Understanding and fulfilling information needs of stakeholders along product lifecycles- Applying data analytics in product life-cycle management

Vom Fachbereich Produktionstechnik

der

UNIVERSITÄT BREMEN

zur Erlangung des Grades

Doktor-Ingenieur

genehmigte

Dissertation

von

M.Sc. Elaheh Gholamzadeh Nabati

Hauptreferent: Prof. Dr.-Ing. habil. Klaus-Dieter Thoben

Korreferent: Prof. Dr. Otthein Herzog

Tag der mündlichen Prüfung: 20.12.2018

To my parents Parvin and Mohsen

Abstract

Improving the management of product lifecycle is becoming increasingly important today. Companies need to do the product lifecycle operations more effectively, share product information and streamline the flow of information between different organizations of the product value chain to stay in a competitive market. At the same time, they need to innovate and design new products or services. Reaching these goals is challenging for most organizations, which are involved in the product value chain.

By considering this, companies are trying to use digital information stored in organizational databases. At the same time, they are trying to access new sources of information. An essential part of the information they want to access concerns the product and its users. In this context, the concept of Closed Loop Product Lifecycle Management (CL-PLM) has been developed to help organizations in the value chain to promote the comprehensive management of product lifecycle information and activities. CL-PLM attempts to manage all processes related to products, from product design to end of product life. However, the development of this concept is not still complete. There is a need for more research to manage the data and information flows and to provide each organization in the value chain with the right data from product Middle Of the Life (MOL).

In this thesis, a new concept is presented, which completes the concept of CL-PLM in terms of information requirements of product beneficiaries in MOL of the product. The research question is as follows:

Which groups and organizations use the products or are affected by it and can be benefited from the product's operational information? The research involves the determination of the information needs of the stakeholders and tests the automatic provision of information to stakeholder organizations (beneficiaries) using data analytics methods.

The methodology used in this thesis is an inductive approach aimed at creating the concept of the information needs of stakeholders from MOL. To this end, two case studies on engineered products are carried out. Based on these two studies, the stakeholders involved in MOL, which benefit from the product and its information, are identified. Next, the dissertation identifies current and future information needs. Interviews and surveys have been used to identify information that MOL stakeholders need. The second part of this thesis examines, which data analytics methods can be useful for supporting each information need. Factors such as the characteristics of the identified information needs, the features of the data available from the product's function and the possible purpose of the decision have been considered to accomplish these goals. Based on these factors, suitable data mining techniques are proposed that can facilitate processing information needs. These techniques are also categorized. The categories are designed based on the area of applications of data analytics in CL-PLM. In other words, the classification is based on the usage of data analytics in the product lifecycle. After the completion of the analyzes, the findings are tested and implemented in real-world scenarios. The scenarios show how these information requirements can be satisfied with data analytics.

The findings of this research are various. First is to find out the MOL stakeholders of the product lifecycle and grouping the most influential ones, which are affected by product MOL data and information. Moreover, identifying their current and potential future information needs from the product MOL. Second is a classification of data analytics tools (suggestions on suitable tools) that can help the stakeholders to meet their new information needs from the product, or improve their access to product MOL information. The third part of the results is the implementation of the mentioned concept in three scenarios. In implementation part, information needs of OEM about the performance of a component in an electric vehicle, information need of maintenance provider from motorboat in respect of spare part supply and information need of wind farm operator from windmills for contact design are selected and modeled. The latter shows that it is possible to automatize the provision of the required information to the stakeholders. These three contributions form the output of this dissertation.

In sum, the findings of this thesis can be helpful for product and service manufacturers, especially those who want to move from pure production to the production of a product with the services, and organizations, who wish to add intelligence to their products. This research can also be useful for PLM software developers. Finally, it is helpful for all companies involved in the realization of engineered products that want to understand and manage the lifecycle of the product.

Keywords: Product lifecycle management, Closed-Loop Product Lifecycle Management (CL-PLM), stakeholders, Middle Of the Life (MOL), big data analytics, product service systems, engineered products

Zusammenfasung

Die Verbesserung des Produktlebenszyklus-Managements wird heute immer wichtiger. Unternehmen müssen die Produktlebenszyklus-Vorgänge effektiver durchführen, Produktinformationen teilen und den Informationsfluss zwischen verschiedenen Unternehmen der Produktwertschöpfungskette optimieren, um im Wettbewerb erfolgreich bestehen zu können. Gleichzeitig müssen sie neue Produkte oder Dienstleistungen entwickeln. Das Erreichen dieser Ziele ist für die meisten Unternehmen, die an der Produktwertschöpfungskette beteiligt sind, eine Herausforderung.

In diesem Zusammenhang versuchen diese Unternehmen, vorhandene digitale Daten zu speichern und zu nutzen sowie weitere neue Datenquellen zu erschließen. Ein wichtiger Teil dieser Datenquellen betrifft dabei das Produkt und seine Kunden in der Nutzungsphase. Mit diesem Hintergrund wurde das Konzept des Close-Loop Product Lifecycle Management (CL-PLM) entwickelt, um Unternehmen in der Wertschöpfungskette zu helfen, die Prozesse des Produktlebenszyklus zu verbessern und Daten und Informationen Durchgängig zu verwalten. Das CL-PLM adressiert dabei alle produktbezogenen Prozesse, vom Produktdesign bis zum Produktlebensende. Die Entwicklung des Konzepts des CL-PLM ist jedoch noch nicht abgeschlossen. Es besteht Bedarf an weiterer Forschung für die Entwicklung von geeigneten Lösungen zur Verwaltung der Daten- und Informationsflüsse. Ebenfalls ist die Entwicklung von Konzepten für die Bereitstellung der richtigen Informationen aus der Nutzungsphase eines Produktes (Middle of life, MOL) eine wichtige Forschungsfrage.

Im ersten Teil dieser Arbeit wird ein neues Konzept vorgestellt, welches den Unternehmen ermöglicht, Informationen aus der Nutzungsphase eines Produktes (MOL) zu erhalten und zu verwenden. Dabei wird folgende Forschungsfrage behandelt: Welche Gruppen und Organisationen (Stakeholder) nutzen die Produkte und können so von den Informationen aus der Nutzungsphase des Produktes (Information aus dem MOL) profitieren? Innerhalb dieser Arbeit werden des Weiteren die Informationsbedürfnisse des Stakeholders analysiert und die Bereitstellung von Informationen an diese unter Verwendung von Datenanalysemethoden getestet.

Die Methodik dieser Arbeit basiert auf einem induktiven Ansatz, um das Konzept der Informationsbedürfnisse von Stakeholdern an der Nutzungsphase (MOL) eines Produktes zu erstellen. Zu diesem Zweck werden zwei Fallstudien zu zwei konkreten technischen Produkten durchgeführt. Auf Basis dieser beiden Studien werden die beteiligten Stakeholder identifiziert, die von Produkten und deren Daten und Informationen aus der Nutzungsphase profitieren. Anschließend wird der aktuelle und zukünftige Informationsbedarf dargestellt. Dafür wurden Interviews und Umfragen durchgeführt, um Produkt-MOL-Interessengruppen und ihre Informationsbedürfnisse zu ermitteln.

Im zweiten Teil dieser Arbeit wird untersucht, welche Datenanalyseverfahren zur Unterstützung des Informationsbedarfs nützlich sein können. Dazu werden unterschiedliche Faktoren betrachtet. Dies sind die Charakteristika der identifizierten Informationsbedürfnisse, die Charakteristika der Daten, die aus der Funktion des Produktes erhältlich sind und das mögliche Ziel der Entscheidung des Stakeholders. Basierend auf diesen Faktoren werden geeignete Datenanalyseverfahren vorgeschlagen. Dabei werden diese in die Klassen Produkt, Betrieb und Geschäft innerhalb des Produktlebenszyklus eingeteilt.

Nach diesen beiden Analysen werden die Ergebnisse in drei Szenarien implementiert und das Konzept des Informationsbedarfs und der -darstellungen sowie die vorgeschlagenen Datenanalyseverfahren (Datenanalyseklassen für CL-PLM) getestet. So lässt sich ermitteln, wie diese Datenanalyseklassen tatsächlich die Bedürfnisse der Stakeholder erfüllen und die Informationsflüsse im Produktlebenszyklus verbessern können.

Die Ergebnisse dieser Dissertation lassen sich wie folgt zusammenfassen. Erstens wurden die Stakeholder an der Nutzungsphase eines Produktes und ihre Informationsbedürfnisse identifiziert. Zweitens wurden Datenanalyseverfahren klassifiziert und Vorschläge für geeignete Datenanalyseverfahren ermittelt, die den Beteiligten helfen können, ihre Informationsbedürfnisse bez. des Produktes zu unterstützen und ihren Zugang zu Produktinformationen zu verbessern. Drittens wurden drei Szenarien zur realen Umsetzung des Konzepts erarbeitet.

Für Produzenten von Produkten und Dienstleistungen können die Ergebnisse dieser Dissertation von zentraler Bedeutung sein. Dies gilt besonders für diejenigen, die von der Produktion von klassischen Produkten zur Produktion von Produkten mit kombinierten Dienstleistungen (sogenannte Produkt-Services) übergehen wollen, sowie für Entwickler von smarten Produkten, also Produkten mit einer gewissen Intelligenz. Die Ergebnisse sind ebenfalls für PLM-Softwareentwickler, sowie alle Unternehmen, die an der Realisierung von technischen Produkten beteiligt sind und den Lebenszyklus des Produkts besser verstehen und verwalten möchten, nützlich.

