieves and Vickers 2016).

In this perspective, the Digital Twin's lifecycle is comprised of four stages, namely:

- Modelling stage: The DT models every aspect of the physical product, including geometry, materials, and behaviors.
- Testing stage: The DT uses simulation to test product performance under a variety of scenarios.
- Operation stage: The DT mirrors the product's condition in real time, predicts failures and generates operational insights, by receiving and analyzing sensor data from the product.
- Post disposal stage: The DT is no longer linked to the physical product but continues to store the information captured during its lifecycle for future access by design engineers.

As previously established, the Digital Twin is an informational construct. Thus, is it important to detail the stages of the information lifecycle over the Digital Twin's lifecycle. The lifecycle of information varies according to the characteristics and purpose of the information, yet it is generally possible to identify five distinct phases.

- Creation: Information is captured from internal or external sources.
- Storage: Information is classified, organized, and stored.

- Retrieval: Information is retrieved by a query mechanism.
- Usage: Information is used to fulfill an information need or is distributed.
- Retirement: Information that is no longer useful is discarded.

During the DT lifecycle, the creation phase includes the capture of both internal and external information. The internal information is captured during product development, production, and operation, and consists of design artefacts, maintenance and service records, and sensor data. On the other hand, exterior information is captured exclusively during product development, including requirements defined by the client, as well as norms and laws that restrict product design.

The storage of this information is challenging, as it varies greatly in format, purpose, and size. For example, information such as norms and design task lists are mostly textual information embedded in pdf files, sensor data is numerical, and mechanical models are usually 3D representations of the product created with CAD. Furthermore, a single document might contain different types of information that are useful for engineers of various specializations, and these engineers might face difficulties in retrieving a given document if they do not understand the metadata associated with it. As a result, documents and data must be organized and classified in such a way that it takes into account all useful information contained therein and permits its fast and efficient retrieval.

During the design phase of the physical product, information retrieval and usage is of utmost importance, as engineers need to access requirements and norms to guarantee the conformity of the design with project specifications, and also check historical information generated during previous product development projects to identify optimized design characteristics. In the case of the Digital Twin, another important source of information are the insights generated from sensor data from the physical twin. This data is mainly used during product operation, to monitor product performance, predict failures and define optimized operation parameters and maintenance schedules

Lastly, the retirement stage of information management does not coincide with the deactivation phase of the physical product, because, as previously stated, the information gathered during its lifecycle can be useful for future product development projects. However, this information can become obsolete if a particular product is discontinued or changes massively over time, in which case deletion may be appropriate.

2.3.4. Implementation of a Digital Twin

2.3.4.1. Methodologies and technologies

Having analyzed the potential of Digital Twin in product design, it is equally important to discern necessary frameworks, workflows and technologies identified in academic literature for the development of these solutions, especially for the modeling of physical components and the capture and transmission of data.

Specifically in the aspect of modeling the physical product to the virtual space, Moreno et al. (2017) enumerate the steps that led to its implementation in a case study of sheet metal drilling machines, detailing the tasks related to the representation of physical components in 3D, modeling of the behaviors of the machines, representation of knowledge and simulation.

For visual modeling, CAD models of the machines were segmented into 3D parts, which were subsequently agglomerated into scene graphs and saved in XML. In this same format, ontologies were created with the purpose of modeling the behavior and parameters of the machines, as well as all the other elements with which they interact. Then, the effects that each machine tool has on the metal sheets were mapped using computational geometry algorithms. Finally, all of these models were integrated and connected in a user interface built using the Qt software, resulting in the possibility of generating reliable simulations (Moreno et al. 2017).

Schroeder et al. (2016) suggest the use of an XML data format called AutomationML for the creation, storage, and transmission of the necessary models in the different phases of the product lifecycle. In a case study related to an industrial valve, they applied a methodology divided into 3 steps, involving the modeling of physical components, the implementation of a communication channel between systems and, finally, the availability of data.

In the first step, modeling was carried out using AutomationML, mapping the physical components whose operation would produce data such as voltage and temperature that would later be transmitted to other systems. Then, it was ensured that the data would be transmitted through an IoT middleware called FIWARE. The last step consisted of receiving this data through a system designed to run on computers.

The authors state that the use of this methodology enables users to model and operate a Digital Twin without programming knowledge (Schroeder, Steinmetz, Pereira, and Espindola 2016).

Additionally, Schroeder et al. (2016) define the Digital Twin architecture from the perspective of data management, focusing especially on how they can be viewed through Virtual Reality. The authors list 5 layers, which are the device, user interface, web services, search, and data repository.

The first layer consists of the devices that run the virtual reality systems, while the second is associated with the interface through which users access the Web services. These services constitute the third layer, and it is where informational resources are available. The last two layers are related to data, incorporating databases, and searching mechanisms using languages such as SQL.

This architecture is presented in a case study that focuses on the visualization of data originating from the extraction and processing of oil and gas on an offshore oil rig. These raw materials are stored in tanks for which a Digital Twin is built, and the tanks transmit data in real time about their gas, oil and water level. This information is sent to the tank control system and is also transmitted to the Digital Twin database and its web interface. Thus, this information, in addition to historical data such as year of construction and maximum capacity, is available online, with communication between the web interface and the VR system allowing the visualization of these data through virtual reality (Schroeder, Steinmetz, Pereira, Muller, et al. 2016).

Another implementation was carried out by Haag and Anderl (2018), who achieved a proof of concept by building a Digital Twin for a beam test platform.

In the first phase, the virtual representation of the platform was developed by CAD software and simulated to ensure that all aspects of its operation were mirrored virtually. The connection between the physical and virtual twin was made through a *publish-subscribe* architecture based on the MQTT message protocol, and an IoT platform was introduced as an intermediary for the control of both physical and virtual components, as well as to present the test results on the beam test platform.

The tests were started with the input of parameters in 1 of 2 variables, namely force applied to the beam or displacement reached, and this information was distributed to the physical twin who subsequently performed the test until the selected value was reached. After the test, the physical twin recorded the value of the variable that was not selected, sending it to Digital Twin. Afterwards, it autonomously analyzed the results, sending them to the IoT platform so that an intermediary could compare them with the physical results.

The authors consider that the concept of Digital Twin was validated with this experience, although further progress is needed in areas such as the capture and analysis of Big Data. In addition, they argue that a product must be accompanied by a Digital Twin template throughout all of its development processes (Haag and Anderl 2018).

Other authors idealize the generic functionalities that Digital Twin should include so that it can be an asset for an organization, without associating them with a specific product or area of activity.

Tao et al. (2019) indicate not only the characteristics that the Digital Twin should include, but also the steps to be taken to build it. Figure 4 specifies the technologies applied in the different stages of the development of a Digital Twin, according to these authors.

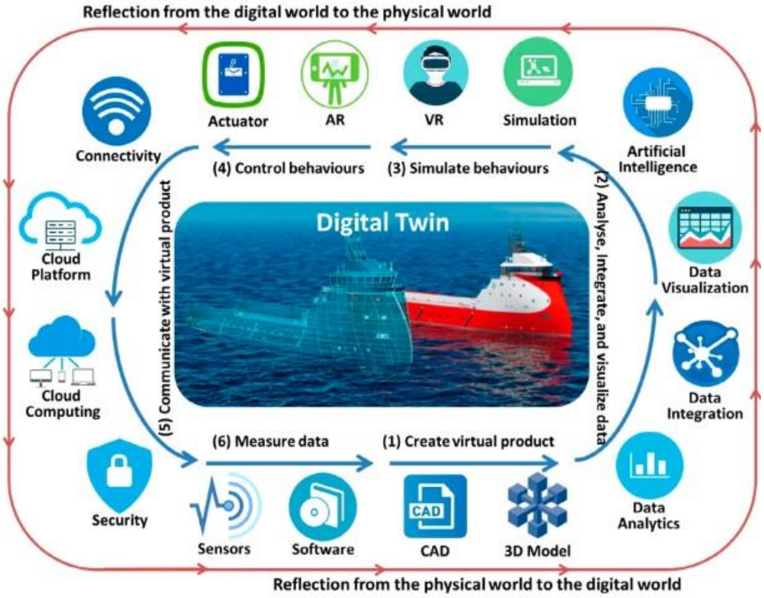


Figure 5- Processes and technologies applied in the development of a Digital Twin according to (Tao et al. 2019)

The authors mention the construction of the virtual representation of the physical product as the first task to be performed, and this representation must include geometric and physical models, the product and user behavior and the forecast, optimization and evaluation models. Then, data from a variety of sources must be integrated and analyzed using data analytics techniques, allowing engineers to use it to assist in decision making. At this stage, Artificial Intelligence can also be implemented to provide simple recommendations. Then, the product's behaviors must be simulated, either through software or through virtual reality technology, and these simulations should result in the real-time transmission of optimized behaviors to the physical product. The development cycle is continued with the continuous collection of physical product data and information related to the customer and their interaction with it. This information is fed to the first step, thus creating a continuous optimization loop (Tao et al. 2019).

2.3.4.2. Potential problems

The complexity of the technologies that Digital Twin hosts and the changes they bring to the way an organization works, imply the need to address a set of potential problems that may arise before, during and after its implementation.

Grieves (2019) points out the main problematic areas, indicating the security flaws against industrial espionage, the excess of non-useful data collected and the non-interoperability of information representations in different Digital Twin as crucial aspects to be solved (MW Grieves, 2019). In turn, Schleich et al. (2017) focus on the problems involving data, namely the difficulty in capturing and treating large data sets, making predictions in complex systems and synchronizing the virtual with the physical world (Schleich et al. 2017).

Other problematic situations involving the organizations themselves have been identified, such as the difficulty of change management, since it is crucial to prepare and educate workers, customers and suppliers for a new way of working, and the uncertainty regarding the total costs of implementing a Digital Twin ("Digital Twins in Logistics" 2019).

2.3.4.3. Implementation cases

Despite its potential problems, the implementation and integration of Digital Twin in the different phases of the product lifecycle is already a reality in some organizations. An example of the application of this solution was carried out by CNH Industria, a multinational that produces vehicles for agriculture, industry, and commerce. In order to improve the reliability of the robots that weld the vehicle chassis, the company implemented a Digital Twin of the production line that included the different types of chassis and the welding requirements of each, the welding stations in the line and the robots installed there. With the input of data collected by sensors and the production planning system, Digital Twin uses simulation and machine learning to predict the probability of each component failing ("Digital Twins in Logistics" 2019).

Another implementation was undertaken by Siemens, one of the most active companies in the Digital Twin area, which used it for the planning, operation, and maintenance of a Finnish power system. This Digital Twin is available in an interface where network analyzes are carried out, ensuring its security and reliability, and it is also possible to view a virtual map of the network that is constantly updated with historical and real-time data ("For a Digital Twin of the Grid" 2017).

In the area of renewable energies, GE Renewable Energy has developed software that creates virtual models powered by data from physical wind turbines. The company uses these

Digital Twins to simulate how environmental factors, such as wind, affect the operation of wind turbines, as well as to control their performance. As a consequence, engineers can make decisions regarding turbine design and operation more quickly and effectively ("Improving Wind Power with Digital Twin Turbines", undated)

2.3.5. Final considerations

Digital Twin is an emerging research and development topic, with most of the literature published after 2016. The concept was originally conceived in the early 2000s, however, it was technological development that aroused greater interest in this area , since only the proliferation of sensors and advances in simulation and data analysis made it possible to transform the Digital Twin from an idea to a possibility. Its definition varies according to the author, however, the basic idea of Digital Twin as a digital version of a physical product, which simulates its visual aspect and behavior based on bidirectional data transmission, is transversal in scientific literature. .

The first proposal for implementing this solution occurred in the aeronautics sector, with the creation of digital twins of critical aircraft components, with the purpose of predicting possible failures and scheduling maintenance in a timely manner.

In product design, the usefulness of Digital Twin materializes in several phases, from ideation to testing and validation. The central idea present in the researchers' work is Digital Twin as a single source of information, aggregating customer requirements, user feedback, data originating from the use of products in the real context and design restrictions and rules. All this information is analyzed, providing the engineer with the diagnosis of contradictions or problems in product design, and recommending optimized characteristics for the product. Ultimately, the simulations supported by Digital Twin allow the product to be validated, turning the need to produce product prototypes for testing and validation purposes obsolete.

Considering the various definitions and potential features attributed to the Digital Twin, its development must necessarily incorporate modeling aspects between physical and virtual twins. This modeling must be carried out on its mechanical properties, behaviors, operating parameters, rules and restrictions, and the tools identified to do so include CAD software, ontologies, and XML formats such as AutomationML.

Furthermore, the transmission and synchronization of data between the physical and digital space must necessarily be ensured, with the case studies analyzed proposing publish-subscribe architectures and IoT middlewares to accomplish this goal.

Moreover, implementation problems are identified in the literature, with some authors reporting technological problems, and others indicating possible organizational problems. The technological problems stem mainly from the fact that the transmission and processing of large amounts of data are still areas that need further development, although there are other critical aspects to be resolved, such as the danger of industrial espionage and the collection of useless data. For the organization, the biggest obstacles are change management and uncertainty about the total investment needed to implement a Digital Twin.