Acknowledgment

- Thanks to Professor Klaus-Dieter Thoben for the supervision of this dissertation. I appreciate his guidance and discussions during the accomplishment of this research.
- Thanks to Dr. Ingrid Rügge for her thorough support, from organizational, technical to mental. Thanks to her for always being there to solve my problems!
- Thanks to Professor Otthein Herzog for accepting the review of this dissertation as the second supervisor.
- Thanks to DAAD (Deutscher Akademischer Austauschdienst) for financing this research. Without their generous support, this dissertation could not form.
- Thanks to IGS (International Graduate School for Dynamics in Logistics) for supports and several offers in terms of the structured doctoral program, courses, financial support and the possibility of networking. All these opportunities had a great positive impact on the outcome of this research.
- Thanks to my colleagues in IGS, BIBA (Bremer Institut für Produktion und Logistik GmbH) and BIK (Institut für integrierte Produktentwicklung) who helped me during my Ph.D. Thanks to Morice Daudi, Karl Hribernik, Stefan Wellsandt and Moritz Von Stietencron for effective discussions and their good feedbacks on my work.
- Thanks to my parents, Parvin and Mohsen; and my brother, Ehsan. Their encouragements and continuous support helped me to calm down and keep on.
- Thanks to Andrea Stenner and Professor Reinhard Schweitzer-Stenner for their support during my Ph.D. life.
- Thanks to all my friends. Without their great companionship I could not be productive during the journey of doctoral studies.
- Thanks to my friends and colleagues who helped me in proofreading of this dissertation.
- Thanks to all professors in IGS for their thoughtful comments on my work. That made it possible for me to see my research problem from other aspects.

Table of content

List of abbreviations

AE	Absolute Error
ARIMA	Auto-Regressive Integrated Moving-Average
BOL	Beginning Of the Life
BOM	Bill Of Material
CART	Classification And Regression Tree
CBM	Condition-Based Maintenance
CL-PLM	Closed-Loop Product Lifecycle Management
CMMS	Computerized Maintenance Management System
CMS	Condition Monitoring System
CPS	Cyber Physical Systems
CRM	Customer Relation Management
CV	Cross-Validation
EEX	European Energy Exchange
EOL	End Of the Life
ERP	Enterprise Resource Planning
EV	Electric Vehicle
GIS	Geographical Information System
IMS	Intelligent Maintenance System
IoT	Internet of the Things
ICT	Information and Communication Technology
MAE	Mean Absolute Error
MCDM	Multi Criteria Decision Making
MLP	Multi Layer Perceptron

MOL	Middle Of the Life
MRO	Maintenance, Repair, Overhaul
MSE	Mean Square Error
NN	Neural Networks
OEM	Origional Equipment Manufacturer
OTC	Over The Counter
PDM	Product Data Management
PLC	Programmble Logic Controller
PLM	Product Lifecycle Management
PPA	Power Purchase Agreement
PSS	Product Service Systems
PUI	Product Use Information
QFD	Quality Function Deployment
RF	Random Forest
RMSE	Root Mean Square Error
SCADA	Supervisory Control And Data Acquisition
SCM	Supply Chain Management
SPC	Statistical Process Control
SVM	Support Vector Machine
SVR	Support Vector Regression
SWOT	Strength, Weaknesses, Opportunities and Threats
TPS	Transaction Processing System

1 Introduction

You cannot make a windmill go with a pair of bellows.
George Herbert

Those who believe that we have reached the limit of business progress and employment opportunity in this country are like the farmer who had two windmills and pulled one down because he was afraid there was not enough wind for both.
Morris S. Tremaine

Industry is one of the pillars of the economy. The manufacturing sector in the European Union solely accounts for 2 million enterprises, 33 million jobs and 60% of productivity growth of the economy (European Commission, 2017). Nowadays the industry is facing a new revolution with the advancement of the digitalization.

Digitalization has hugely influenced products and affected markets, which are involved in the development of products and their services. For example, digitalization has contributed to the connectivity of engineered products, availability of vast amounts of data from products, processes and user experiences, increase in E-commerce as well as the emergence of new technologies in product design and development. Technologies such as virtual reality, 3D printing, and digital twins have been newly introduced in product design and development. They help to connect designers with the products. Moreover, they help design engineers to improve communication with marketers and consumers.

Therefore, organizations have realized the opportunities for collaboration that this connectivity can bring to them (McKinsey & Company, 2014). However, Current procedures for the realization of product and services lack a straightforward approach for sharing product data and multidisciplinary collaboration between product stakeholders. Thus, recently different surveys have been done to redeem this. For instance, a survey has been done from 850 Product Managers and Product Management team leaders (Lawley, 2016). Based on this investigation (Lawley, 2016), profit increases by 34% on average if the product management procedures are optimized.

The concept of holistic product lifecycle management provides an opportunity for improving product related procedures. When holistic product lifecycle management is combined with advances in digitalization, it can bring an opportunity for better product management. That means, managing the product lifecycle is heading into a new phase defined by soaring flows of data and information (Manyika, et al., 2016).

This chapter provides an overview of holistic product lifecycle management and its importance for the organizations. Motivated by its importance and by the need to explore holistic lifecycle management problems, research questions are stated. Afterward, the methodology of the conducted research is delineated. The last section illustrates the structure of this dissertation.

1.1 Motivation

In an increasingly competitive and complex environment with more product and service options than ever (McKinsey & Company, 2014), a new opportunity for companies to stay competitive is to manage the lifecycle of their products or services from beginning to end. This dissertation identified the following most important driving factors that push companies to adapt to the overall management of product lifecycle:

- Adaption to technology advancement
- Competitive market, which meets customer expectation, and arranges better collaboration among organizations
- Environmental issues and product sustainability

In the following, these factors are explained in detail.

Firstly, technology advancement is one of the most influential factors in the overall management of lifecycle, i.e., increasing the amount of data, digitalization and Internet of Things (IoT). Furthermore, developments in the fields of industry 4.0, cyber-physical systems, techniques of data processing, ICT, sensors and wireless technologies and IoT keep changing the product lifecycle. Therefore, they affect the physical products, product data and information, organizations involved in using and managing products, and the networks in which the products are connected. The capabilities of engineered products together with IoT have provided the opportunity in sales to be linked to product development processes. This connection and the information exchange between two departments can lead to innovations for including features in products that allow on-demand upsells. Alternatively, connecting products through IoT networks provides the opportunity of learning from the operating condition that can directly be used for the improvement of products and portfolio (Zügn, 2017).

Secondly, a competitive market is another feature that produces the necessity of complete product lifecycle management. Companies must respond quickly to customer needs. They have to control and coordinate the product value chain effectively. Based on the customer's taste, they have to design new products and arrange for quick delivery. Preserving the required high response rate needs that activities throughout its lifecycle are managed more efficiently.

Thirdly, one of the environmental issues associated with the product lifecycle is the need to control the pollutants. Pollutants partly result from industrial processes that are caused by the production and transportation of the goods. Besides, the need to recycle products, the implementation of national and international resolutions and regulations, such as the Kyoto Protocol and Paris Agreement (United Nations, 1998; United Nations, 2015) are some of the requirements that companies must fulfill. Without having a holistic view of the product lifecycle, companies cannot reach the required level of sustainability and fail to comply with the regulations.

The challenges mentioned above motivate the research for this dissertation, which was aimed at investigating the enhancing of the holistic management of product lifecycle. Definition and characteristics of comprehensive product lifecycle management are presented in chapter 2.

1.2 Problem statement and research questions

Based on the motivating factors mentioned in the previous section, companies seek to manage the entire lifecycle of their products. Adapting lifecycle management perspectives can help companies to overcome the challenges of protecting the environment. Moreover, it provides them with more collaboration, opportunities, and gains from the market that may offer them economic advantages. Therefore, companies that are involved in the value chain of a product show more interest in managing the entire product lifecycle.

To enable companies to perform the overall management of the lifecycle of products, the engineering community put forward a solution, namely, to integrate the various stages of the product lifecycle. The Closed-Loop Product Life-Cycle Management (CL-PLM) concept conveys this philosophy. CL-PLM considers the complete lifecycle of the product. It controls all the processes that a product undergoes in its lifetime as well as stakeholders' activities during this cycle. However, the CL-PLM concept has not yet been completely put into practice. This dissertation addresses the research gap to complete CL-PLM in terms of data and information requirement.

Managing the product lifecycle during product design, development, and manufacturing has already been well investigated inthe product lifecycle. Namely, several information systems exist, which integrate the data of product design, development, and manufacturing. However, managing the product after it is sold to customers have been demanding and is still a critical challenge. For more information refer to (Jun, et al., 2007). After the product starts its operation, the manufacturer can rarely trace the product afterward. As a result, in terms of product lifecycle management, companies have little or no access to relevant information. This information could be operating condition of the product; the kind of maintenance or service the product receives; users or organizations, which use the product or benefit from its services. This type of information has been hardly accessible so far.

Thus, holistic management of product lifecycle could not be done in practice. However, current advances in research and technology have facilitated accessing product information. It is necessary to examine the sources of data, accessible through these technological advancements, and to examine the contribution of data and information for better holistic control of the product lifecycle. To this end, this dissertation focuses on the use of data extracted from the middle of life. Middle of life (MOL) starts when the product begins its operation. This life period of product lifecycle includes operation, use, maintenance, services, inspection and overhaul of a product. This dissertation examines data from MOL, such as operating condition, user behavior, maintenance and service information, failures, etc. Moreover, this research examines how value chain organization can be provided with the right data that they need (from MOL). The results of this research can enhance the insight or perspective, which organizations have at MOL, helping them to manage product lifecycle activities better.

To better understand the dimensions of the problem, the following research questions have been put forward. These research questions are therefore designed to take a broad view of the research area.

RQ1: What are the sources in which the data are generated and where this information is needed?

RQ2: What is the aim of using information from MOL? Who is interested in gaining knowledge from product MOL?

RQ3: Can the information support decision-making in lifecycle if the information from lifecycle phases are provided to beneficiaries? If not, what is needed to be accomplished?

RQ4: Can enhancing the richness of data improve the quality of decisions? If so, what is the appropriate mechanism?

More technical information and in-depth investigation regarding research problem, challenges and research gap against the problem of holistic management of product lifecycle is explained in chapter 2. Moreover, refinement of research questions regarding a suitable approach for providing the organization with relevant data is discussed in chapter 3.