Even though the Digital Twin is a recent concept, its use in a business context is already a reality today. The real applications found are related to the construction of vehicle chassis and the maintenance and operation of a power grid and a wind farm. However, it was not possible to find a concrete case of implementation of Digital Twin in design processes, which suggests organizations have so far prioritized the development of these solutions for other stages of the lifecycle such as the manufacture and service of the product.

3. Methodological Approach

The methodology employed for this dissertation encompassed three main iterative phases, namely:

1. Elaborating the state of the art;
2. Analyzing Power Transformer development processes; *and*
3. Developing the Digital Twin concept.

The analysis of power transformer development processes was supported by semi-structured interviews with several Efacec stakeholders. These interviews were conducted following a 4-step methodology:

1. Defining information needs;
2. Identifying stakeholders;
3. Scheduling interview; *and*
4. Conducting interview.

The definition of information needs considered the dissertation objectives, as well as the gaps and inconsistencies identified following each interview. Because of this, all interviews had the purpose of not only acquiring new knowledge regarding Efacec's processes, but also of building on topics discussed in previous interviews.

After a set of information needs were defined, Efacec was contacted in order to identify the stakeholder most qualified to answer questions related to those needs, and also to schedule the interview.

The interviews were conducted in a semi-structured manner, as general questions were asked, and answers given by the interviewee served as the foundation for the elaboration of increasingly specific questions.

Following the identification of information and knowledge management processes, which included establishing what information is created and captured during PT lifecycle, and what KM instruments are used, the development of the Digital Twin concept begun.

The application of the Digital Twin concept to PT development was accomplished through an iterative methodology that was supported by the state of the art and the knowledge acquired during PT development process analysis. Thus, bottlenecks identified in current Efacec information and knowledge management processes, were juxtaposed with the features and applications proposed for the Digital Twin in scientific literature.

This task resulted in the definition of features and technologies that the power transformer Digital Twin should incorporate to be beneficial to PT development, and in the proposal of a set of KM instruments supported by the DT.

The final task was the development of an information classification scheme, as a tool to organize DT information. This was achieved through the analysis of the types of documentation that the DP captures and generates, specifically the information and concepts contained in each document, as well as information formats and purposes.

In sum, the development of the Digital Twin concept involved defining functionalities, technologies, and information architecture. Moreover, it included detailing modifications to current PT development processes, so that they could be supported by the DT. These objectives were materialized through the analysis of the state of art regarding KBE and DT application in product design, and the study of Efacec IM and KM processes in PT development.

4. The development of power transformers at Efacec

4.1. Organization description

Efacec Power Solutions is a Portuguese company that operates in the energy, mobility, and engineering sectors, and is present in more than 65 countries.

Its origin dates to 1905, the year in which A Moderna, Mechanical Sawmill Society was founded. This company evolved to Electro-Moderna, Lda. in 1921, starting the production of motors, generators, transformers, and electrical accessories, at that time.

The name Efacec emerged in 1962, and the company has since experienced a gradual growth that has resulted in the agglomeration of a group of companies under the name Efacec Power Solutions, covering all means of production, technologies and technical and human skills in the areas of energy, engineering, environment, transport and electric mobility.

Currently, the organization invests in a strategy of innovation and constant adaptation to maintain its status as a large industrial company at the national and international level, which is reflected in its mission:

We create value with Energy, Environment and Transport solutions that improve everyone's daily lives, through the integration of different skills and the most innovative technologies. We develop people in a learning and continuous improvement organization.

In Portugal, Efacec operates in 3 industrial centers, namely the Polo da Maia, Lagoas Parque and Parque da Arroteia, accounting for a total of 2400 employees.

4.2. Power transformer description

A power transformer, also called three-phase transformer, is a device that transmits electrical current from one voltage level to another without changing the circuit's power.

The principle that governs the operation of the machine is not the movement of its components, since they are static, but the electromagnetic induction between a main coil and one or more secondary coils wrapped in a ferromagnetic core. This results in the reception of alternating current electrical energy by the main coil, which converts it into magnetic energy, followed by the reconversion of this energy into electrical current by the secondary coil.

The voltage conversion can be carried out from the lowest to the highest, or vice versa, with the transformers responsible for this conversion being called step-up or step-down transformers, respectively.

The most frequent locations where step-up transformers are installed are electrical substations, with the purpose of efficiently transmitting energy over long distances, while step-down transformers are often used in factories, shopping centers and homes to reduce voltage to the level of operation of appliances and other equipment (Georgilakis 2009).

4.3. Organization structure and task delegation

The power transformer development projects at Efacec involve a set of iterative processes that require interaction and collaboration between members of the product development team, the departments linked to the different phases of the TP lifecycle, and the customer. Because of this, the design activities of transformers are transversal to the organization, involving the commercial, project management, research and development, product development, and production departments.

The commercial department is primarily responsible for the initial communication with the customer, obtaining orders and carrying out a basic survey of requirements and technical specifications.

Project management works together with the product development department to create and potentially complement the project dossier, contacting the client after each milestone of the work to inform him of the status of the project and to collect his feedback.

The product development department performs all engineering studies and calculations that contribute to the design of the transformer, and then transmits all the necessary specifications to the production team. This division of the company is composed of a leading engineer, 3 senior consultants, and a set of engineering cells composed of one team coordinator and a varying number of engineers.

If there is no suitable calculation method to match the specifications of a given transformer, the product development team collaborates with the R&D team to create a new method or modify an existing one. Following this, the R&D team is also responsible for testing the transformer design.

Finally, the production department translates the specifications provided by the product development department into a transformer ready for operation.

The following figure shows the communication and collaboration between the different teams involved in the development of power transformers at Efacec, as well as with the customer.

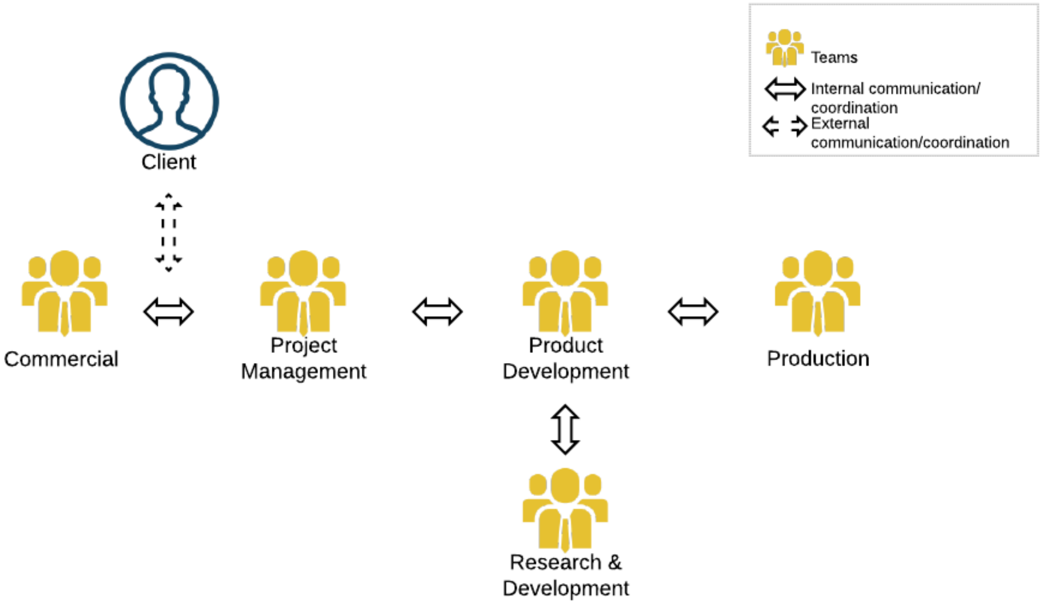


Figure 6- Communication and collaboration between Efacec departments

4.4. Power transformer development workflow

Four categories of tasks can be distinguished in the development of power transformers at Efacec, the first relating to requirements gathering and definition of project characteristics, followed by those that focus on engineering design and laboratory testing, and ending with the construction of the transformer. Mapping these tasks to the product development steps defined by Tao et al. (2019), it is possible to discern the phases related to product design, specifically task clarification, conceptual design and virtual verification, and the production phase (Tao et al. 2019).

4.4.1. Task clarification

The task clarification phase aims to identify and validate all the requirements of the transformer, as well as allocate the necessary resources, in order to correspond to the customer's expectations.

This stage begins with the elaboration of a project proposal by the commercial department, which involves close collaboration with the client in order to raise all his requirements, needs and expectations for the transformer, as well as to define other aspects of the project such as deadlines and budget.

In case the proposal is accepted by both parties, the commercial team sends all the necessary information to start the transformer design to the Project Management Department, which then clusters all information in a project dossier. Following this, the Project Management team appoints an engineering cell to carry out the project, assigning a coordinator and a group of electrical and mechanical engineers to the team.

The coordinator's first responsibility is analysing the information contained in the dossier to identify flaws or errors that make the project, product definition or planning unworkable. If this is the case, the coordinator works together with the project management team to review and complete the document until a satisfactory version is reached, informing the client of the changes made.

Subsequently, task clarification ends with a project presentation meeting for the team, where roles are assigned to each employee, and task planning is carried out.

4.4.2. Engineering design

The design of the transformer includes the activities of engineering calculus and mechanical design and has the purpose of creating a set of electrotechnical and mechanical specifications that are in accordance with the requirements of the project and whose construction is feasible.

The initial conceptualization of a transformer is achieved through the engineering calculus, from which parameters such as number of coil turns, size of the magnetic circuit and electromagnetic induction values are determined.

This task is performed by a single electrical engineer, who is assigned the responsibilities of identifying the calculation method most appropriate to the project specifications and performing the necessary calculations from it. Eventually, an optimized calculation method may not exist for the specifications of a given transformer, in which case, the engineer asks the R&D department to develop a new one.

After calculating the theoretical parameters, the mechanical design is performed using CAD and CAE software, developing the aspects of the transformer related to the dimensions, positioning and behavior of the machine components.

Finally, the information created during the design tasks, specifically the electromagnetic and mechanical models of the transformer, are aggregated in order to be tested.

4.4.3. Virtual Verification

Design testing and validation is accomplished through a set of simulations based on the Finite Element Method, a technique that assesses the behavior and structural integrity of the machine, under different operating conditions.

Based on the results of the analysis, a report is created containing a set of recommendations for the design of the transformer. These recommendations make it possible to optimize the design, for example, in terms of the mass of the materials or the size of the reinforcements, as well as ensures that the machine will meet the requirements and expectations of the customer when it is in operation.

4.4.4. Production

Production processes convert models created during PT development, into an operational machine. Production is done gradually, component by component, and after each stage, errors or problems are reported to the product development team.

Based on encountered problems, which may result from machinery limitations or inconsistencies in the design, changes are made to the transformer design to adapt problematic aspects.

4.5. Information management

Information management practices during PT development projects start in the task clarification stage. Here, the project coordinator gathers all requirements negotiated between Efacec and the client, along with other project aspects such as schedule and budget, and creates a document called Project Dossier. This document has both a digital and paper copy.

Following this, the calculus and mechanical design generate a set of design artefacts that describe the power transformer. This information is stored in a software developed inhouse called WinTree, and exported and sent via email to the client, in pdf format. Other information created during PT development, such as purchase lists, are also stored in this software.

In the WinTree application, information is organized by project, and each project has 4 folders containing information about each aspect of the power transformer, namely about active components, exterior components, welded construction, and purchase lists.

As such, this information base is restricted to information that the production team uses to construct the power transformer, and thus, does not include information captured and created during other development tasks that are not part of engineering design. Consequently, information such as client feedback, is lost in instruments that do not support the storage or query of information, such as email.

4.6. Knowledge management strategies and instruments

Knowledge management at Efacec encompasses a set of instruments designed to identify the mistakes made during the process of developing transformers, and to enhance and share the acquired knowledge to various stakeholders.

These instruments prioritize knowledge management through the implementation of instruments that require interaction, communication, and collaboration, between members of each engineering team. As a result, it is possible to state that Efacec employs a personalization strategy for knowledge management.

4.6.1. Personalization strategy

The personalization strategy is characterized by its strong human component, focusing on the interaction between employees of the engineering team, especially between more experienced and higher ranked members, with lower level ones. This contact occurs on three occasions:

- Kaizen meetings: Weekly meetings between the team coordinator and engineers.
- Design Review Meetings: Meetings held after the completion of the design of the interior and exterior components of the transformer, with the participation of senior consultants and the team coordinator.
- Training / Preventive Actions: Training meetings held after feedback is sent by the production department, with senior consultants, the team coordinator, and the engineers present.

Kaizen meetings are part of a philosophy of continuous improvement, involving interaction between all members of the team. The central purpose of this instrument is to ensure that everyone is informed about the status of the project, specifically the tasks and objectives that must be achieved in a given week, and also to identify and jointly solve potential obstacles to its realization.