1.3 Research goal and approach

The goal of this research is to support the development of a holistic concept for the management of product lifecycle. To this end, the research approach is to develop a concept for the integration of organizations, which are beneficiaries of product MOL, with the newly available sources of product use information (PUI). The objective of this integration is to provide relevant information for stakeholders; so that they can gain more insight for better decision-making (decisions regarding product, its operation, and collaboration with other beneficiaries). To understand the requirements (intents, expectations, and needs) of stakeholders and challenges in providing useful data, several case studies from MOL of products are conducted and analyzed. The first result of these studies is a conceptual model for defining "MOL stakeholders", their roles, and data needs. The second result of this research is appropriate data analytics techniques, which can support meeting the data needs of each stakeholder. For this purpose, data mining techniques are analyzed and grouped. Based on the informational features that stakeholders need and the specification of data sources, an appropriate tool is proposed for each information need. In the end, three examples are conducted in the form of case studies demonstrating the implementation of the concept in practice. For each case study, information needs are shown using proper data mining technique. Finally, this dissertation specifies and examines areas, where data analytics is useful for CL-PLM.

1.4 Procedure and structure of the dissertation

In this section, the overall structure of this dissertation is explained, which is divided into nine chapters. Table 1 displays this structure.

The first chapter provides an overview of the research topic. It states the research motivation followed by the research problem and the research questions. Subsequently, a brief explanation of the research approach and the leading research goals are presented.

The second chapter illustrates the extended problem of the conducted research. It describes the environment in which product lifecycle is performed, and introduces the elements of product lifecycle as well as their characteristics. Later it discusses the driving factors for this study. Subsequently, challenges to reach the goal are listed.

Chapter 3 describes approaches for using data and information in the lifecycle. First, it is discussed how stakeholders of product lifecycle have gained the information they needed. In this regard, main concepts and approaches for CL-PLM data management, as well as big data analytics are cited from state of the art. Moreover, details of data analytics approaches and its relevance to CL-PLM data are discussed.

Chapter 4 describes a concept for integrating the stakeholders of MOL in CL-PLM. Required data, information needs and a mechanism to accomplish these needs are suggested. Different types of analytics, which can support each stakeholder, enhancing its role in decision-making, in the product lifecycle, are also shown.

Chapter 5 argues the data analytics methods, which suit exploring data needs of stakeholders. This chapter extends the concept of the previous chapter. However, the technical layer is considered in chapter 4.

Table 1: Structure of the dissertation

	Chapter	Content			
1.	**Introduction**	Definitions	Motivation	Problem statement	Research goal
2.	**Extended problem statement**	Engineered products	Closed loop product lifecycle management	Challenges regarding digitalization and data and information systems, stakeholders and requirements	
3.	**State of the art for approaches to fill the knowledge gap**	State of the art of technologies in CL-PLM	IoT, big data, data analytics	Capability of data analytics to model PUI and answer the information needs of stakeholders	
4.	**Concept development**	Research methodology of the dissertation	Model stakeholder analysis in MOL	Identify data and information needs	Potential future changes in roles and data needs
5.	**Application of data analytics to solve the research problem**	Classification and selection of methods of data analytics for CL-PLM	Suggestions on classification of data analytics for use in CL-PLM	Integration of data analytics techniques with identified information needs	
6.	**Concept implementation**	Implementation Scenario 1	Implementation Scenario 2	Implementation Scenario 3	
7.	**Evaluation and discussion**	Evaluation of proposed stakeholder data needs model	Evaluation of data mining models from chapter 6	Discussion	
8.	**Conclusion and outlook**	Conclusion	Limitations	Answering research questions	Future research

Based on the results of the previous chapters, chapter 6 tests the implementation of the model. The implementation is done in three case studies of engineered products. The selected products are a leisure boat, electric automobiles and wind turbines. This chapter addresses useful information for product lifecycle improvement in each case study and models them with data mining techniques.

Chapter 7 discusses the results gained by this research. It also provides the validation of models implemented in chapter 6.

Chapter 8 summarizes the work done and provides an answer to research questions. It also provides the limitations of the research and puts forward suggestions for future research.

2 Evolution of the concept of product lifecycle management: Using all the lifecycle information

This chapter provides essential background information relevant to the basic idea of this dissertation. The content of this chapter is twofold. First, it gives the definition and necessary background and state of the art of research and development related to the concept of CL-PLM. Second, it shows the research gaps and challenges in filling the research gap. Section 2.1 defines engineered products and presents their developments. Section 2.2 discusses the management of product lifecycle. Moreover, it describes the concept of CL-PLM. Section 2.3 defines the elements of the product lifecycle and its characteristics. Section 2.4 discusses the research gaps in the management of product lifecycle. It lists the missing information sources and flows in the concept of CL-PLM. Moreover, it presents the importance of stakeholders as a major element of the product lifecycle. Based on the research gaps identified in this section, as well as on the motivation and the problem statement from chapter 1 (section 1.1 and section 1.2), challenges are identified and discussed at the end of this chapter in section 2.6.

2.1 Advances in the design and manufacturing of engineered products

The manufacturing perspective of products varies based on customer needs. Some products are mass-produced while others are produced in fewer quantities. On the contrary to mass-produced products, engineered-to-order products have high values and are designed on a smaller production scale, based on the order of the customer. For creating these products, the customer requirements are considered in the design of each part and component. Even the production process and parts list are designed individually for that specific product. The resulting individualized designs as well as fulfillment of the customer requirements cause a lot of cost and effort.

Recently, trends with the potential to change costumer involvement in the design and production of engineering-to-order products have emerged. Mainly, there are trends in additive manufacturing and digitalization. Resulting from these trends, the products and their respective components have received a high level of smartness and connectivity. In this context, there is a demand for more investigation of how to better involve customers in the product design. Thus, the cost decreases and at the same time customer needs are met.

As mentioned in the last paragraph, several engineered-to-order products are experiencing changes in the design such as gaining a kind of smartness. A group of products (with the focus on engineered-to-order), which are experiencing these changes is addressed in this dissertation. This dissertation calls these products "engineered products". Engineered products are defined as industrial products, which are complicated from an engineering design point of view and they should be built based on world-class quality standards (Rasmussen, 2017). The engineered products considered in this research have a longer MOL lifetime comparing other products, which is usually between 20 to 50 years. Different kinds of engineered products can be wind turbines, ships, aircraft, industrial production machinery and automobiles. They are used in industries, such as transportation and logistics, energy and process utilities, manufacturing and aerospace.

One particular characteristic of engineered products is that they can collect data about their usage and operation. Several monitoring and control systems are installed on these products nowadays to collect needed data effectively (Nabati, et al., 2017). For example, PLC (programmable logic controller) on manufacturing machinery and condition monitoring systems on vessels and wind turbines are some of these monitoring and control systems. A major goal of organizations to install such systems on these products is to provide them, or the product user, with information on product functionally. Information about product functionally can help organizations that design or operate these products. For example, the information can help to ensure a level of safety in using products, or monitoring products' performance and service needs.

Nowadays, emerging technological advances, such as IoT and industry 4.0, are supporting the collection of products' data. The products with these capabilities are often regarded as Cyber-Physical Systems (CPS), or in simple words, smart products. When the captured data is used together with the rest of the data in the lifecycle, several benefits can emerge. It is possible, for instance, to improve product capabilities (product upgrade). A production line can announce when it needs to be repaired. Alternatively, this information can also be used for the design of products such as autonomous cars that drive themselves. Although some advances have been made to integrate data of the entire lifecycle, still more attention should be paid to this field of technology development and research. Figure 1 shows the evolution of engineered products from physical product to connected systems.

Another characteristic of engineered products is that it is possible to offer these products together with services. Product service systems (PSS) determine the way by which products can be provided jointly with the services to the customer in a complete package (Baines, et al., 2007). Under this scenario, the user can use the product without caring for the maintenance service or caring for buying other extra services he may need. This concept can be exemplified under a car-sharing program (Baines et al., 2007). Leasing the components, such as engines, is an example of PSS for engineered products (Horton, 2016; Rolls Royce Corp., 2016; Eyer, 1979).

As per definition, there are mainly three types of PSS: product-oriented, use-oriented, and result-oriented PSS. The product-oriented PSS involves selling a product (as a product) in a traditional manner although additional services are included. Such services usually include contracts for services, maintenance, and recycling. The use-oriented PSS emphasizes on selling the use or availability of the product rather than the physical product itself. In the use-oriented PSS like sharing, pooling, and leasing a product, the manufacturer tries to maximize the use of the product. Finally, the result-oriented PSS refers to selling the result, capability or performance of the product. An instance of result-oriented PSS is selling the laundered clothes instead of selling a washing machine (Baines, et al., 2007).

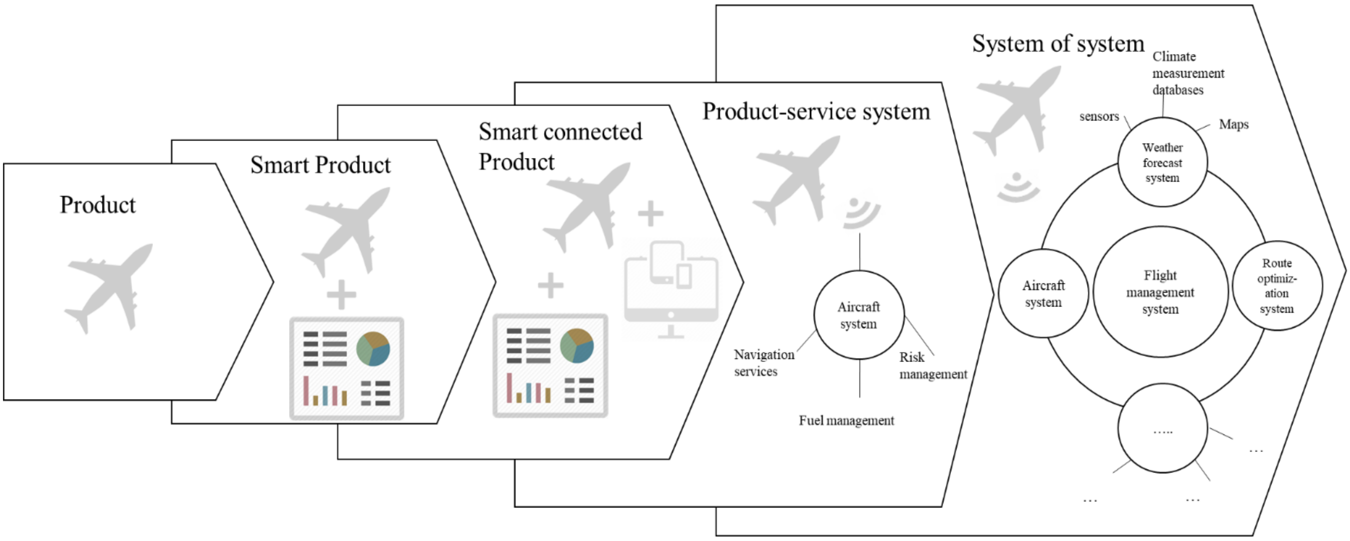

Figure 1: Development of products to connected systems (from (Schallmo, 2016))

The emergence of PSS together with digitalization advances in the engineered products have provided the opportunity to connect better to customers. The smart engineered products can collect and consequently provide required data, about the availability of a product, even while the product is operating for the customer. However, organizations of a product lifecycle need to develop more collaboration as well as innovation in the services, to realize the potential of mentioned changes in the improvement of products and customer needs.

In summary, characteristics of engineered products such as the capability of data collection and their integration with service provide opportunities to reduce costs of product design and interact with the customer for product realization more than ever before. This dissertation contributes to this issue from the perspective of the product lifecycle. It identifies customer requirements and their involvement in the creation of engineered-to-order products. In other words, this dissertation addresses the identification of customer's requirements as well as other product stakeholders' needs from the engineered product and integration of these requirements in the processes of product lifecycle such as product design and manufacturing.

In this research, we consider a group of smart engineered products, which are products equipped with the sensors and embedded technologies and they can collect data from their usage environment as well as, communicate and process these data (Jun, et al., 2007). For more information, please refer to (Lee, et al., 2015; Meyer, et al., 2009). In this focus, it is further assumed that decision-making is still done by human beings as key stakeholders, although it is not yet known who they are.

Next section describes current practices of managing the lifecycle of engineered products. Moreover, it presents special characteristics and effects that smartness in products has caused.

2.2 Product lifecycle and its management

Product lifecycle encompasses all the activities done by lifecycle actors as well as the physical product, its components and all the information related to a product throughout its lifespan. The following subsections explain the main scientific concepts concerning the management of the lifecycle of the product.

2.2.1 Closed-loop product lifecycle management (CL-PLM)

CL-PLM is a concept for managing the whole lifecycle of a product. It considers product lifecycle from a broad point of view. Before the emergence of CL-PLM, managing lifecycle of a product was seen under the concept of PLM. A principal difference between both is that, while PLM considers the product design and development, CL-PLM considers all the stages of the product lifecycle from design, development, manufacturing, and use (usage-utilization) until recycling and disposing at the end of life of the product. CL-PLM seeks effective management of product lifecycle. As stated by Jun et al. (Jun, et al., 2007) it is a new strategic approach to manage the product- related information efficiently over the whole product lifecycle. Its objective is to provide more product-related information over the entire product lifecycle to the organizations involved in the product value chain.

Recently several research works and projects (Wellsandt, et al., 2016; Borsato, 2014; PROMISE, 2010; Jun, et al., 2007) have proposed three phases f the product lifecycle. Beginning Of Life (BOL) includes product early stages of development, from idea generation, design, manufacturing until product sales. Middle Of Life (MOL) refers to the phase in which the product has started its operation. It can be studied by the time the product is sold to the customer. End Of Life (EOL) is the period in which the product terminates its operation and is dismantled, recycled or disposed. Figure 2 shows the main processes and phases of CL-PLM.

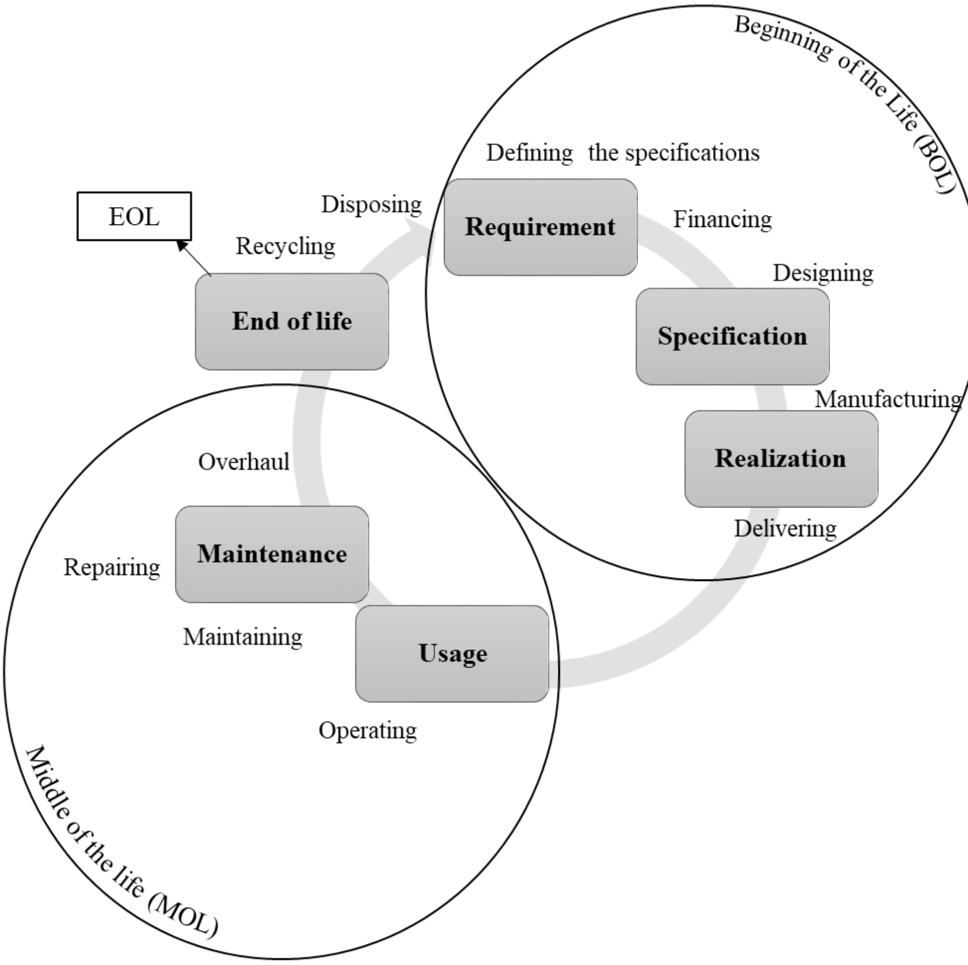

Figure 2: View of CL- PLM with major processes and phases

In order to study the effect of capabilities of a new generation of smarter engineered products on their lifecycle, it is necessary to have a better understanding of product lifecycle characteristics. So that one can take the opportunities for innovation or improvement in the product lifecycle. The following section introduces the elements of the product lifecycle.

2.3 Characteristics of the product lifecycle

The results obtained from the study of (Wellsandt, et al., 2016) suggests that product lifecycle contains some specific elements. Main elements, which can be observed in most of the scientific works, are as follows:

"Processes, stakeholders, material flow, energy flow, information flow, waste & releases flow, geometry and number of cycles."

For more information about these elements, see (Wellsandt, et al., 2016). In the following, primary definitions of the elements, which are used in this dissertation, are provided.

Processes: Lifecycle processes are set of activities that provide value. Every process receives input and transforms the inputs into an output. From the lifecycle view, the main tasks and activities that a product goes through are also referred to as processes (Wellsandt, et al., 2016). For a comprehensive list of the typical lifecycle processes, please see (Wellsandt, et al., 2016).

Stakeholders: These are the groups of beneficiaries including organizations or individuals who influence the product or are affected by the product.

Sharp, et al. (1999) refer to stakeholders as people, groups or organizations, which influence or are influenced by the organization's objective. This is a standard definition of stakeholders. This concept initially comes from management science, but it has been applied to several different domains of science afterward. Building on this definition, stakeholders of the lifecycle of a product can be referred to as people or organizations that influence or are influenced by the product, with their decisions or actions taken, along with the entire lifecycle of that product. From this point of view, the person who creates, builds, delivers, uses, services and disposes the product is a stakeholder in a product. More information about the importance of stakeholders is provided in section 2.5 of this chapter.

Information flow: Transfer of information from point x to point y. Information flows in the CL-PLM can happen between lifecycle phases, activities or stakeholders.

This dissertation addresses the elements mentioned above because current digitalization advances can enormously affect stakeholders, processes and information flow for smart engineered products. Relation of these three elements gives unique characteristics to the lifecycle of a new generation of engineered products (Wiesner, et al., 2014; Tempest, 2015). Firstly, advances in digitalization have caused a change in the "processes of CL-PLM", because new concepts of automation and new business models, such as data-driven business models influence the lifecycle processes. Secondly, "Stakeholders" of the product lifecycle are affected because their expectations change and their roles in the lifecycle alter (more discussion about stakeholders is presented later in chapter 4). In addition, "information flows" are changing, as new lifecycle-data sources are collected and the information sources are connected.

Following subsections explain data, information and information flows in the lifecycle of engineered products in more details. In addition, these subsections show the relevant state of the art.

2.3.1 Product lifecycle data and information flows

In order to discuss CL-PLM data, it is crucial to provide a clear definition of data in the product lifecycle. In this context, it is helpful to distinguish between data, information, and knowledge. Data are symbols referring to the facts as stated by Thoben et al. (Thoben, et al., 1999). Information is the reduced data that are processed and can provide a meaning for the user. It reduces the uncertainties of data and has the potential to support the user in decision-making. Knowledge is an understanding of information and is gained by learning, perceiving, experiencing or discovery. It enables someone to perform a task through the context-dependent selection, interpretation, and valuation of data. In addition, it takes time before knowledge can be used effectively, because the learning process is time-consuming.

In this context, a smart product is a product, which records its status; generates data, but does not create information or knowledge. When these data are integrated and processed, they are converted to information. A human being can understand information. Having relevant information, knowledge can be produced by human beings and not by-products.

In state of the art, product lifecycle data have been viewed by different researchers, sometimes as raw data (which contains no meaning), and in some cases as similar as information. In some other cases, data have been referred to all the three types of data, information and knowledge.

In general, information flows are the information or knowledge that is communicated between the stakeholders or transmitted in the processes (exists in the processes- streams into the processes). There are also information feedbacks, which refer to backward information. They are generated as a response to the completed task or process.