These interactions allow the sharing of knowledge related to the solution of a given problem and ensure that all members are aware of the responsibilities and competences of their colleagues, facilitating collaboration and coordination between the team.

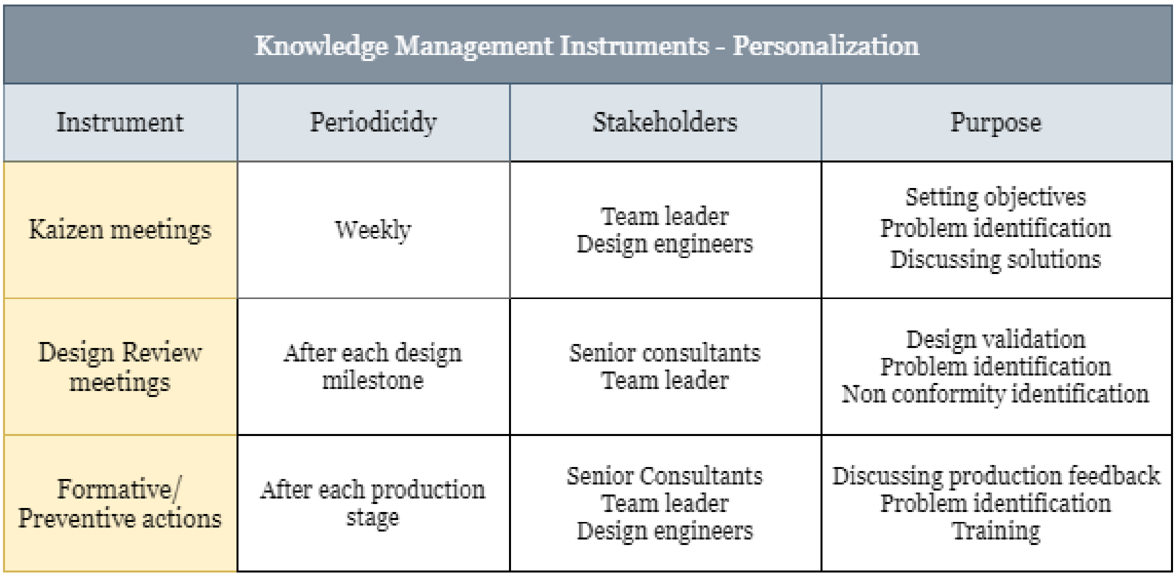
Another instrument that is inserted in Efacec's knowledge management strategy is the Design Review meetings. These have the purpose of validating the design of the transformer, with the identification of problems or nonconformities in charge of senior consultants, the most experienced members of the engineering department. After this validation, the design moves on to production, if there are no errors, or returns to the engineering team to make the necessary changes.

Finally, the central instrument of knowledge management in the organization are the training actions, carried out after receiving feedback from the production department. This feedback consists of listing all the problems encountered during the construction of the transformer, which may be the result of equipment limitations or flaws in the product design itself.

Following the identification of the problems, the senior consultants and team coordinator define the changes or adaptations to be implemented in the design of the transformer. Then, they carry out a training action for the remaining members of the team, describing the problem, the reason for it to have occurred and how it should be solved.

Consequently, these actions are a means of transmitting knowledge, serving a preventive function, as the acquired knowledge is used to prevent the same errors from occurring in the future.

The following table describes the knowledge personalization instruments employed in power transformer development processes at Efacec.



Knowledge Management Instruments - Personalization			
Instrument	Periodicidy	Stakeholders	Purpose
Kaizen meetings	Weekly	Team leader Design engineers	Setting objectives Problem identification Discussing solutions
Design Review meetings	After each design milestone	Senior consultants Team leader	Design validation Problem identification Non conformity identification
Formative/ Preventive actions	After each production stage	Senior Consultants Team leader Design engineers	Discussing production feedback Problem identification Training

Table 3- Efacec knowledge personalization instruments

4.7. Bottlenecks

The analysis of PT development processes revealed bottlenecks that can have a negative impact on the final design of the machines, as well as missed opportunities in the utilization of organizational knowledge to decrease lead time and improve PT performance and design. Among these critical issues are:

- Data generated during PT operation is not captured.
- The search of instructional documentation such as manuals is not possible, as this information is not indexed, and no metadata is associated with it.
- Sharing of acquired organizational knowledge such as lessons learned is mostly done through personalization strategies, and as such, this knowledge is not formalized in documents.

Operational data captured by sensors installed in Power Transformers is a valuable source of knowledge, as the insights generated by data analysis algorithms can be effective in predicting failures and recommending optimized operational parameters and maintenance schedules. Furthermore, these insights, combined and contextualized with other design information, could be made available to design engineers, which would then have a centralized knowledge base that supported their decision-making during PT design.

However, current design processes at Efacec are not supported by operational data, as this data is managed solely by the entities who control its operation. As such, the opportunity to create an important knowledge source for continuous improvement of PT design, should be exploited.

Another problematic aspect of PT development is the difficulty in retrieving manuals and other instructional documentation. This information is made available in the software engineers use during PT design and is only accessible as a pop-up window that opens after a “Help” button is clicked. Thus, no search mechanism is offered, which creates an inability for engineers to manually search instructional information regarding a given topic, component, or process.

Finally, knowledge management instruments implemented by Efacec are focused on the sharing of knowledge through in-person interactions between engineers, and the formalization of this knowledge is not standard practice.

The codification of this knowledge in document format could allow engineers to quickly search organizational knowledge, creating benefits such as improved decision making and more frequent and efficient knowledge sharing.

5. Applying the Digital Twin concept to power transformer development

The application of the Digital Twin concept to power transformer development involved identifying key features and capabilities of this emerging technology, followed by the analysis of current Efacec processes, and finally, coupling the identified DT capabilities with Efacec KM and IM bottlenecks, to establish PT development processes that actively employ organizational knowledge to improve PT design.

Two types of DT features are proposed, namely those whose main purpose is to use organizational knowledge to improve PT development and operation, and those that focus on capturing and disseminating knowledge, and as such, can be seen as KM instruments supported by the DT.

An overview of the proposed DT functionalities in each PT lifecycle phase, is presented in the following figure:

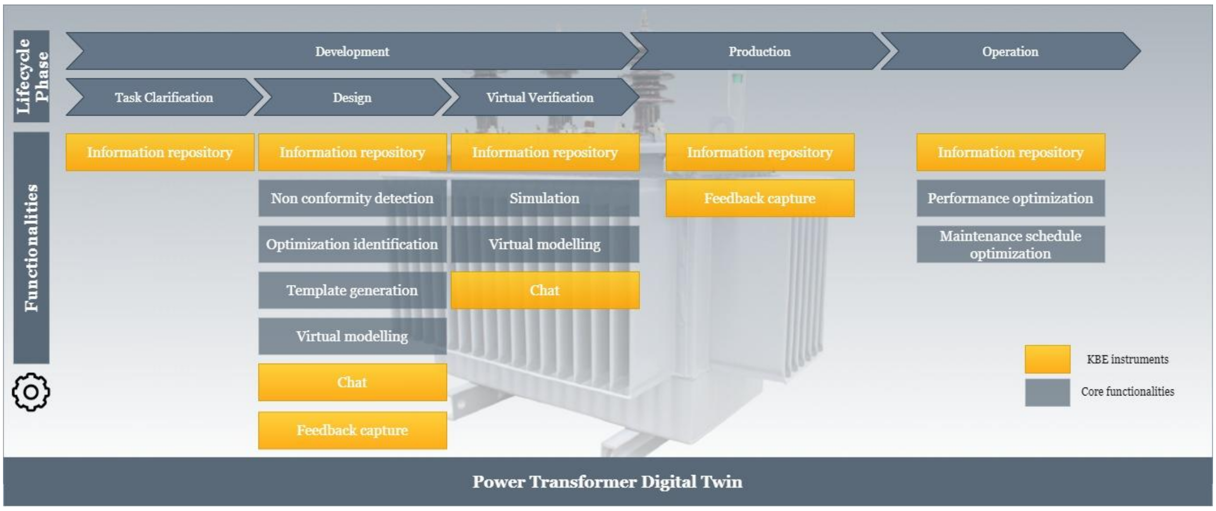


Figure 7- Overview of power transformer Digital Twin Functionalities

The proposed DT features imply that this technology serves a supporting role in PT development, helping engineers as they perform mechanical design and virtual testing, as well as optimizing PT performance after it is operational. This, along with functionalities that function as KM instruments, means that the DT can be seen as a Knowledge-based Engineering System able to capture, employ, and make available organizational knowledge to automate, optimize and streamline PT development tasks.

The following sections describe each proposed DT functionality and specify the workflow of Digital Twin supported PT development tasks.

5.1 Core Functionalities

5.1.1. Power transformer virtual modelling

Given the definitions attributed to the Digital Twin concept by academics, it can be asserted that this technology is composed of three main components, namely the physical power transformer, a digital model that accurately represents it, and a bidirectional data transmission channel that connects both. Thus, the Digital Twin is required to include a virtual modelling functionality.

In order to virtually model a physical product in a reliable and exact way, the Digital Twin must be able to replicate all its properties. In the case of the power transformer, this involves modelling the machine's geometric, mechanical, and electromagnetic characteristics.

Geometric modelling refers to the appearance, dimension, and position of the components, while the mechanical properties include the movements each component is able to make, and the type and mass of the material that it is composed of. Finally, the electromagnetic properties of the PT should also be replicated, and this involves modelling the behavior of the electric circuit and magnetic forces that the machine handles while in operation.

Furthermore, a crucial part of the implementation of the Digital Twin is the establishment of a bidirectional data transmission channel between the virtual and the physical power transformer, which enables the Digital Twin to receive sensor data from the physical transformer and send real time commands to it. This entails information security risks, as this data is extremely valuable, and can be subject to industrial espionage. Consequently, steps should be taken to ensure that the communication between the DT and the PT are encrypted.

The final aspect regarding DT modelling is the platforms in which this application can be visualized and utilized from, as the contexts in which the Digital Twin will be accessed require that it should be able to be visualized through multiple devices.

For instance, a design engineer will most likely work in an office and use a desktop computer as his main working tool, while a process engineer's work involves him being mobile on the shop floor and therefore, requires him to use portable electronic devices such as laptops, tablets and smartphones.

Moreover, virtual reality can enhance the Digital Twin's functionalities, especially those related to KM and collaboration management. By allowing its visualization through this type of display, the Digital Twin improves the ability of design engineers to understand the PT design and behaviors, decreasing potential mistakes and making the design process more interactive.

Because of this, the Digital Twin should be a focal point of current Efacec KM instruments, with Kaizen, Design Review and Training meetings involving the visualization of the Digital Twin through VR headsets. Ultimately, the visualization of the PT in a context that resembles its actual operational context, allows engineers to better understand and explain design decisions, which culminates in the improvement of knowledge sharing.

Additionally, if this feature is made available to the client, his understanding of the design can also improve, making the identification of needed changes or nonconformities easier, compared to current practices, which involve using 2D design models to update the client on the project status.

5.1.2. Design support

A common feature attributed to the Digital Twin in academic literature is its ability to actively support product design tasks. Works like those of Tao et al. (2019), Canedo (2016), and Schleich et al. (2017), propose several functionalities that use operational data, requirements, and restrictions, to detect nonconformities, suggest optimizations and validate product design.

Power transformer design can benefit from these functionalities, through 3 distinct DT features:

- Creating a PT design template according to engineering calculus specifications;
- Identifying design nonconformities, according to norms, legislation, and requirements; *and*
- Suggesting optimizations supported by sensor data.

Current PT design processes involve an initial phase of engineering calculus, in which a set of specifications are defined for the machine, and a mechanical design phase, where the geometric and mechanical properties of the PT are created according to calculus specifications.

This process is routine regardless of project, as although no PT conceptions are the same, machine specifications generally lead to similar designs. Because of this, an opportunity arises to use historical data to automate and streamline PT development tasks, which results in the decrease in time and monetary costs.

The benefits of this opportunity can be materialized through a DT supported design workflow, composed of 4 main tasks:

1. Engineers perform engineering calculus;
2. Based on calculus specifications, the DT automatically lists previous PT development projects, according to similarity;

3. Users select a previous design to serve as a template for the current project; *and*
4. Users adapt nonconformities.

This process allows engineers to quickly reach an initial PT design, ultimately freeing them to focus on customized design features, specific to the current project. To achieve this, the DT requires Artificial Intelligence algorithms to recognize similarities between PT specifications, and previous PT designs.

Furthermore, since design templates will usually have minor nonconformities with current project specifications, they will have to be corrected, and thus, the DT should also be able to recommend design adaptations that accomplish this goal.

Identifying and recommending changes to PT design based on project requirements will entail the coding of this information on the DT platform, which is usually done in a "If x, then y" format.

Supported by this DT functionality, potential mistakes, inaccuracies, or nonconformities in PT design, are resolved as follows:

1. DT detects design problem;
2. DT alerts user; *and*
3. User manually adapts the design or commands the DT to autonomously do so.