Previously, information flows in the product lifecycle were focused on BOL. For instance, the flow of information has been exchanged between designers and the flow of design information has been passed to the manufacturer. Recently, however, a full flow is considered between actors of all CL-PLM phases.

2.3.2 Channels of information in product lifecycle

Information sources have several different aspects. Typical information in the product lifecycle, which are managed by classical PLM systems based on Rachuri et al. (Rachuri, et al., 2008) can be considered as the form data, functional data, and lifecycle data. The form data are specifications relating to the physical product, such as features, 2D/3D models, surface models. Functional data (information) are about the functions expected from a product or services, which should be integrated into the product. The third part is lifecycle information such as information about lifecycle processes and phases of maintenance, design and disposal.

By including MOL in management of product lifecycle, several other channels of information emerged. These channels include:

- Data of product service and repair
- Log files (e.g., alarms), which are stored in a maintenance information system
- Web data generated by users regarding products
- Social media, which shows the opinion of users about the products
- Product use information that engineered products can collect

The data and information from the channels mentioned above are called “MOL data and information sources” hereafter in this dissertation.

2.3.3 Product Use Information (PUI)

PUI is a term for describing the data from a product when the product is put into operation (Abramovici, et al., 2008). Products that have sensors or measurement technologies can display or transmit their PUI digitally; such as measurement of temperature on main parts, speed, and alarms. If no measurement technology is installed on the product, the human operators might record the history of the function of the product. In this case, maintenance reports can be named as a type of PUI data. Other sources of PUI data contain environmental data, use condition data, product repair logs as well as data generated by the product user.

This work focuses on using PUI and this term is being used often in this dissertation. Although PUI is a part of “MOL data and information sources” (see subsection 2.3.2), in this dissertation, both terms have the same implication.

2.4 Developments in product lifecycle information flows and utilization of PUI

Supply of product-related information between phases of CL-PLM is a key to close the information gaps, putting the concept of holistic product lifecycle management (CL-PLM) in practice. Figure 3 shows the primary information flows between phases of the product lifecycle. PUI are generated in MOL phase. PUI can be used (a) to improve the processes inside the MOL phase. This information flow is presented as “Adjustment flow” in Figure 3; for example, in order to support the maintenance and repair process. (b) PUI can also be used to improve BOL. When PUI are exploited to design a new product or service this flow is addressed (“Product upgrade or product service” in Figure 3).

In recent years, there has been an increasing interest in modeling these information flows ((Hribernik, et al., 2017; Deng, et al., 2017). However, the PUI using flows, are still incomplete and need more investigations. This dissertation addresses the shortcoming of the incompleteness of information flows in the CL-PLM concept and contributes to completing information flows from MOL by identifying the needs of lifecycle stakeholders from the recently available data of product MOL.

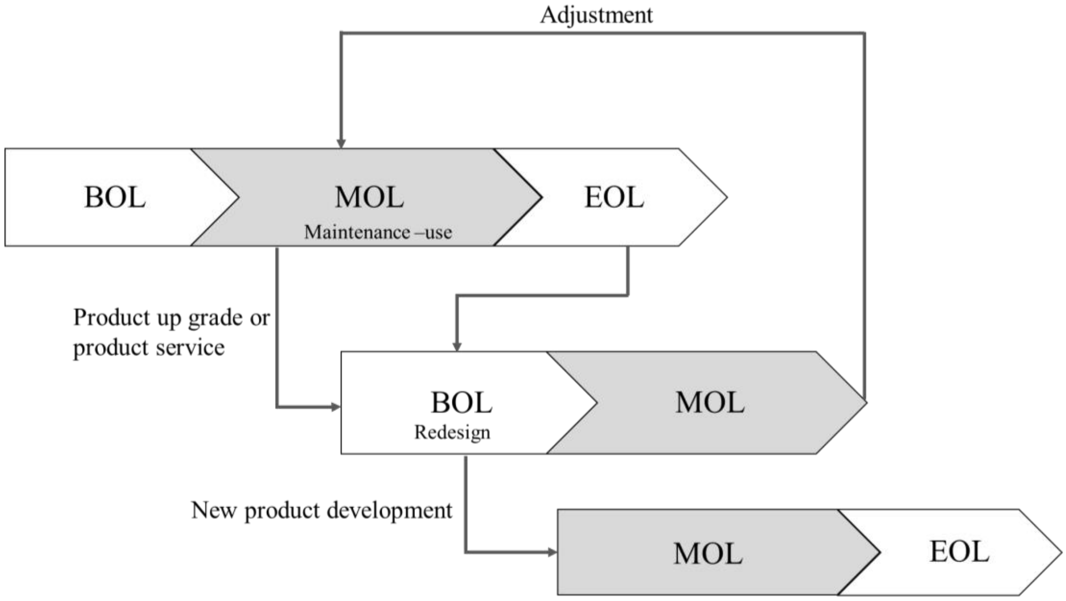

Figure 3: Major information flows between lifecycle phases (adapted from (Hribernik, et al., 2017))

2.4.1 Problem with product use information to integrate it in CL-PLM

It can be inferred from the definition of PUI (in subsection 2.3.3) that it comes from diverse data and information sources. However, data and information sources from MOL have not yet wholly been recognized for utilization in CL-PLM. The problems that make it difficult to recognize and use PUI sources is that they are still spread around the lifecycle processes, that several actors own them and that they are located in different organizations without being linked together. Therefore, there is a need for further investigations to recognize, use and add these data and information sources in the product lifecycle. Data management systems for integrating these channels of data are still under development (Kiritsis, 2011).

Our current knowledge about PUI and its integration in CL-PLM is primarily based on studies of (Abramovici, et al., 2008; Wellsandt, et al., 2015; Fathi & Holland, 2009; Hribernik, et al., 2017). Their research paves the way for an insight in of MOL and raises many questions at the same time in need of further investigation, e.g., what organizations do after collecting PUI? How can these data and information benefit them?

This dissertation focuses on the PUI, which are collected through engineered products. Since this is a new and emerging source of information for the CL-PLM, it can provide several promising feedbacks on improving the management of the product lifecycle and benefit product lifecycle organizations.

2.5 Importance of provision of right data to the right stakeholder

From CL-PLM perspective, reaching the goal of reducing environmental issues, improving product market and integrating technologies (as mentioned in chapter 1) may be possible by using a fundamental idea of the concept of CL-PLM. Utilizing this idea could close information loops between all the phases and processes of the product lifecycle. In this integration process, it is essential to identify and manage the activities, processes, and stakeholders of these phases. In order to achieve this, there is a need to select (use) and manage relevant data and information.

In this context, this dissertation addresses completing the concept of CL-PLM by focusing on better using PUI sources and understanding the data and information needs of MOL stakeholders. Better utilization of PUI sources can support organization in decision-making and it can facilitate the design of new services for the products. To this end, this research connects MOL data sources (with focus of PUI) to the stakeholder organizations, who can benefit from them and suggest relevant tools to support this. In the following importance of this focus is discussed.

Importance of MOL. MOL of a product is the longest time in the lifespan of the product. Apart from the fact that MOL is long, it contains valuable information about the product and its users (product customers). In order to make holistic management of lifecycle possible, MOL should be understood from different aspects, i.e., that of major MOL stakeholders, MOL processes and available information.

Importance of stakeholders in CL-PLM. On a broader sense, CL-PLM is a concept, whose aim is assisting stakeholders in getting a better understanding of the product and its related activities. Therefore, on the one hand, the primary focus of further development in CL-PLM should be, serving the organizations who are beneficiaries of the product lifecycle. On the other hand, stakeholders' behavior, cooperation, decisions, and needs affect the product lifecycle. Thus, characteristics of stakeholders, their needs and wishes should be identified and considered. An example of the interest of a stakeholder can be the wishes of automobile users which can lead (or convince) automobile manufacturer to design a new product. Alternatively, the use of specific material in a product, by the manufacturer, can affect the lifetime of the product and this effect can be identified by integrating PUI data in product maintenance or design process.

Researchers such as (Nilsson & Fagerström, 2006; Abramovici, et al., 2013) have addressed stakeholders in the product lifecycle. However, still more research is needed to understand the roles of product lifecycle stakeholders, their needs, and decisions.

Importance of MOL data sources. Importance of MOL data and information sources, particularly PUI, arises from the fact that using MOL data and information is accompanied by complications that include lack of channels for accessing data, the sparsity of data, high volumes of data and user privacy issues. More explanations in this regard are provided in chapter 3.

The challenges that organizations face in achieving the goal of holistic lifecycle management and integrating data, information and stakeholders of product MOL into CL-PLM concept is addressed in the next section, 2.6.

2.6 Challenges of organizations for getting relevant data from product lifecycle

Based on the importance of the problem stated in section 2.5, this section mentions the challenges in supporting product lifecycle stakeholders and alleviating their problems, by adopting a CL-PLM perspective.

Based on available research results (Wuest, 2014; Jun, et al., 2007; Hribernik, et al., 2017; Lindström, 2016) the following challenges exist in the lifecycle of engineered products. The following options should be taken into account to improve management lifecycle of engineered products:

1. Utilizing emerging technologies and new information sources in the product lifecycle
2. Need for innovation in engineered products and their related services
3. Improving the collaboration between stakeholders of different phases of the lifecycle (mainly, MOL/ EOL with BOL)
4. Integration of useful information feedbacks from MOL in the product lifecycle
5. Secure and practical information sharing between the organizations of the product value chain
6. Increasing sustainability and reduce the environmental impacts of products and processes.

These key challenges highlight the status of the holistic management product lifecycle and show that levels of linking the elements of the lifecycle are increasing. Not only the processes but also the organizations and products are connected. Focusing to tie the elements in the lifecycle, to provide streamlined information, process, and flow, following detailed challenges should be considered:

- Better and easier information management between the stakeholders of the product lifecycle
- Providing stakeholders with the right information
- Automatizing the information flows
- A better understanding of MOL, its mechanisms, data sources, beneficiaries and their needs

- Adopting the right technologies and tools for serving stakeholders, processes and information sharing.