In this point of view, as PT mechanical design is performed, the DT analyzes the design and checks for compliance with design rules and requirements previously coded in the application. When a nonconformity is detected, the DT notifies the user with an alert that includes the type of problem encountered, such as "client requirement nonconformity", the components that are affected by this problem, and a link that opens a pop up containing the norms or other documentation that the infringed design rule originates from. Lastly, the user can opt to manually change the design, or instead, command the DT to autonomously do so.

The following figure presents a use case diagram of the PT mechanical design process, illustrating design tasks supported by the Digital Twin's design template generation and nonconformity detection features, from the user's perspective:

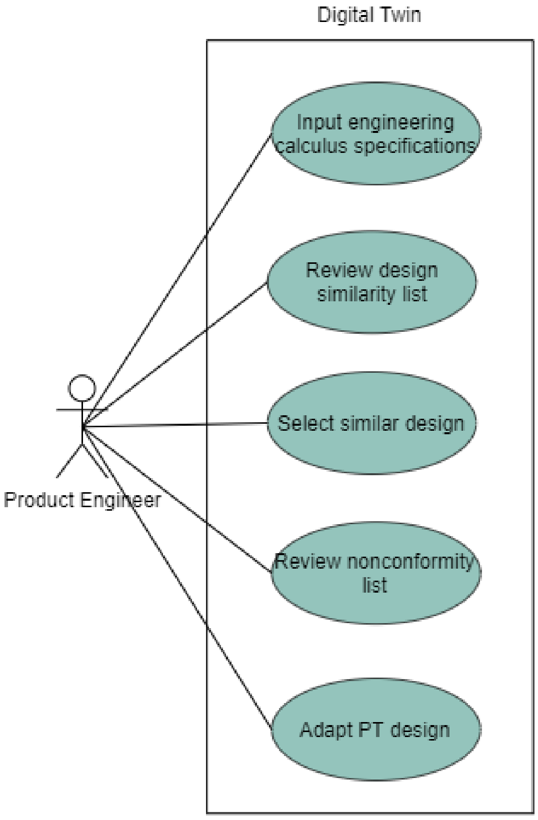


Figure 8- Use case diagram of the Digital Twin's design template generation and nonconformity detection features

This process can also be employed for design optimization. As previously stated, the DT must be connected to the physical PT through a data transmission channel, and as such, it is able to capture sensor data during operation. This sensor data, which is unuseful by itself, can be analyzed by the DT through data analysis algorithms and integrated with other PT lifecycle information, resulting in insights regarding the optimized design of each component of the machine. As a result, the DT design optimization feature has the following workflow:

1. DT detects optimization opportunity;
2. DT alerts user; *and*
3. User manually adapts the design, commands the DT to do so, or opts not to take action.

Similarly to the nonconformity identification functionality, the DT optimization feature analyzes PT design as it is being performed, comparing the current design with the insights gathered during the operation of power transformers that share key design specifications. Supported by Artificial Intelligence, the DT can recommend changes to design, and this is done through an alert that notifies the user of the components and parameters that should be adapted. Finally, based on his judgement and experience, the user decides whether to make the changes recommended by the DT, or not.

5.1.3. Simulation

Historically, product testing implied the development and experimentation of prototypes and usage environments, in an attempt to detect and predict possible design flaws and unwanted behaviors. This often led to the destruction of the prototype, as well as failures in the detection of detrimental behaviors, which would only be discovered when the product was already in operation.

Recently, technological development has made it possible for simulation software to accurately mimic the behavior of a product in an array of operating conditions. This makes the production of prototypes almost entirely obsolete, which consequently reduces monetary costs. Additionally, as the complexity of products becomes greater, their behaviors, specifically behaviors that are unexpected and unwanted, can be efficiently identified through simulation (M. Grieves and Vickers 2016).

A Power Transformer is a complex machine, as it is composed of a large number of components, and its correct operation requires a precise balance between its mechanical and electromagnetic properties. Because of this, the PT can exhibit many unexpected behaviors, and consequently, simulation presents itself as an efficient method for PT testing.

As a virtual model of the PT, the DT is able to perform these simulations, thus fulfilling the primary purpose of this technology according to works such as Weyer et al. (2016), Gabor et al. (2016) and Glaessgen and Stargel (2012).

The DT supported virtual verification stage of power transformer design, has an iterative nature, and begins after the first iteration of PT design is finished. In total, 4 core tasks can be identified to achieve the goal of validating PT design through simulation:

1. Simulation is performed on the first iteration of PT design,
2. Simulation insights are employed to optimize the first design iteration;
3. New design iteration is tested; *and*
4. Previous steps are repeated until an optimal PT design is reached.

The DT functionalities that support these simulation tasks are explicit in the following use case diagram:

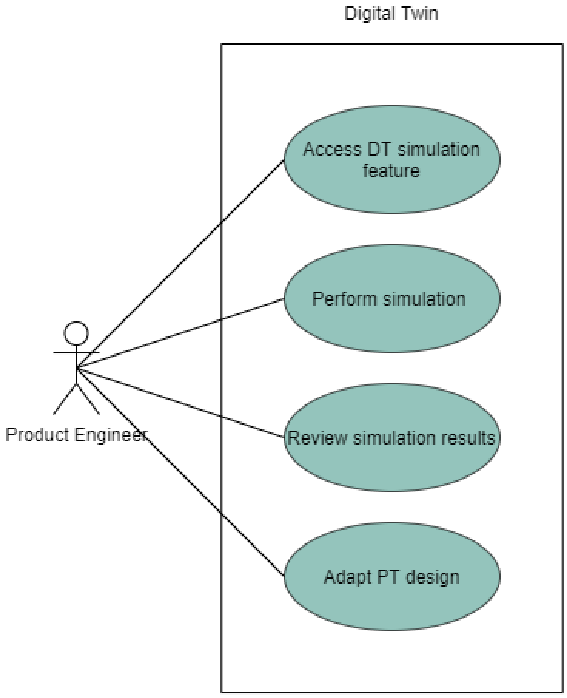


Figure 9- Use case diagram of the Digital Twin's simulation feature

As a result of this virtual verification workflow, designs are gradually upgraded as the DT simulation feature identifies design flaws, optimization opportunities, or unwanted PT behaviors.

5.1.4. Operation optimization

The DT's ability to capture sensor data yields benefits not only for DT design optimization, in the manner described before, but also to optimize PT performance when it is already in operation. This is especially relevant for maintenance scheduling and real time monitoring, where several shortcomings hinder PT operation.

Present-day maintenance scheduling is usually defined according to PT specifications, as expected failures or the need for replacement of components varies according to the characteristics of the power transformer. This does not consider the actual performance of the machine nor the environment in which it operates, and consequently, the PT may perform sub optimally for a period of time before the scheduled maintenance tasks are performed.

The Digital Twin's capabilities present an opportunity to solve this problem, by employing data analysis algorithms and artificial intelligence to provide optimized maintenance schedules based on PT sensor data. As such, a DT supported maintenance schedule can be defined through the steps listed below:

1. Product Engineers draft a first iteration of the maintenance schedule;
2. PT starts operation and DT begins capturing and analyzing sensor data;
3. DT integrates sensor data insights with the PT specifications;
4. DT employs Artificial Intelligence to optimize maintenance schedule; *and*
5. Client is notified of updated maintenance schedule;

This workflow can be further streamlined after the DT has captured data over a period of time, from a range of PT's with different specifications. Supported by this data, the first step may be replaced by a DT functionality that uses Artificial Intelligence to automatically define a maintenance schedule, before the DT has started operation. Hence, this functionality uses historical sensor data from PT's with similar designs, and integrates this information with other PT specification data such as the materials and dimensions of the components, allowing the algorithm to infer a first iteration of an optimized maintenance schedule.

An additional DT functionality that can potentially improve PT operation is real-time monitoring. Since the DT captures Big Data originating from various sensors installed on the power transformer, this data can be streamed with little to low latency to Efacec and to the client, allowing for the quick detection of performance issues.

5.2. KBE supporting functionalities

5.2.1. Information repository

Power transformer development processes involve decision making moments that require stakeholders to have access to the appropriate information that enables them to make the right decisions, thus ensuring that the PT design meets the client's expectations.

One of these critical decision making moments is the start of the engineering calculus task. Here, engineers need to consider all PT requirements defined in the task clarification stage, so that the type of engineering calculus method applied, as well as the final calculus specifications, are in accordance to them. Currently, this information is initially provided to engineers in a project introduction meeting, and then sent to them via email. This does not safeguard this information for future storage, nor does it provide a centralized and structured information access instrument.

Another development stage that encompasses important decision making aspects is mechanical design. During this stage, engineers must make decisions regarding design properties such as dimensions and shape of each PT component, in an attempt to reach an optimal design that complies with the restrictions arising from project requirements. Because of this, having access to project information that influences design specifications, is also crucial in this stage.

Moreover, additional decision-making support can be found in design artefacts created in previous PT development projects, as engineers can compare current project specifications with previous ones and draw insights from them.

Consequently, it is proposed that the Digital Twin include an information repository that stores all information captured in external sources, or generated by the DT, during the entire PT lifecycle. This grants engineers a centralized information base that expands along the development and operation of the power transformer and facilitates the use of information as a major supporting tool for engineering design.

This feature has already been proposed in academic literature in works such as (Tao et al. 2018) and (Boschert and Rosen 2016), although these authors limit the scope of DT captured information to design artefacts and other information useful for design tasks.

This dissertation assesses the Digital Twin as an informational entity that replicates the entirety of the PT's lifecycle, and consequently, the information stored in this platform is not restricted to the development phase, instead encompassing all phases, from task clarification to disposal.

Further work related to the DT information architecture is detailed in chapter 6, specifically the identification of all captured and generated information, along with the proposal of an information classification scheme for the DT, which also functions as the information retrieval query mechanism of the DT information repository functionality.

5.2.2. Interactive comments/feedback

Like previously stated, Efacec's knowledge management instruments heavily rely on person-to-person interactions for knowledge sharing purposes. Therefore, most of the knowledge acquired in each PT development project is present solely in engineer's minds, as opposed to being formalized in a physical or digital document. This entails problems such as loss of organizational knowledge due to employees leaving the organization, as well as a difficulty in identifying stakeholders who possess potentially relevant knowledge for a given design task.

Because of the DT's capability of capturing and storing PT lifecycle information, this application can also be a crucial instrument to create documents that formalize organizational knowledge and make them available to engineers during PT development.

During design tasks, product engineers often use their experience and tacit knowledge to support their decision making, with the goal of achieving a design that satisfies project requirements. As a result, the formalization of the thought process behind PT design decisions can be extremely valuable for future projects, allowing engineers to use those knowledge assets to support their own decision making, and ultimately, optimize PT design.

Considering this, it is proposed that the DT include a feature that enables both product and process engineers to leave comments regarding different aspects of PT design. In this perspective, product engineers will formalize their thought process according to these steps:

1. User performs PT design;
2. User selects PT component; *and*
3. User completes *Insight* form.

This DT feature is prompted by a product engineer to describe the thought process behind a certain design aspect that he deems essential or unusual. To successfully share the knowledge that guided his decision making, the engineer selects the PT components in question, followed by the *Insight form* option. This form should include information about which components it relates to, as well as the name and team of the stakeholder that created it. This metadata can be automatically filled by the DT, based on the components that the user selected, along with the information in his DT profile. Moreover, the form should allow for the user to write his thought process in natural language, as this facilitates comprehension by other Efacec stakeholders.

Not only does this DT feature promote knowledge sharing between product engineers during TP design, but it can also be a key collaboration instrument for product and process engineers, during PT production.

The collaboration between product and process engineers is focused on the identification of problems that affect PT production, and the communication of these issues to product engineers so that PT design can be adapted. Currently, this is accomplished through email communication, which offers a reliable communication channel, yet does so in a noninteractive way.