There is a research gap to connect MOL elements in terms of data sources and stakeholder together when we consider the challenges mentioned above in the current practices of the product lifecycle. Therefore, this research work is undertaken to provide a better understanding of MOL for engineered products in terms of stakeholders and their needs from emerging data sources.

The main goal of chapter 3 is to identify a suitable tool or approach that can be employed as a supportive measure to connect lifecycle stakeholders to data and information that they need. Chapter 3 describes appropriate tools for supporting lifecycle stakeholders as well as the characteristics of these tools. Subsequently, one of these tools is investigated.

As shown in chapter 4 of this dissertation, stakeholders of MOL are identified, as well as their data and information needs from PUI sources are presented. Chapter 5 explores the characteristics of a suitable tool for processing the identified requirements of stakeholders (data and information requirements). Chapter 5 selects data analytics as it can be an appropriate medium to connect data sources and stakeholders. Chapter 6 demonstrates the implementation of the use of data analytics tools for providing CL-PLM stakeholders with relevant information from PUI. This demonstration is done with real-world case studies.

3 Approaches for providing and using product lifecycle data

Throughout time, stakeholders have used several methods to find the necessary information they require regarding product lifecycle. For example, if they want to buy an engineered product they can perform a Lifecycle Costing method (Estevan, et al., 2018) to decide the most cost-effective option for purchasing. Alternatively, they have used simulation or system dynamics to gain an understanding of the real condition of the use as well as improving the design of the product. This chapter presents the current approaches and solutions, which can enhance value from product data and information. For this aim, first, the state of the art of data and information management systems in the product lifecycle is described (subsection 3.1.1 and 3.1.2). Then, other relevant approaches, apart from lifecycle data and information management systems are discussed (subsection 3.1.3). Next, section 3.2 compares the current approaches and a suitable approach for processing MOL data. The candidate techniques, which suits MOL data, is data analytics. Section 3.3 explains data analytics. Finally, the challenges of applying data analytics to product lifecycle are presented in 3.4.

3.1 Using data and information to satisfy information requirements

Based on the challenges, which are described in the previous chapter for holistic management of product lifecycle, this chapter answers the following questions; how is it possible to provide stakeholders with the right information? Moreover, which tools and approaches are currently used? To answer these questions, this chapter investigates the IT-based approaches that stakeholders often have used for accessing, managing and processing data and information in the lifecycle of products. This chapter shows that the selection of tools used by stakeholders depends on the task they perform (their roles) and the characteristics of data sources. Therefore, [big] data analytics is selected as an appropriate tool, which matches characteristics of PUI sources, and is flexible enough to be used by different stakeholders for various decision-making purposes. Finally, this chapter presents the challenges of applying data analytics on PUI. Figure 4 gives an overview of approaches, which is addressed in this chapter

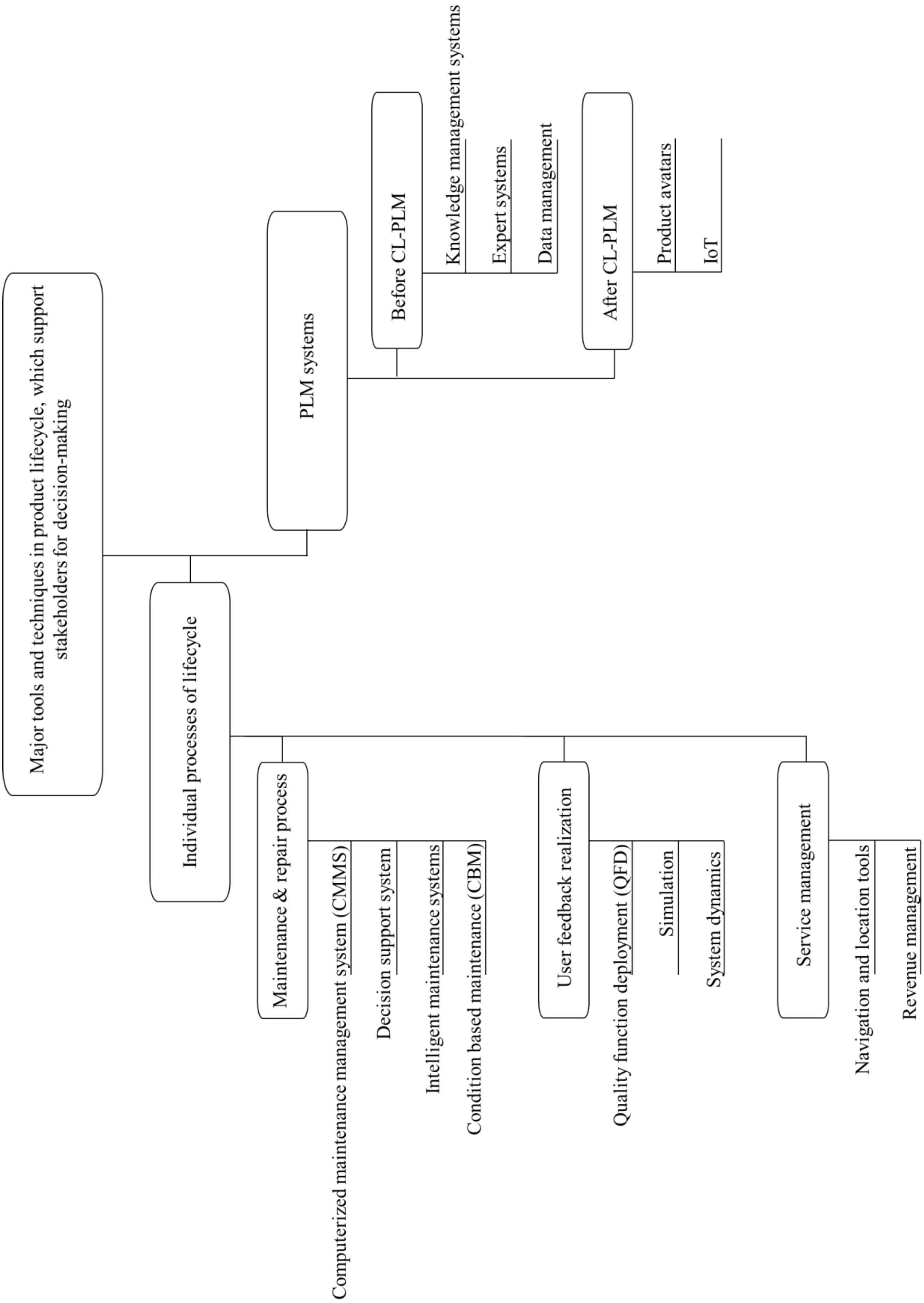

Figure 4: Approaches for supporting decision-making with and without lifecycle perspective

3.1.1 Product Lifecycle Management systems

PLM systems are one of the central IT-based systems that have been used by organizations to manage product-related data. PLM systems are originated from Product data management systems (PDM). PDM systems cover data from the design and development of products. They have functions for managing products during the design and development process. These functions include components and material flow control, or so-called bill of material (BOM), as well as data and information about physical shapes of the product, design files, BOM and version control.

Over time, however, the capabilities of PLM systems increased. Management of product lifecycle developed in the direction of becoming an integrated and comprehensive solution. This view of PLM systems has become very popular. For example, as stated by Stark (Stark, 2006), PLM systems are a solution that not only integrates complex and tailored products, based on customer wishes; but they also take into account the challenges of a more competitive international market and the need of companies to be competitive. Therefore, advances in PLM systems have made it easier for the involved organization to access lifecycle data and information. This advancement has supported organizations by providing them the possibility of getting a better insight into processes and making faster decisions. The primary information systems that are currently used in connection with PLM software are such as the ones listed in (Werner, et al., 2004; Müller, 2016). Despite the large scale of these systems, these information systems still mainly support stakeholders involved in product design and realization.

Data and information modeling is the dominant approach in PLM systems to control and enhance the value of product-related data and information. Stakeholders can then retrieve the structured product-related data via reporting tools and database system queries.

In terms of data and information modeling, researchers have modeled lifecycle data and information from different perspectives. For example, from the standpoint of design related data (e.g., product function and form) (Rachuri, et al., 2008), or from the perspective of modeling and management of information exchange, between manufacturer and suppliers (Belkadi, et al., 2008; Schilli, 2006). Some approaches also consider standardization of product data and information management. For instance, Surdasan et al. (Sudarsan, et al., 2005) present a product lifecycle information model, which integrates suggestions, assemblies, and functions. They developed a comprehensive model based on NIST Core Product Model. The model can catch product, design, and assembly information from the initial design process and make the information available for the other lifecycle processes and users. Although the research works provide valuable models for sharing product-related information they do not cover information exchange and management in the whole product lifecycle. Moreover, despite the fact that data and information modeling and management are practical approaches for enhancing product lifecycle data and information, applications of them are mainly investigated in BOL (such as for product designer). Little attention has been paid to manage PUI, coming from processes such as maintenance and product operation. In other words, developing relevant tools for integrating information from other processes of the lifecycle such as maintenance, logistics, and product operation process into CL-PLM are missing.

From another point of view, researchers proposed several additional capabilities and functionalities for PLM systems. For example, the addition of the traceability feature of the product has been proposed (Jansen-Vullers, et al., 2003; McFarlane & Cuthbert, 2012). Xiao et al. proposed linking asset management solutions to the PLM systems (Xiao, et al., 2010). Luh et al. presented a methodology for a distributed PLM platform, which can serve globally distributed enterprises (Luh, et al., 2011). More discussion about this issue is presented in section 3.1.3. Adding new functions to PLM systems made it possible to connect useful information systems to PLM, e.g., decision support systems, knowledge-based systems, and expert system. Appendix A provides a review of these approaches. These systems support stakeholders (mainly stakeholders from BOL) to define and access their information needs. Although all of these systems are advantageous and necessary, they are designed for a specific area of knowledge or application. Moreover, the performance of the systems is negatively affected by an increase in the size of input datasets.

In summary, this section reviewed advances in PLM systems in terms of data and information management. Apart from focusing on PLM systems, stakeholders can benefit from the holistic product lifecycle management system, which is still under development. The current state of the art of CL-PLM systems and the extent, to which they can currently cover stakeholders of all lifecycle phases, are discussed next, in section 3.1.2.

3.1.2 Holistic management of product lifecycle (CL-PLM systems)

Unlike PLM systems, few IT-based solutions are available, when it comes to managing the whole lifecycle of the product. However, in the last decade, there has been a rapid development of research in holistic product lifecycle management systems and technologies. For example, one of the available systems introduced in the PROMISE project (PROMISE, 2010). Product Knowledge and Data Management (PKDM) is an IT-based system used for integration and managing the data and knowledge from the lifecycle processes. Cassina et al. (Cassina, et al., 2006) provided a system architecture for PKDM that supports analysis of lifecycle data in a closed loop format. Therefore, the system can be used by the organizations across the respective product lifecycle and they can gain insight from the whole product lifecycle phases.

Apart from this research, scientists worked on using cloud services for sharing information sources across the lifecycle and connect PLM systems to other information systems in the product lifecycle. Shilivitsky (Shilivitsky, 2014) mentioned that PLM systems are transformed from single and isolated information management systems to a new version, which took into account the cloud services. Moreover, companies such as PTC also provided services for PLM software users to use PLM cloud (CIMdata, 2017).

Lately, there exist other research approaches, which were aimed at closing the information gaps in the product lifecycle. Table 2 lists important research works in this respect. However, in spite of the availability of these studies, little research has been done so far on developing a system for holistic lifecycle information management of products.

From the perspective of tools and techniques, which are used in CL-PLM systems, there are still works to do. Data and information management in CL-PLM is not as easy as it was in PLM systems, because CL-PLM considers all the lifecycle processes. The dataset has a considerable size and cannot be easily stored. Moreover, structuring and modeling data for inclusion in database systems may not be possible, due to memory issues. Researchers believe that using new technologies can help overcoming the challenges of developing CL-PLM systems. In the following, some of the solutions are introduced.

Table 2: Data, information and knowledge management in CL-PLM

Paper	Description	Studied aspect
(Ameri & Dutta, 2005)	Role of PLM as knowledge management system	Emphasis on closing knowledge loops, the role of PLM in supporting knowledge-intensive processes throughout the product lifecycle.
(Terzi, et al., 2007)	A meta-model for CL-PLM (called Holonic)	Integrating product traceability along the product lifecycle
(Fathi & Holland, 2009)	Metamodel for MOL data. Challenges of PUI	Emphasis on feedback integration from lifecycle to knowledge-base system
(Abramovici, et al., 2009)	Metamodel for an industrial product-service system in the lifecycle	
(Taisch, et al., 2011)	Standard development for PLM data management	Integrated management of all product-related information. Developing PDKM a semantic data model, which is written in XML language

Paper	Description	Studied aspect
(Siemens PLM Software, 2011)	Product lifecycle data sharing framework	Focus on open data as a solution to make product-related data accessible
(Kiritsis, 2011)	Ontology-based semantic standards framework for product lifecycle management together with related knowledge management and sharing in the lifecycle	CL-PLM Ontology-based model proposal
(Umeda, et al., 2012)	Lifecycle engineering by considering product sustainability	EOL integration in the lifecycle

Product Embedded Information Device (PEID). PEID is responsible for making a product uniquely identifiable. PEID can not only locate the product, but it can also collect data about it and its environment (PROMISE PEID Grouping, 2009). PEID allows temporary storing data and even has a limited ability to process data.

Product avatars. A product avatar or a digital twin is a virtual model of the product. they work together with PEID. PEID make linking the physical product with its virtual version possible. Digital twins not only facilitate accessing lifecycle data but also the analysis of data and monitoring of systems to head off problems before they even occur. They can reduce downtime and even plan for the future by using simulations (Marr, 2017). Although the concept of the digital twin is available, its real integration in CL-PLM is still not completely realized. It is the focus of currently ongoing research.

IoT. IoT considers a network of smart connected products. Similarly, a product with a digital twin can be a candidate to operate in IoT network. IoT takes into account not only products and product-related data, but it also considers several aspects, such as connected infrastructure, data transmission, remote monitoring and interconnection of the applications with each other and with the users.

Still, several critical issues (e.g., sensor deployment, big data and analytics, information management, information services, query processing, network security, software system) (Maglaras, et al., 2017) are not addressed in the IoT. That is true for the application of IoT in CL-PLM. There are currently several undergoing research projects to resolve this issue.

Big data technology. Big data (data analytics) technologies facilitate storage and processing vast and complex amounts of data. Big data uses decentralized data storage to overcome problems with the size of data. It also provides the ability to process data with large volume and sophisticated structure, for example, advanced analytics tools, which can handle datasets with multiple parameters. Section 3.3 explains why big data technology can be a suitable approach to PUI sources. This research work investigates the applications of this technology to overcome the challenges of holistic lifecycle management when using PUI sources.

To summarize, this section introduced the approach to holistic lifecycle management. This area still needs further research. Apart from PLM and CL-PLM systems, that are explicitly designed to manage the lifecycle of the product, organizations use several tools and techniques for accessing and enhancing the information they need for each process in the product lifecycle. Tools and techniques to use data from products or improving data to extract valuable information are based on the process and its actors. An overview of important tools in individual processes is described next, in subsection 3.1.3. These are ones, which stakeholders have used to assist decision-making. Product lifecycle consists of several processes; however, this dissertation focuses only on tools and techniques used in processes inside the MOL phase as well as information flow from user feedback to the product design process. The reason for this choice is that PUI data can play a more critical role in improving MOL processes rather than other lifecycle processes.

3.1.3 Current techniques and tools for improving information in the product lifecycle (individual processes)

Processes, such as, product commissioning and setup, product use or operation, maintenance and service, product repair and overhaul, form MOL of a product are in the focus of this subsection. This subsection aims to present the (state of the art of) primary tools for decision support or for enhancing information, which the actors have typically used to decide on the data from the product or its services. In this context, PUI is a part of data from the product or its services. It is remarkable that the tools mentioned here are often developed for a particular process and support the role of the decision maker. Thus, they are not designed to support the perspective of holistic lifecycle management.

Considering using data from MOL, the major processes that benefited from product related data are maintenance and repair as well as user feedback to the design process. State of the art of research shows that these processes have gained more attention compared to other MOL processes, such as product setup, product operation or product overhaul. In the following, first, a general characteristic of organizational IT-based systems is introduced. Then, primary IT-based and non-IT based tools for enhancing information and support decision-making in these processes are described.

Organizational decision-making and information systems (IS). A characteristic of decisions in organizations is a dependency on the corporate level. Based on the literature, there are three major decision types based on organizational levels (Montana & Charnov, 2008). The operational decisions, tactical decisions and strategic decisions. Operational decisions are day-to-day decisions, which usually concern the daily operation and functionality of the product. A tactical decision is related to planning the activities. They involve a longer time horizon rather than the operational level. They can consider one or several products. Strategic decisions are related to long-term decisions. These decisions take into account long-term changes such as market positioning, technology adaption, and capital investment. This characterization of these decisions also applies to decision-making in the lifecycle. To this end, the tools and techniques, which are available for improving information in every process of the product lifecycle, depending on whether the decision maker comes from strategic, operational or from tactical level. In the following, some of these tools are illustrated and explained.

Major IT-based and non-IT based tools for enhancing information for individual processes of the lifecycle. In this subsection, tools used for decision- making in maintenance process and process of user feedback to design are discussed.

Maintenance process. Primary tools for supporting stakeholder in the maintenance process address either process of fault diagnosis, detection or repair. Therefore, different decision support systems and knowledge management systems have been developed for maintenance, service and repair processes. The following lists major information systems in maintenance; Computerized Maintenance Management System (CMMS), Intelligent Maintenance System (IMS) (Abramovici, et al., 2013) (Zuccolotto, et al., 2013), Condition-Based Maintenance (CBM) (Gulledge, et al., 2010; Rahej, et al., 2006), Condition Monitoring (CM) and Supervisory Control and Data Acquisition (SCADA) (Hameed, et al., 2009). See Figure 5 for major tools.

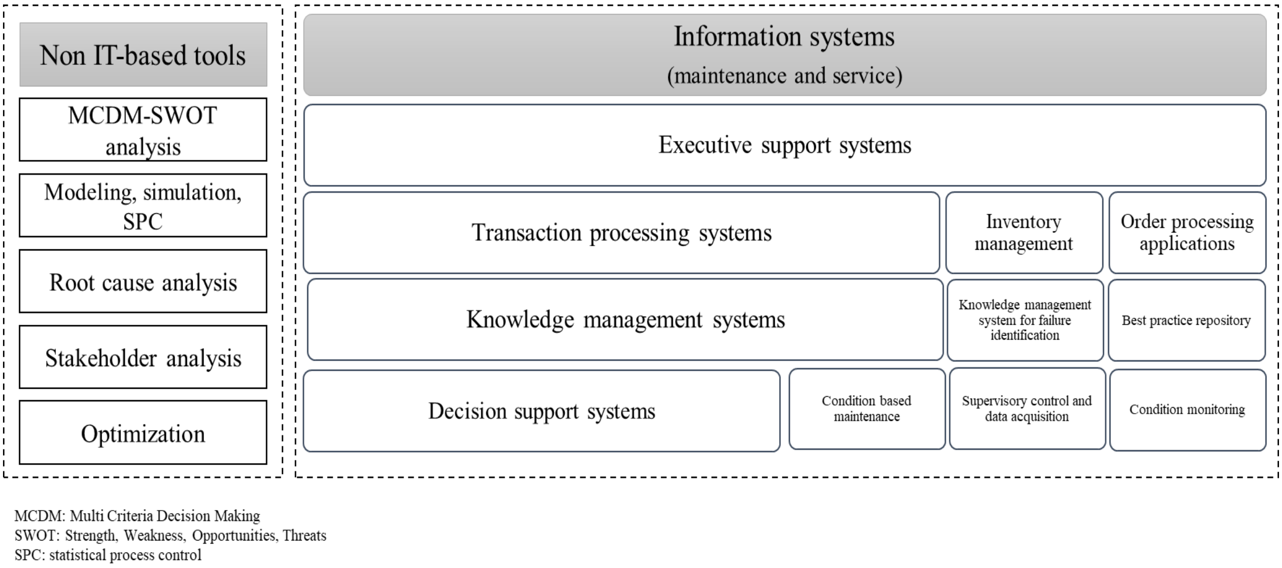

Figure 5: Examples of tools for decision-making and information enhancement in the maintenance process

User feedback to design. These tools allow the product user to communicate with designers. In the following, the major groups of these tools are listed. For more description regarding each group and conventional software and applications, please refer to Gilliam (Gilliam, 2017).