To provide a more efficient and interactive collaboration process, product engineers should also have access to the DT platform, and the identification of production problems related to PT design should have the same workflow as the previously described product engineer comment functionality. DT supported collaboration tasks between product and process engineers are explicit in the use case diagram below:

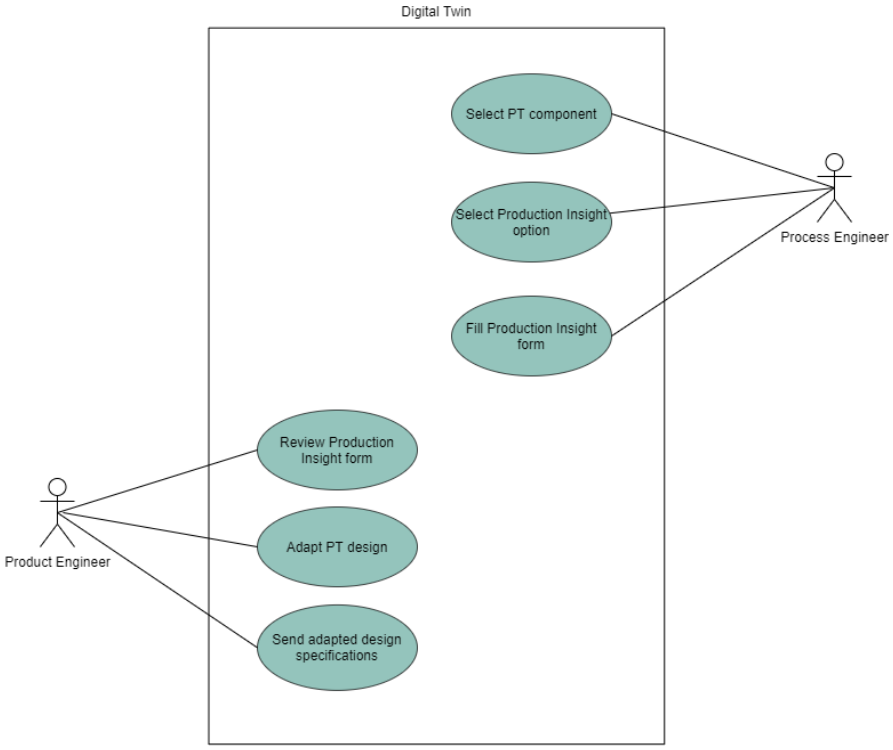


Figure 10- Use case diagram of the Digital Twin's production insight feature

5.2.3. Integrated chat

The collaboration between members of an engineering team is crucial for the success of a PT development project, as it guarantees that all members are aware of their role within the team and the status of the project. Furthermore, team interactions are an important instrument for knowledge sharing because they generally lead to the more experienced members naturally passing on their knowledge to others, over time.

Presently, Efacec relies on in-person interactions to accomplish these goals, both in formal meetings, and in informal conversations, in person or by email. The latter involves a period of time between responses, and as a result, is not an appropriate medium for fluid conversations. Moreover, it requires users to have the email address of the person they want to communicate with, and if the email is meant for multiple people, it may be unclear for the receivers, who exactly it is being sent to.

An instrument that can solve these problems is a chat feature that enables users to communicate in real time with other members of the engineering team, while PT development tasks are being performed. Through this feature, team members can follow the work that others are performing in real time and provide their insights regarding design features.

To enhance this feature, the chat feature should display the name, role, and team to which the user is assigned. Additionally, it should be possible to access the user's profile through the chat feature. This allows other users to actively associate what they are reading, to the skills and role of the person writing, facilitating collaboration during design tasks.

6. The Digital Twin Information architecture

The Digital Twin informational architecture is characterized by the information that it captures and generates, along with the purpose that each informational artifact serves, the methods and criteria by which they are organized, and the technologies that either create or are fed this information.

Given the applications and functionalities attributed to the Digital Twin in the previous chapter, this platform will capture and generate information with different sources, formats, and purposes, during the power transformer lifecycle. The following figure presents an overview of DT information capture and generation in all stages of PT lifecycle.

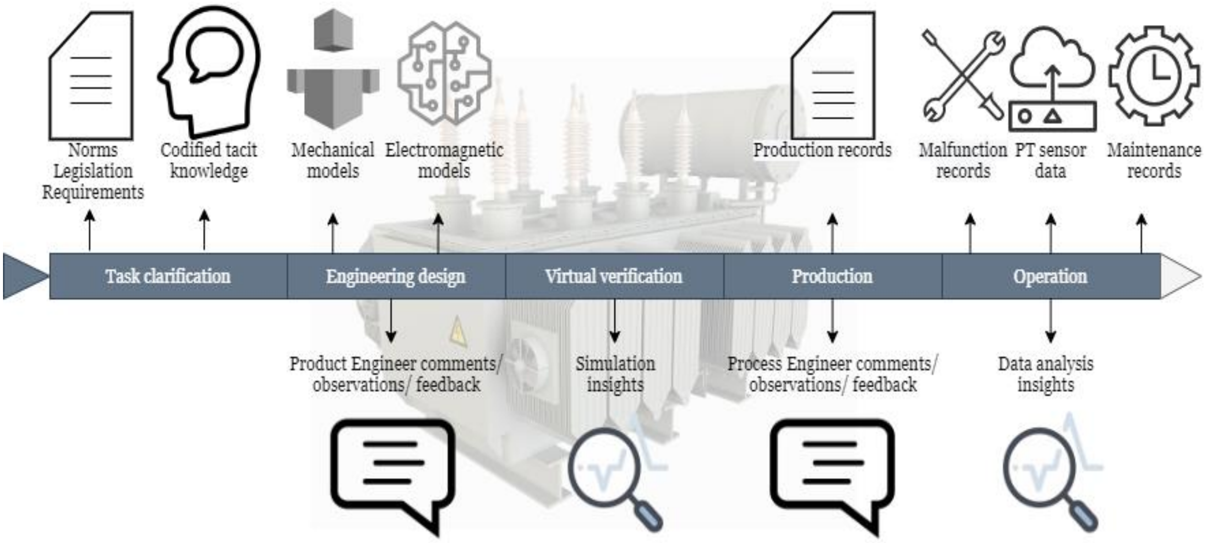


Figure 11- Overview of the Digital Twin information architecture

This information is stored and made available, not only as a way to enable PT functionalities to autonomously improve or correct non conformities in the design, but also to provide engineers with a searchable knowledge base that they can use to support and justify their decision making. Ultimately, this improves PT design, reduces lead time, and decreases costs.

Consequently, organizational knowledge becomes the foundation of the power transformer design process, thus achieving the primary goal of Knowledge-based Engineering.

6.1. The Digital Twin document corpus

6.1.1. Captured information

The information captured by the Digital Twin originates in the task clarification, engineering design, production, and operation stages of the PT lifecycle.

During task clarification, the DT captures information that restricts PT design, such as normative and legislative rules, as well as client requirements. Furthermore, tacit knowledge present in engineers' minds is also extracted and codified in the DT platform. This information is used to identify contradictions and non-conformities, and to optimize PT design.

In the design phase, models created in the engineering calculus and drawing tasks are captured by the DT, enabling the platform to accurately represent every property of the physical PT that will later be produced. Moreover, as the client is regularly sent design models to keep him informed of the progress of the project, his feedback is also stored in the DT.

Accordingly, during the design and production phases, the feedback of product and process engineers is also captured. Both documents contain feedback and comments regarding PT design, help formalize organizational knowledge, enabling engineers in future projects to understand design choices made in previous PT designs, and prevent making similar mistakes.

Finally, as PT operation begins, so does the capture of sensor data by the Digital Twin, which stores it through data warehouse technology and analyzes it with data analysis algorithms and artificial intelligence.

6.1.2. Generated information

Digital Twin generated information refers to all information that is autonomously created by DT functionalities and tools, instead of being captured in external sources, or created by stakeholder feedback.

This type of information is mostly generated by the data analysis and simulation features of the DT, and as such, it involves the virtual verification and operation lifecycle stages.

In the virtual verification stage, the DT performs Finite Element Analysis on the PT design, and this produces insights regarding the performance of each component when it is tested in several working conditions and parameters. Accessing this information during design tasks allows the engineers to support their decision-making regarding PT design.

Additionally, sensor data analysis through data analysis algorithms and artificial intelligence, also generates insights that enable the improvement of PT design or otherwise motivate changes made to it.

6.2. Defining a query mechanism for DT information

6.2.1. Enumerative and faceted classification schemes vs term-based query

As it has been established, the Digital Twin will function as a knowledge base that engineer's access to support their design tasks. As a result, it is crucial to implement a query mechanism that allows the fast and effective retrieval of the information stored in the platform. This involves describing, classifying, and organizing all captured and generated artifacts, as well as attributing metadata to them.

Traditionally, the retrieval of digital document-based information is done through term-based querying, a method that involves the return of a set of documents based on search terms input by users. Although this can be an effective way of retrieving information if engineers are familiarized with the terminology and contents of the document corpus, the complexity and longevity of power transformer engineering information makes this almost impossible, as the documents contain information related to multiple engineering branches and tasks. This is compounded by the lack of context offered by term-based querying, which generally relies on the user's understanding and interpretation of their informational need, to successfully retrieve a target document (Giess, Wild, and McMahon 2007).

This issue can be avoided with the implementation of a search mechanism that provides a predetermined structure in which documents are organized, as it enables the user to understand the different concepts and contexts that each document encompasses, allowing for a more effective information retrieval.

Two types of query mechanisms that permit information retrieval through a predetermined structure exist, namely the enumerative scheme, and the faceted classification scheme.

An enumerative scheme partitions concepts in a hierarchical arrangement until each describes a specific topic. A problem that arises with this type of scheme is the difficulty in selecting the top and lower level concepts, and the criteria that governs these choices has a deep impact on information retrieval, as it can steer a scheme towards users from a particular branch of engineering, compared to others.

Conversely, a faceted classification scheme involves the development of facets that describe a concept, and these facets can be combined to describe all the concepts that a specific document has information about. As a result, no viewpoint is prioritized by the query structure.

For this reason, a faceted classification approach is better suited to describe Digital Twin information.

6.2.2. Developing a Digital Twin information classification scheme

The development of a classification scheme involves identifying key concepts within a document that can be combined in order to describe the entirety of information contained in it, and then developing categories based on those concepts.

In the case of PT Digital Twin information, the development of the classification scheme was carried out by first describing the information generally contained in each type of engineering document, followed by developing categories that accurately describe the documents.

These tasks took into account the main goal of providing an easily understandable and efficient method for information retrieval by PT development stakeholders, regardless of field of expertise or role in the organization.

6.2.2.1. Describing Digital Twin information

To describe Digital Information, each type of document was envisioned regarding the concepts it contains, how it is organized, what format it is in, what task or process created it, and what purpose it serves. This resulted in the identification of six main types of documents, which are described in this section.

Information that restricts PT design: This type of documentation contains textual information regarding rules and restrictions of power transformer design, which can originate from norms, legislation, requirements, or tacit knowledge. These rules govern several design parameters related to PT components, such as dimensions, material and mass of the core, and oil levels in the oil tank, and as such, documents of this nature are generally organized by component, parameter, or functionality.

Information that specifies and formalizes PT design: This information originates in the engineering calculus and mechanical design tasks and includes all specifications regarding PT design. This includes 3D models, information about the materials, dimensions, shapes, and mass of each component, and electromagnetic models that specify voltage values and the behavior of magnetic forces in the machine.

Data captured by sensors: Datasets contained in .csv files. Each file contains sensor values and metadata related to them, such as the PT component that it was captured from, the parameter that was captured, and the time in which it was captured. Parameters include temperature, voltage, noise, and vibration.

Insights that support PT design improvements: Documents containing insights about PT design, production, or operation. PT design and production insights are created by product and process engineers, respectively, and consist of the comments and feedback posted on the DT platform regarding each PT component. Each of these documents contain information about the component in question, the type of problem that was encountered, and comments that explain the problem in further detail, and/or propose a solution. Moreover, operation insights are generated automatically by data analysis algorithms, which combine sensor data with design models, design rules and other historical information, to identify possible optimizations to operational parameters, predict failures, and define maintenance schedules. While the information in design and production insights is mostly in natural language, operation insights are generally composed of numerical values, graphs, and charts, and are usually organized by PT component, or PT functionality.

Information that guides engineers through the PT development process: This information contains instructions regarding product development tasks, and consequently, documents of this type are mostly textual information in natural language.

Administrative information: Documents that are created to support PT development, production, and operation, or are a result of the development process. This includes documents that support PT production, such as purchase, material allocation, and warehouse booking lists, as well as documents created by services done on the PT, such as maintenance and malfunction records.

6.2.2.2. Defining categories

The development of categories that provide a query structure for information retrieval, was supported by the identification of the concepts, formats, purposes, and origins of the Digital Twin document corpus.

The overall goal of this task was to create a robust classification scheme that can accurately describe any DT document, allowing for informational retrieval through a variety of query strategies. Each defined category is described in this section, which also includes the thought process and reasoning that led to its implementation in the scheme, as well as examples of possible query strategies.

- *Design rule*

To describe documents that contain information that restrict PT design, the *Design Rule* category was created. This allows engineers to retrieve the entirety of documents of this nature, and then refine their search by combining other categories related to a PT component or parameter, as needed.

Furthermore, as this category has four subcategories, namely *Normative Rule*, *Legislative Rule*, *Rule of Thumb* and *Requirement*, users can segment PT design restriction information according to its origin. Normative and legislative rules originate from norms or laws, requirements are agreed upon by the client and the organization based on client needs, and rules of thumb are elicited from the tacit knowledge of engineers.

As an example, a norm that details magnetic core dimensions according to target voltage values, can be retrieved by the following combination of categories:

NORMATIVE RULE – MAGNETIC CORE – DIMENSION
VOLTAGE

- *Design model*

Design artefacts created specifically during engineering design are crucial to design engineers, as this information accurately represents the mechanical, electric circuit, and magnetic circuit properties of the power transformer, and enables them to have an understanding of the designs created in previous PT development projects. Thus, it is crucial to define a category and set of sub-categories that allow users to retrieve design models relative to each main area of PT design. To achieve this, the *Design model* category was created, which is divided in the *Mechanical Model*, *Electric Circuit*, and *Magnetic Circuit* subcategories, respectively.

Consequently, a mechanical model of the primary winding, which also specifies dimensions, material, and mass of the component, could be described by:

MECHANICAL MODEL – PRIMARY WINDING – DIMENSION
MATERIAL
MASS

- *Insight*

The *Insight* category addresses the need for engineers to retrieve a set of documents that contain information that may lead to alterations and improvements in PT design or operation, thus effectively using captured knowledge to support their decision making regarding PT design.

These insights can be generated automatically, through Finite Element Analysis or DT sensor data analysis, or manually created, either by the design and production teams, based on their experiences while designing and building the machine, or by clients, by comparing project results with their needs and expectations. The *Human Insight* and *Automated Insight* subcategories distinguish insights generated by people, or automatically created by software or algorithms.

For instance, a document containing comments made by process engineers during production, regarding the welding in the oil tank, may be described by the following categories:

HUMAN INSIGHT – WELDING – OIL TANK – PRODUCTION

- *Instructional documentation*

The *Instructional Documentation* category is used to classify documents that contain information that help engineers determine what, how or when to do a certain task during design processes. Calculus manuals guide engineers through the calculus process, while project manuals encompass information related to the workflow of the transformer development process.

Supported by this category, engineers can quickly recover information that helps solve problems that are hindering PT development. An example of this is the classification of a document that details how to calculate the number of coils in the primary winding of the power transformer, depending on the voltage values required. In this case, employed categories could be:

CALCULUS MANUAL – WINDING – NUMBER OF COILS

- *Administrative documentation*

The Administrative Documentation category describes documents that support the PT development process. This includes documents detailing the purchase of materials or components, lists of materials to be allocated to every station on the shop floor, warehouse booking documents for different materials and components, transport plans for the complete PT, and type plates detailing PT specifications and production information. Accordingly, this category encompasses the subcategories *Advanced Purchase List*, *Material Allocation List*, *Transport Plan*, *Type Plate*, and *Warehouse Booking List*.

Given the nature of this type of documents, the *Administrative documentation* category will frequently be combined with the *Component* and *Parameter* categories, such as in the following example, which describes a purchase list containing the materials to be bought for the production of a PT oil tank:

ADVANCED PURCHASE LIST – STEEL – OIL TANK

IRON

- *Service Information*

The *Service Information* category describes documents that originate from tasks done on the PT during its operation, specifically maintenance tasks, and information created by PT operation itself, including sensor datasets and malfunction records. The thought process behind the creation of this category revolves around the need to clearly distinguish information that is created during PT development, from the information that is created during its operation.

Although this purpose is also fulfilled by the *Lifecycle Phase* category, since the information captured and generated during operation varies in type and origin, it was deemed best for it to be described by a set of subcategories that classify the tasks and activities that they result from.

As a result, information such as sensor data captured by the Digital Twin and stored in datasets, can be described by the following scheme:

SENSOR DATA – COMPONENT - PARAMETER

- *Component*

The *Component* category details the PT component that a given document has information about. To accommodate this classification scheme with Efacec practices, current PT component terminology was used to classify each component. For this reason, instead of using terminology such as *Primary Component* or *Secondary Component,* the defined subcategories were *Active Component*, which includes the components *Winding* and *Core*, and the *Accessory* subcategory, which lists secondary components such as *Oil Tank* and *Conservator*.

Given the extensive number of documents containing information about each component of the PT, a query using only these categories will return many results. Thus, this category will generally be combined with other categories that limit the scope of the search.

The following combination is an example of the classification of a document containing requirements for oil levels in the oil tank:

OIL TANK – OIL LEVEL - REQUIREMENT

- *Clamping element*

Similarly to the *Component* category, the *Clamping Element* category describes physical parts of the power transformer, although in this case, it refers to the devices or processes that support, divide, or connect PT components. Specifically, this category is used to classify documents that contain information about nails, nuts, reinforcements, washers, valves, or welding between components.

Supported by this category, a user is able to search for documents containing information about individual clamping elements, as well as to retrieve information about clamping elements located in the various components of the power transformer, through a combination of the *Clamping* element category, with the *Component* category. The combination below describes a document whose content is a mechanical model of the valves located in the oil tank:

VALVE – MECHANICAL MODEL – OIL TANK

- *Parameter*

The *Parameter* category is used to describe documents that detail properties of PT components. These can be dimensions, electric current voltages, magnetic energy intensity, mass, shape, type of material, noise, number of coils, oil levels, pressure, vibrations, or temperature. As such, a query involving the *Mass* subcategory would result in the retrieval of all documents that contain information about the mass of a given component, and the same applies to the other subcategories of this nature.

These subcategories will be associated with several types of documents, such as mechanical models, datasets, and simulation insights. An example is presented below, in the context of the results of a simulation test on the core temperature according to voltage levels:

TEMPERATURE – CORE
VOLTAGE

- *Lifecycle phase*

The *Lifecycle phase* category relates to the stage of power transformer development in which the document or data was produced or captured. These stages were defined according to the point of view of Tao et al. (2019), which proposed the partition of DT supported product development processes in the task clarification, engineering design, virtual verification, production, and operation stages (Tao et al. 2019).

The *Task Clarification* subcategory refers to the tasks related to the definition of requirements and other project aspects such as budget and deadlines. *Engineering Design* encompasses the engineering calculus and drawing tasks. *Virtual Verification* includes simulation and other testing done on PT design, and *Production* relates to the tasks of building the PT.

These categories allow users to have an overview of the entirety of information produced in each stage of PT development. This is especially important for design engineers before and during design tasks, as information is their main resource to guarantee a design that meets client and project expectations. The main benefit achieved through this classification scheme, is that an engineer can recover all information gathered before PT design begins, thus making it easier to satisfy informational needs. Because of this, a query of this type might simply consist of a single category, such as this:

TASK CLARIFICATION

- *Format*

The *Format* category describes the format in which a document or data is presented, with subcategories including *.csv*, *.pdf*, *.txt*, and *.xlsx*. Although it does not offer the user an indication of the content of the document, this category provides an additional path for informational retrieval. For instance, an engineer looking specifically for sensor data, regardless of component or parameter, may simply select the *.csv* category. Since sensor data is contained in .csv datasets, this would allow him to retrieve all datasets related to a specific power transformer.

Conversely, this category could also be combined in a query with *Component* and *Parameter*, which would result in the return of the same information as using the *Sensor Data* category, with the possible addition of other unrelated documents in this format. The category combination below exemplifies this, in the case of a core temperature dataset:

.CSV – CORE – TEMPERATURE

6.3. Final considerations

The Digital Twin virtually and accurately replicates all aspects of the physical power transformer, whether it is its geometric, mechanical, or electromagnetic properties and behaviors. The main driving force behind the link between physical and virtual space, is information, because it enables the accurate description of all its properties, and the employment of tools to simulate its actual expected and unexpected behaviors, and to identify nonconformities and optimization opportunities, entirely on the virtual realm.

In addition to this, the DT lifecycle not only matches the lifecycle of its physical counterpart, but also predates and supersedes it, as it is created before production, and continues to exist after disposal. Consequently, it can store all information that is captured or created during the development, production, and operation of the PT.

As such, the DT is a digital informational construct, which encapsulates information about every task, communication, operation, or action done to the physical power transformer, relative to it, or originating from it.

Considering this, an overview of the Digital Twin information architecture is envisioned based on the physical characteristics of the PT, the information that originates from development processes, the data captured in the physical space, and the information generated by DT features.

Furthermore, an information classification scheme that structures the DT information architecture, is proposed. This scheme enables users to query and retrieve information that empowers them to make informed decisions during PT development. This scheme should be further developed to ensure that it is faceted in nature, as this permits the creation of a predetermined query structure that allows for various points of view and information retrieval strategies, to access a certain document.

7. Conclusions and future work

The work carried out in this dissertation resulted in the development of a Digital Twin concept for power transformer development. The DT concept and functionalities were developed based on scientific and professional literature, along with the analysis of PT development processes at Efacec, a Portuguese company of the energy sector.

The overall goal of this dissertation was to define DT functionalities that lessened the impact that bottlenecks in current Efacec processes have, specifically those that relate to information, knowledge, and collaboration management. As a result, a crucial objective was to develop DT supported PT design processes, which are driven by organizational knowledge to automate, optimize, and streamline PT design tasks.

Among the bottlenecks identified in current Efacec processes was the inability to capture PT operation data, and a knowledge management strategy that relies solely on interactions between collaborators for knowledge sharing, and thus lacks mechanisms for knowledge formalization.

The proposed DT concept encompasses features that support PT design, testing and operation, by employing four different sources of organizational knowledge, namely the design artefacts captured in previous PT development projects, the insights generated by the capture and analysis of power transformer sensor data, testing data created by the DT simulation functionality, and finally, the formalized tacit knowledge of Efacec product and process engineers.

During PT development, the DT can provide engineers with design templates based on engineering calculus specifications and previous PT designs, as well as detect design nonconformities through the coding of project requirements on the platform. Furthermore, this technology can analyze sensor and simulation data through data analysis algorithms, to iteratively identify and recommend optimization opportunities in PT design.

Throughout operation, the DT enables real time monitoring of PT performance, which allows users to take action and send appropriate commands to the machine in case of an unexpected event or behavior. Crucially, the DT also employs sensor data analysis in this lifecycle stage, specifically to define optimized maintenance schedules and predict possible component failures.

These functionalities help engineers reach an optimized design faster and more efficiently, which yield benefits such as reduction of project lead time and monetary costs. Additionally, they assure the correct and optimized performance of the power transformer, during operation. Thus, some of these benefits result from the DT taking advantage of PT sensor data, an asset that has yet to be exploited by Efacec.

Other DT features focus on knowledge and collaboration management, with the goal of capturing, storing, and making available all PT lifecycle information, along with providing KM instruments that motivate knowledge formalization and sharing.

These goals are accomplished by the DT's interactive comment, integrated chat, and information repository functionalities.

The interactive comment feature allows users to select a given PT design attribute and create a record that describes the thought process behind the design decision making. This allows engineers to learn from previous designs and use the acquired knowledge to support their decision making in their current projects. Because of this, the DT becomes an important knowledge sharing instrument.

Moreover, the DT also provides an integrated chat feature that permits constant interaction between elements of an engineering team, or alternatively, between different engineering teams. As a collaboration instrument, this feature is bound to stimulate knowledge sharing, as teammates can share opinions and insights while design or testing is being performed, in a single platform.

Finally, as the Digital Twin captures and generates a large amount of product lifecycle information that either feeds its functionalities, or is created by them, it is appropriate for the platform to provide an information repository that makes the entirety of this information available to engineers, allowing them to perform queries and satisfy their information needs.

As a follow-up to this feature, the information architecture of the DT was envisioned, and this included defining a classification scheme for DT information.

Three different query mechanisms were considered, namely term-based query, enumerative classification schemes, and faceted classification schemes. Because of the interdisciplinarity of PT development, which involves both mechanical and electromagnetic engineering, and the variety of sources, formats, and purposes of PT lifecycle information, a query mechanism that provides redundancy was favored, as it allows the retrieval of a document through various query strategies and points of view.

Consequently, this dissertation proposes a classification scheme for the DT information architecture, which includes a range of categories that can be combined to describe the entirety of information in a given document or dataset.

Together with the proposed DT functionalities, this information classification scheme forms a conceptual and theoretical base for the future implementation of the Digital Twin in power transformer development, yet, some more work is needed in this area before a successful practical implementation can take place.

Specifically for the classification scheme, further work is required to ensure that it is able to accurately describe every document, and that it provides an efficient query mechanism for information retrieval during PT development. This entails examining every document produced throughout the PT lifecycle, identifying concepts that describe the information contained therein, and comparing them to the categories developed during this dissertation. Ultimately, this test should verify that the scheme is able to describe every identified concept, or otherwise motivate the development of additional facets.

Furthermore, Efacec stakeholders should have a role in defining additional DT functionalities, as well as in adapting those that were proposed in this dissertation. As such, requirement elicitation supported by techniques such as interviews, questionnaires, or observation, should be pursued in order to obtain a complete understanding of the DT concept, when applied to PT development.

From the perspective of information and knowledge management, the implementation of the DT will involve close collaboration with software developers, with the goal of coding relevant organizational knowledge in the DT platform. This includes identifying and assembling every design rule that is present in norms and legislation, and communicating this information to programmers, who will then code it in software to enable DT functionalities such as the nonconformity identification feature. Additionally, the extraction of tacit knowledge present in engineers' minds should be attempted, and this can be guided by methodologies like CommonKADS or MOKA. As a result, this knowledge, which will generally be presented as design rules of thumb, could also be coded in the DT platform.

In addition to these tasks, IM stakeholders will play a critical role in linking upper management stakeholders with software developers during the DT implementation, as their skills include proficiency in both the technological and human realms. As a result, upper management stakeholders, who generally do not have technological competencies, can be properly updated on the status of the DT implementation project, as well as have a good understanding of the purpose, characteristics, and scope of the DT application.

Finally, following the conclusion of the DT implementation, substantial training efforts should be undertaken, as this technology will bring significant change to current PT development processes.

8. References

Aamodt, A, and E Plaza. 1996. “Case-Based Reasoning: Foundational Issues, Method Ological Variations, and System Approaches.” *Artificial Intelligence Communications* 7 (1): 39–59. https://doi.org/10.3390/s120811154.

Ammar-Khodja, Samar, Nicolas Perry, and Alain Bernard. 2008. “Processing Knowledge to Support Knowledge-Based Engineering Systems Specification.” *Concurrent Engineering Research and Applications* 16 (1): 89–101. https://doi.org/10.1177/1063293X07084642.

Baumeister, Joachim, Jochen Reutelshoefer, and Frank Puppe. 2011. “Know We: A Semantic Wiki for Knowledge Engineering.” *Applied Intelligence* 35 (3): 323–44. https://doi.org/10.1007/s10489-010-0224-5.

Bermell-García, Pablo, and Ip-shing Fan. 2002. “A Kbe System for the Design of Wind Tunnel Models Using Reusable Knowledge Components.” *Proceedings of the VI International\nConference on Project Engineering*, 594–604.

Boschert, Stefan, and Roland Rosen. 2016. “Digital Twin - The Simulation Aspect.” In *Mechatronic Futures: Challenges and Solutions for Mechatronic Systems and Their Designers*, 59–74. https://doi.org/10.1007/978-3-319-32156-1_1.

Calkins, Dale, Nathaniel Egging, and Christian Scholz. 2000. “Knowledge-Based Engineering (KBE) Design Methodology at the Undergraduate and Graduate Levels” 16 (1).

Canedo, Arquimedes. 2016. “Industrial IoT Lifecycle via Digital Twins.” *2016 International Conference on Hardware/Software Codesign and System Synthesis (CODES+ISSS)*, 1.

Chapman, C. B., and M. Pinfold. 1999. “Design Engineering - a Need to Rethink the Solution Using Knowledge Based Engineering.” *Knowledge-Based Systems* 12 (5–6): 257–67. https://doi.org/10.1016/S0950-7051(99)00013-1.

Cooper, Stephen, Ip-shing Fan, and Guihua Li. 2001. “Achieving Competitive Advantage Through Knowledge-Based Engineering A Best Practice Guide,” 21.

Curran, Richard, Wim J.C. Verhagen, Michel J.L. Van Tooren, and Ton H. Van Der Laan. 2010. “A Multidisciplinary Implementation Methodology for Knowledge Based Engineering: KNOMAD.” *Expert Systems with Applications* 37 (11): 7336–50. https://doi.org/10.1016/j.eswa.2010.04.027.

“Digital Twins in Logistics.” 2019. DHL. https://www.logistics.dhl/content/dam/dhl/global/core/documents/pdf/glo-core-digital-twins-in-logistics.pdf.

Elst, S. W.G. Van Der, and M. J.L. Van Tooren. 2008. “Application of a Knowledge Engineering Process to Support Engineering Design Application Development.” In *Collaborative Product and Service Life Cycle Management for a Sustainable World*, 417–31. Springer, London. https://doi.org/10.1007/978-1-84800-972-1.

Emberey, C. L., N. R. Milton, J. P.T.J. Berends, M. J.L. Van Tooren, S. W.G. Van Der Elst, and B. Vermeulen. 2007. “Application of Knowledge Engineering Methodologies to Support Engineering Design Application Development in Aerospace.” *Collection of Technical Papers - 7th AIAA Aviation Technology, Integration, and Operations Conference* 1 (September): 83–95. https://doi.org/10.2514/6.2007-7708.

“For a Digital Twin of the Grid.” 2017. https://assets.new.siemens.com/siemens/assets/api/uuid:09c20834-4ed4-49d8-923d-ebcc541cab37/inno2017-digitaltwin-e.pdf.

Gabor, Thomas, Marie Kiermeier, Michael Till Beck, and Alexander Neitz. 2016. “A Simulation-Based Architecture for Smart Cyber-Physical Systems.” *2016 IEEE International Conference on Autonomic Computing (ICAC)*, 374–79. https://doi.org/10.1109/ICAC.2016.29.

Georgilakis, Pavlos S. 2009. *Spotlight on Modern Transformer Design*. *Power Systems*. Vol. 38. https://doi.org/10.1007/978-1-84882-667-0.

Giess, M. D., P. J. Wild, and C. A. McMahon. 2007. “The Use of Faceted Classification in the Organisation of Engineering Design Documents.” *Proceedings of ICED 2007, the 16th International Conference on Engineering Design* DS 42 (January 2007).

Girard, John, and JoAnn Girard. 2015. “Defining Knowledge Management : Toward an Applied Compendium.” *Online Journal of Applied Knowledge Management* 3 (1): 1–20.

Glaessgen, E.H., and D.S. Stargel. 2012. “The Digital Twin Paradigm for Future NASA and U . S . Air Force Vehicles.” In , 1–14.

Grieves, Michael, and John Vickers. 2016. “Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems.” *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, 2016. https://doi.org/10.1007/978-3-319-38756-7.

Grieves, Michael W. 2019. “Virtually Intelligent Product Systems: Digital and Physical Twins.” In *Complex Systems Engineering: Theory and Practice*, 175–200. American

Institute of Aeronautics and Astronautics, Inc. https://doi.org/10.2514/5.9781624105654.0175.0200.

Haag, Sebastian, and Reiner Anderl. 2018. “Digital Twin – Proof of Concept.” *Manufacturing Letters* 15: 64–66. https://doi.org/10.1016/j.mfglet.2018.02.006.

Hansen, M. T., N. Nohria, and T. Tierney. 1999. “What’s Your Strategy for Managing Knowledge?” *Harvard Business Review* 77 (2).

“Improving Wind Power with Digital Twin Turbines.” n.d. Accessed November 29, 2019. https://www.ge.com/renewableenergy/stories/improving-wind-power-with-digital-twin-turbines.

“Knowledge Based Engineering (KBE) – CADVision Engineers.” 2019. 2019. http://cadvisionengineers.com/services/kbe-knowledge-based-engineering/.

Lin, Bor Tsuen, Chian Kun Chan, and Jung Ching Wang. 2008. “A Knowledge-Based Parametric Design System for Drawing Dies.” *International Journal of Advanced Manufacturing Technology* 36 (7–8): 671–80. https://doi.org/10.1007/s00170-006-0882-y.

Moreno, Aitor, Gorka Velez, Aitor Ardanza, Iñigo Barandiaran, Álvaro Ruíz, and De Infante Raúl. 2017. “Virtualisation Process of a Sheet Metal Punching Machine within the Industry 4 . 0 Vision.” *International Journal on Interactive Design and Manufacturing (IJIDeM)* 11 (2): 365–73. https://doi.org/10.1007/s12008-016-0319-2.

Narayan, K. Lalit, K. Mallikarjuna Rao, and M.M.M Sarcar. 2008. *Computer Aided Design and Manufacturing*. PHI.

Oldham, Keith, Stephen Kneebone, Martine Callot, Adrian Murton, and Richard Brimble. 1998. “MOKA-A Methodology and Tools Oriented to Knowledge-Based Engineering Applications.” *Proceedings of the Conference on Integration in Manufacturing. Göteborg, Sweden*, 198–207.

Reddy, E Jayakiran, C N V Sridhar, and V Pandu Rangadu. 2015. “Knowledge Based Engineering: Notion, Approaches and Future Trends.” *American Journal of Intelligent Systems* 5 (1): 1–17. https://doi.org/10.5923/j.ajis.20150501.01.

Riege, Andreas. 2005. “Three-Dozen Knowledge-Sharing Barriers Managers Must Consider.” *Journal of Knowledge Management* 9 (3): 18–35. https://doi.org/10.1108/13673270510602746.

Riesbeck, Cristopher, and Roger Schank. 1989. *Inside Case-Based Reasoning*.

Rosen, Roland, Georg Von Wichert, George Lo, Kurt D Bettenhausen, Roland Rosen, Georg Von Wichert, Kurt D Bettenhausen, Roland Rosen, and D Germany. 2015. “About The Importance of Autonomy and Digital Twins for the Future of Manufacturing.” *IFAC-PapersOnLine* 48 (3): 567–72. https://doi.org/10.1016/j.ifacol.2015.06.141.

Sandberg, Marcus. 2003. “Knowledge Based Engineering-In Product Development.”

Santos, Vitor Ricardo, António Lucas Soares, and João Álvaro Carvalho. 2012. “Knowledge Sharing Barriers in Complex Research and Development Projects: An Exploratory Study on the Perceptions of Project Managers.” *Knowledge and Process Management* 19 (1): 27–38. https://doi.org/10.1002/kpm.1379.

Schleich, Benjamin, Nabil Anwer, Luc Mathieu, and Sandro Wartzack. 2017. “CIRP Annals - Manufacturing Technology Shaping the Digital Twin for Design and Production Engineering.” *CIRP Annals - Manufacturing Technology* 66 (1): 141–44. https://doi.org/10.1016/j.cirp.2017.04.040.

Schreiber, Guus, Hans Akkermans, Anjo Anjewierden, Robert De Hoog, Nigel Shadbolt, Walter Van De Velde, and Bob Wielinga. 2000. *The CommonKADS Methodology*.

Schroeder, Greyce, Charles Steinmetz, Carlos Eduardo Pereira, Ivan Muller, Natanael Garcia, Danubia Espindola, and Ricardo Rodrigues. 2016. “Visualising the Digital Twin Using Web Services and Augmented Reality.” *2016 IEEE 14th International Conference on Industrial Informatics (INDIN)*, 522–27. https://doi.org/10.1109/INDIN.2016.7819217.

Schroeder, Greyce, Charles Steinmetz, Carlos Pereira, and Danubia Espindola. 2016. “Digital Twin Data Modeling with AutomationML and a Communication Methodology for Data Exchange.” *IFAC-PapersOnLine* 49 (30): 12–17. https://doi.org/10.1016/j.ifacol.2016.11.115.

Shafto, Mike, Chris Kemp, Mike Conroy, Rich Doyle, Ed Glaessgen, Jacqueline LeMoigne, and Lui Wang. 2010. “DRAFT Modeling , Simulation , Information Technology & Processing Roadmap.” https://www.nasa.gov/pdf/501321main_TA11-MSITP-DRAFT-Nov2010-A1.pdf.

Shankar, Ravi, Singh, Amol Gupta, and Rakesh Narain. 2003. “Strategic Planning for Knowledge Management Implementation in Engineering Firms.” *Work Study* 52 (4): 190–200. https://doi.org/10.1108/00438020310479036.

Sobieszczanski-Sobieski, Jaroslaw, Alan Morris, and Michel J.L. van Tooren. 2015. *Multidisciplinary Design Optimization Supported by Knowledge Based Engineering*.

Multidisciplinary Design Optimization Supported by Knowledge Based Engineering. wiley. https://doi.org/10.1002/9781118897072.

Tao, Fei, Jiangfeng Cheng, Qinglin Qi, Meng Zhang, He Zhang, and Fangyuan Sui. 2018. “Digital Twin-Driven Product Design, Manufacturing and Service with Big Data.” *International Journal of Advanced Manufacturing Technology* 94 (9–12): 3563–76. https://doi.org/10.1007/s00170-017-0233-1.

Tao, Fei, Fangyuan Sui, Ang Liu, Qinglin Qi, Meng Zhang, Boyang Song, Stephen C Lu, and A Y C Nee. 2019. “Digital Twin-Driven Product Design Framework.” *International Journal of Production Research* 7543: 1–19. https://doi.org/10.1080/00207543.2018.1443229.

Tuegel, Eric. 2012. “The Airframe Digital Twin: Some Challenges to Realization.” In *53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference*. https://doi.org/10.2514/6.2012-1812.

Uhlemann, Thomas H, Christian Lehmann, and Rolf Steinhilper. 2017. “The Digital Twin : Realizing the Cyber-Physical Production System For.” *Procedia CIRP* 61: 335–40. https://doi.org/10.1016/j.procir.2016.11.152.

Ulrich, Karl, and Steven Eppinger. 2012. *Product Design and Development*. 5th ed. McGgraw-Hill.

Valentim, Celso Ricardo Salazar. 2019. “Engenharia Do Conhecimento e a Metodologia CommonKADS.” HUMANTECH. https://humantech.ws/humantech/ebooks/ebook-Engenharia-do-Conhecimento.pdf.

Weyer, Stephan, Torben Meyer, Moritz Ohmer, Stephan Dominic, Torben Meyer, Moritz Ohmer, Stephan Dominic, Torben Meyer, Moritz Ohmer, and Dominic Gorecky. 2016. “Future Modeling and Simulation of CPS-Based Factories : An Example from the Automotive Industry.” *IFAC PapersOnLine* 49 (31): 97–102. https://doi.org/10.1016/j.ifacol.2016.12.168.

Zack, Michael. 1999. “Developing a Knowledge Strategy.” *California Management Review* 41 (3): 125–45.

Zhang, Hao, Qiang Liu, Xin Chen, Ding Zhang, and Jiewu Leng. 2017. “A Digital Twin-Based Approach for Designing and Multi-Objective Optimization of Hollow Glass Production Line.” *IEEE Access* 5: 26901–11. https://doi.org/10.1109/ACCESS.2017.2766453.

Annexes

Annex 1– Interview scripts

Interview 1 - 02/07/2020

Location: Efacec
Start Time: 10 am
Interviewee role: Transformers Business Unit Principal Technology Expert
Interview Objectives:

- Obtain an overview of how the power transformer development processes are carried out.
- Realize what information sources are consulted, what kind of information is generated, how it is made available.
- Understand when team members collaborate, and through what channels they communicate.
- Identify the moments in which the tacit knowledge of engineers is crucial and how their sharing is encouraged.

Questions:

General / organization / collaboration

1. What is the methodology that governs the management of PT development projects?
2. How is a PT development project initiated?
3. Through which channel is communication with the customer done?
4. When is communication with the customer made?
5. What is the layout of PT development teams?
6. What aspects of development is each team responsible for?
7. Are DT activities carried out simultaneously by the different teams?
8. How are the activities of the different DT teams coordinated?
9. When do DT teams communicate?
10. By which channels do the different teams communicate?
11. How is communication / coordination done with the production teams?

Information Management

12. What sources of information / databases do engineers consult during the development of a transformer?
13. In what formats is this information found?
14. What tools / software do teams use?

15. What design artifacts / documents are produced in each project?
16. In what formats are the artifacts / documents in?
17. Where are design artifacts stored?
18. How are design artefacts organized?
19. What metadata is assigned to the artifacts / documents?
20. How are the artifacts made available?
21. How long are design artefacts stored?
22. What documents / information originate from design validation activities?

Knowledge Management

23**.** How is knowledge sharing encouraged between teams?
24. In what aspects of PT development is the tacit knowledge and experience of engineers most important?

Interview 2 – 19/03/2020

Location: Remote (Skype)

Start Time: 11:30 am

Interviewee role: Power Transformer Lead Engineer

Interview Objectives:

- Understand the layout of engineering teams/cells.
- Identify what information is accessed by engineers during design.
- Understand which tools / mechanisms are used to manage knowledge and organizational learning.
- Identify other relevant stakeholders in the development of transformers.

Questions:

Personal questions

1. Academic background?
2. Professional background?
3. What is your role at Efacec?
4. How long have you held these responsibilities?

Team structure

5. How many members compose each engineering cell?
6. What responsibility does each member have?
7. Is the cell static or do elements rotate?
8. What are the criteria for the selection and distribution of members to the engineering teams?
9. Is each element assigned a special role, responsibility or ranking according to their skills, years in the company, interests, etc? For example, someone experienced with the responsibility to guide someone inexperienced during a given project?

Project start/ validation

10. By what means do you receive the project dossier?
11. What information does this document contain?
12. What information is most often incomplete or wrong?
13. What is the purpose of the handover meeting?
14. If project specifications are invalid or are incomplete, do you report these deficiencies to the project management department?
15. Which stakeholders are responsible for making the necessary project specification changes?

16. How is the project presented to the engineering team?

17. What information is made available to the engineering team when starting a new transformer development project?

Power transformer Development

18. What studies are carried out before engineering calculus?

19. What information is generated from these studies?

20. How do the results of these studies affect the final design of the transformer?

21. What does the engineering calculus consist of?

22. What are the criteria for selecting the most appropriate calculation method?

23. What information is generated from engineering calculus?

24. What is the purpose of internal Design Reviews?

25. How do External Design Reviews differ from Internal Design Reviews?

26. What information is generated during PT mechanical design?

27. Where is it stored?

28. How is it made available?

29. By what means is this dossier transmitted to production?

Collaboration Management / Information Management in General

30. How does customer feedback impact the development of the power transformer?

31. How often do errors in the development of the transformer (design errors, missed deadlines, errors in model generation, failures in parameter input), occur?

32. Are there other methods (lessons learned, others) that aim to identify / capture acquired? Organizational knowledge?

33. Is any of this knowledge encoded in software?

34. Is any of the acquired knowledge recorded in documents such as manuals, etc.?

34. Are there meetings/moments where employees are motivated to express possible innovative ideas, suggestions, or improvements?

Interview 3 – 07/04/2020

Location: Remote (Skype)

Start Time: 5 pm

Interviewee role: Engineering team coordinator

Interview Objectives:

- Understand what knowledge management instruments are employed during power transformer development.
- Understand what information is accessed to support design decision making.

Knowledge management

1. What meetings are held to ensure team members are aware of the status of the project and the tasks they must perform?
2. How are mistakes made during PT development identified?
3. What KM instruments are employed to learn from these mistakes?

Power transformer Development

4. Which tasks are done manually, and which are done autonomously?
5. What software is used for engineering calculus?
6. What software is used for mechanical design?
7. What software is used for design validation?
8. During PT development, what information sources do engineers access?
9. Do engineers access supporting documentation such as manuals?
10. What format is this information in?
11. Is there information that is only in paper format?
12. How is design validated? What kind of tests/simulations are performed?
13. What information is provided to the client during PT development?

Annex 2- Digital Twin requirements

FUNCTIONAL REQUIREMENTS

1. **The Digital Twin should be able to virtually replicate the physical power transformer**

    1.1 The Digital Twin should be able to virtually replicate the geometric properties of the power transformer

    1.2. The Digital Twin should be able to virtually replicate the mechanical properties of the power transformer

    1.3. The Digital Twin should be able to virtually replicate the electromagnetic properties of the power transformer

2. **The Digital Twin should be able to communicate with the physical transformer through a data transmission channel**

    2.1. The Digital Twin should be able to receive sensor data originating from the physical power transformer

    2.2. The Digital Twin should be able to send commands to the power transformer

3. **The Digital Twin should be able to be accessed through multiple devices**.

    3.1. The Digital Twin should be able to be accessed through a computer

    3.2. The Digital Twin should be able to be accessed through portable electronic devices

    3.3. The Digital Twin should be able to be visualized through virtual reality

4. **The Digital Twin should provide an information repository**

    4.1. The Digital Twin information repository should provide a query mechanism

    4.2. The Digital Twin information repository should be able to store power transformer design artefacts

    4.3. The Digital Twin should be able to store Big Data

5. **The Digital Twin should be able to identify similar power transformer designs based on engineering calculus parameters**

    5.1. The Digital Twin should list similar power transformer designs based on engineering calculus parameters

    5.2. The Digital Twin should be able to generate a percentage of similarity between power transformer designs

6. The Digital Twin should be able to identify nonconformities in power transformer design based on project requirements, rules and restrictions

6.1. Requirements, rules, and restrictions should be able to be coded in the Digital Twin

6.2. The Digital Twin should notify the user when a nonconformity is detected

6.3. The Digital Twin should be able to autonomously correct design nonconformities

7. The Digital Twin should be able to perform simulations

7.1. The Digital Twin should be able to perform Finite Element Analysis on the power transformer design

7.2. The Digital Twin should be able to recommend changes to the power transformer design based on Finite Element Analysis results

7.3. The Digital Twin should be able to generate a report containing the results of the Finite Element Analysis

7.4. The Digital Twin should be able to store Finite Element Analysis reports

8. The Digital Twin should be able to analyze sensor data with data analysis algorithms

8.1. The Digital Twin should be able to recommend optimizations to power transformer design based on sensor data analysis

8.2. The Digital Twin should be able to predict power transformer failures based on sensor data analysis

8.3. The Digital Twin should be able to predict optimized maintenance schedules based on sensor data analysis

9. The Digital Twin should be able to monitor the physical transformer's condition in real time

SECURITY REQUIREMENTS

1. Each user of the Digital Twin platform should have a profile

2. Each profile should have a login username and password

3. Access to the Digital Twin platform should require the user to login

4. The Digital Twin should allow the assignment of different permissions to each user

5. The Digital Twin should allow view only permissions

6. The Digital Twin should allow view and comment permissions

7. The Digital Twin should allow view and modify permissions

8. View and comment permissions should be allocated to process engineers and members of the engineering department

9. View and modify permissions should only be allocated to members of an engineering team

10. The communication between the Digital Twin and the Power Transformer should be encrypted

PERFORMANCE REQUIREMENTS

1. The Digital Twin should be accessible 24 hours per day

2. The Digital Twin should be accessible 365 days per year

3. The Digital Twin must take a maximum of 1 second to execute a command

Annex 3- Digital Twin information classification scheme

CATEGORY	SUBCATEGORY	DESCRIPTION
DESIGN RULE	NORMATIVE RULE LEGISLATIVE RULE RULE OF THUMB REQUIREMENT	The *Design Rule* category describes documents that contain restrictive information regarding power transformer design. It has four subcategories, namely *Normative Rule*, *Legislative Rule*, *Rule of Thumb* and *Requirement*. Normative and legislative rules originate from norms or laws, requirements are agreed upon by the client and the organization based on client needs, and rules of thumb are elicited from the tacit knowledge of engineers.
DESIGN MODEL	ELECTRIC CIRCUIT MODEL MAGNETIC CIRCUIT MODEL MECHANICAL MODEL	The *Design Model* category describes models created during engineering design. The models accurately represent the mechanical, electric circuit, and magnetic circuit properties of the power transformer. These aspects are described by the *Mechanical Model*, *Electric Circuit*, and *Magnetic Circuit* subcategories, respectively.
INSIGHT	HUMAN INSIGHT AUTOMATED INSIGHT	The *Insight* category describes documents that contain information that may improve or otherwise modify PT design. These insights can be generated automatically, by Finite Element Analysis or DT sensor data analysis, or manually created, either by the design and production team, based on their experiences while designing and building the machine, or by clients, by comparing the design with their needs and expectations for the product. The *Human Insight* and *Automated Insight* subcategories distinguish insights generated by people, or automatically created by software or algorithms.
INSTRUCTIONAL DOCUMENTATION	CALCULUS MANUAL PROJECT MANUAL	*Instructional Documentation* is used to classify documents that contain information that help engineers determine what, how or when to do a certain task during design processes. Calculus manuals guide

	SOFTWARE MANUAL	engineers through the calculus process, while project manuals encompass information related to the workflow of the transformer development process. Finally, software manuals detail how to use software related to PT design, such as CAD applications.
ADMINISTRATIVE DOCUMENTATION	ADVANCED PURCHASE LIST	The Administrative Documentation category describes documents that support the PT development process. This includes documents detailing the purchase of materials or components, lists of materials to be allocated to every station on the shop floor, warehouse booking documents for different materials and components, transport plans for the complete PT, and type plates detailing PT specifications and production information. Accordingly, this category encompasses the subcategories *Advanced Purchase List*, *Material Allocation List*, *Transport Plan*, *Type Plate*, and *Warehouse Booking List*.
	MATERIAL ALLOCATION LIST	
	TRANSPORT PLAN	
	TYPE PLATE	
	WAREHOUSE BOOKING LIST	
SERVICE INFORMATION	MAINTENANCE RECORD	The *Service Record* category describes information that originates from tasks done on the PT during its operation, or are created by PT operation itself. These can be maintenance or malfunctions records, and sensor data datasets.
	MALFUNCTION RECORD	
	SENSOR DATA	
COMPONENT	ACTIVE COMPONENT	The *Component* category details the PT component that a given document has information about. The *Active Component* subcategory includes the subcategories *Winding* and *Core*, while the Accessory subfacet lists secondary components such as *Oil Tank* and *Conservator*.
	ACCESSORY	
CLAMPING ELEMENT	NAIL	In this classification scheme, *Clamping Element* refers to the devices or processes that support, divide, or connect PT components. Thus, this category describes documents that contain information about nails, nuts, reinforcements, washers, weldings or valves.
	NUT	
	REINFORCEMENT	
	VALVE	

	WASHER	
	WELDING	
PARAMETER	DIMENSION	The *Parameter* category is used to describe documents that detail properties of PT components. These can be dimensions, electric current voltages, magnetic energy intensity, mass, shape, type of material, volume of noise, pressure, number of coils, oil levels in the oil tank, vibrations, or temperature.
	VOLTAGE	
	MAGNETIC ENERGY	
	MASS	
	MATERIAL	
	NOISE	
	NUMBER OF COILS	
	OIL LEVEL	
	PRESSURE	
	TEMPERATURE	
	VIBRATION	
	SHAPE	
LIFECYCLE PHASE	TASK CLARIFICATION	The *Lifecycle phase* category relates to the stage of power transformer development in which the document or data was produced or captured. The *Task Clarification* subcategory refers to the tasks related to the definition of requirements and other project aspects such as budget and deadlines. *Engineering Design* encompasses the engineering calculus and drawing tasks. Virtual Verification includes simulation and
	ENGINEERING DESIGN	
	VIRTUAL VERIFICATION	

	PRODUCTION	other testing done on PT design. Finally, Production relates to the tasks of building the PT.
	OPERATION	
FORMAT	.CSV	The *Format* category describes the format in which the document or data is presented. Subfacets include *.csv*, *.pdf*, *.txt*, and *.xlsx*.
	.PDF	
	.TXT	
	.XLSX	