- Voice of the customer tools
- Survey
- Online review tools
- User testing tools
- Visual feedback tools
- Community feedback tools

Apart from IT-based tools, other techniques such as simulation, optimization, root cause analysis, SWOT matrix are the major tools used by stakeholders.

Overall, this section studied the tools, which organizations use in maintenance process and user feedback to the design process. A vital limitation of the mentioned tools for this research is that they lack a holistic lifecycle view of the product. On the one hand, the use of these approaches is limited to specific tasks or processes. On the other hand, knowledge generated by the use of some tools is only available to a particular group of stakeholders. Usually, these are just the people who deal with this tool. For example, maintenance information and insights are not shared with manufacturing organizations. If the mechanism for effective and secure sharing of data would be available, processes in the lifecycle could be optimized easier. Therefore, there is a need for holistic approaches.

To wrap up and to arrive at a conclusion with regard to the currently available approaches for using product data, the following statements can be made. The tools (section 3.1) for improving data can be divided into two groups. A group of tools specifically addresses product lifecycle. In this group, software and systems for CL-PLM are still under development. The other group of tools does not provide a holistic lifecycle perspective (as shown in subsection 3.1.3). In order to improve problems of managing product lifecycle information (as stated in section 2.4) it is advantageous to examine the possibility of integrating processes, data, and stakeholders of MOL, which currently lack a holistic lifecycle perspective in CL-PLM. As mentioned in section 2.1 and 3.1.2, advances in smart engineered products and technologies of PIED, digital twin have made it possible to share data from MOL to the other phases of the lifecycle. As stated above, the data sources, which are collected by these technologies, are called PUI. However, it is still unclear what kind of characteristics PUI sources have, and how it is possible to provide stakeholders with the right PUI information. The next section discusses the aspects of PUI and limitation of current available CL-PLM systems to deal with them.

3.2 Limitation of approaches to answer the needs of stakeholders when considering PUI

PUI was described in subsection 2.3.3 and challenges that organizations face to use it and make it transparent was discussed in 2.4.1 and 2.6.

Chapter 2 stated that PUI could be gathered by smart products along the different processes and phases of CL-PLM. The sensors installed on smart products, can stream the data, such as environmental condition, the status of product and history of changes, the type of use as well as performance of the product. There are also other sources of product usage data, which can be gathered from mobile applications, social media and websites. These types of data can show the user's opinions about the product or problems with the product.

To benefit from PUI data in the product lifecycle, using existing solutions (subsection 3.1.1 and 3.1.2), first data and information modeling and then the data and information management tools should be implemented. Specifically, product-related data should be entered into database systems. In addition, users should be able to benefit from the data. However, PUI has deficiencies that make it difficult to implement the above-mentioned data modeling and management tools. Generally, PUI sources are generated very quickly. For example, in the case of sensor data, the measurements can be done every few minutes. As a result, tools must be capable of receiving data every few minutes.

Moreover, PUI has various formats and characteristics. For example, PUI data from sensors can be in the form of numeric values in log files, while the data from maintenance reports are unstructured and presented in textual format. The diversity of file formats challenges data and information modeling. Since the variety of data sources are high, structuring is not easy. In addition, current databases in PLM systems are hardly compatible with different file formats.

Furthermore, PUI usually consist of batches of data for every measurement interval. That increases the dimensionality of datasets and makes the definition of the data models in data and information management systems more complex.

Therefore, based on the mentioned three challenges of PUI sources, having the amount of the PUI and the speed of their generation, we are facing a significant amount of data. It can be said that PUI sources have the characteristics of variety, velocity, and volume (3V), similar to the characteristics in the context of big data analytics (Nabati & Thoben, 2016).

Thus, a solution for overcoming the challenges to the integration of PUI into CL-PLM systems can be to use big data technology for PUI. To that end, after processing PUI with this technology, it is possible to integrate the results into CL-PLM system. Therefore, applying big data technology is a relevant approach and this dissertation tests this possibility and examines how it works. Section 3.3 explains big data technology.

3.3 Big data and data analytics technologies for covering PUI characteristics

Big data is a term that recently has gained popularity. It refers to an increase in the volume and complexity of data and information in the industrial and economic systems. Hashem et al. (Hashem, et al., 2014) state "Big data are characterized by three aspects: (a) data are numerous, (b) data cannot be categorized into regular relational databases, and (c) data are generated, captured, and processed rapidly".

Some authors extend the definition of 3V. Such as, Fosso-Wamba et al. (Fosso-Wamba, et al., 2015) who defined big data as a ''holistic approach to manage, process and analyze 5Vs (i.e., volume, variety, velocity, veracity and value) in order to create actionable insights for sustained value delivery, measuring performance and establishing competitive advantages" (Fosso-Wamba, et al., 2015).

The data analytics is a part of big data technology, which aims to convert the data into useful information. This information has the potential to provide insight into decision-making. For example, in the case of maintenance, it can help to find the failures before they happen (Nabati & Thoben, 2016). The terms big data analytics, data mining and data analytics are used interchangeably in this dissertation in spite of slight differences. Next, big data technology is explained.

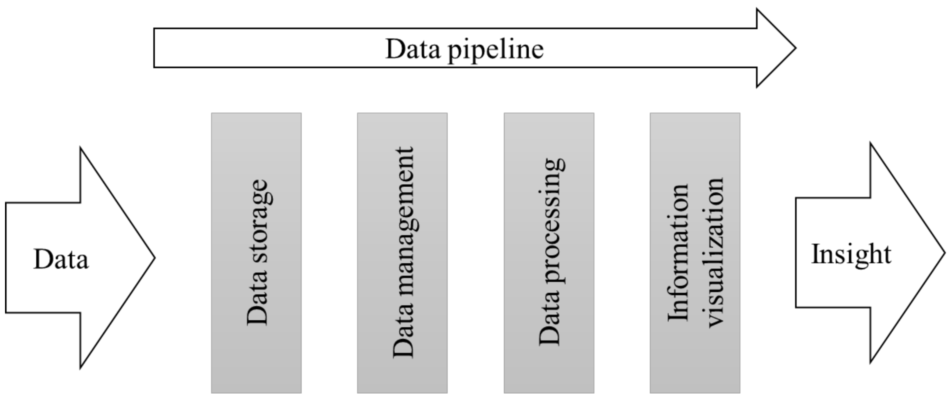

Figure 6: Steps of big data technology

The available technology for harnessing big data has four main parts (Law, et al., 2014). The first part contains, data storage appliances, either centralized or decentralized. The second part involves parallel processing and contains a distribution module, which is the medium to take the data from storage to analysis systems. The third component is data processing. Processing contains ingest and data analytics. This tool is applied to obtain insight from data. The fourth component is visualization, which is used for illustrating, reporting and distributing the results (Figure 6).

Some research also considers "Infrastructure" as an essential part of big data technology (see Table 3).

Table 3 shows different open source and commercial software/ hardware products used to handle big data. Later in chapter 6, this dissertation uses R software.

Table 3: Big data technologies and example of software (Law, et al., 2014)

Category	Used for	Open-source software	Commercialized software
Data Storage	Store data and meta-data	MySQL, Cassandra	Oracle, Marklogic, Teradata
Ingest (Middle ware)	Transformation, normalization, ingest	Hadoop/ Map reduce, Storm	Oracle, SAP, IBM, Amazon web service
Analytics	Data mining, machine learning	R, Hive, Rapid miner	SPSS, SAS, Matlab
Visualization	User interface, dashboard, Web based tools	Ozone, Gephi-Three.js	Tableau, SAS
Infrastructure	Operating system, computers, cloud, networks	Linux, Openshift, Openstack, Oozie	Microsoft Azure, Red Hat, Cloudera

Storage. Storage for the big data platform can have different forms. Centralized storage and distributed storage are the major accepted architectures. In the centralized storage, all data are transferred from the data generator to a central repository. For the big data, this is not always possible, because of the massiveness of datasets and the need for scalability of the data repositories. Scalability is necessary to accommodate the data, which is increasing in time. To overcome the problem with scalability, the decentralized storage is more common for the big data.

Ingest. Data ingestion is performed in order to capture and transfer the big data into the system and make it ready to use. "Data ingestion is the process of obtaining, importing, and processing data for later use or storage in a database. This process often involves altering individual files by editing their content and/or formatting them to fit into a larger document. When numerous data sources exist in various formats (the sources may number in the hundreds and the formats in the dozens), maintaining the data at a reasonable speed and efficiency can become a major challenge" (Rouse, 2016).

Analytics. Once the data are stored, distributed and ready to use, the next step is to process the data. The data analytics can here support processing the data, transform it into information and help to find potentially useful insights for the decision makers. More information about data analytics is provided in subsection 3.3.1.

Visualization. Data visualization is the next step in the workflow of data analytics. The aim is to present an efficient picture of insight or abstracted information. The visualization is done usually by graphs, images, and charts, which are easily understandable. Enhanced techniques of data visualization have recently gained popularity. As a good example, several innovative software developed for visualization can be mentioned. Tableau and Grafana are two of these software. Comprehensive visualization can lead to a better understanding of the problem and consequently contribute to more realistic decision-making.

3.3.1 Data analytics: Approaches and techniques

Data analytics can be considered as a part of big data technology, which supports decision making by discovering knowledge from the data and information. This technology contains several algorithms and approaches. The approaches are originated from machine learning, statistical analysis, mathematical optimization and database management. Recently, authors also proposed hybrid approaches, which take simulation into account. However, mathematical optimization and simulation are not considered in this dissertation. From the perspective of sophisticated analysis, KPMG provided an excellent classification of data analytics techniques. Figure 7 shows these groups. This figure also describes the meaning of different levels of analytics. For example, descriptive analytics is done to explain what is currently happening in a system, such as the current state of orders in an organization. Alternatively, predictive analytics can show, what is happening in the system. In this case, predicting the lifetime of a component in a product, or prediction of disturbances in the supply chain are two examples.

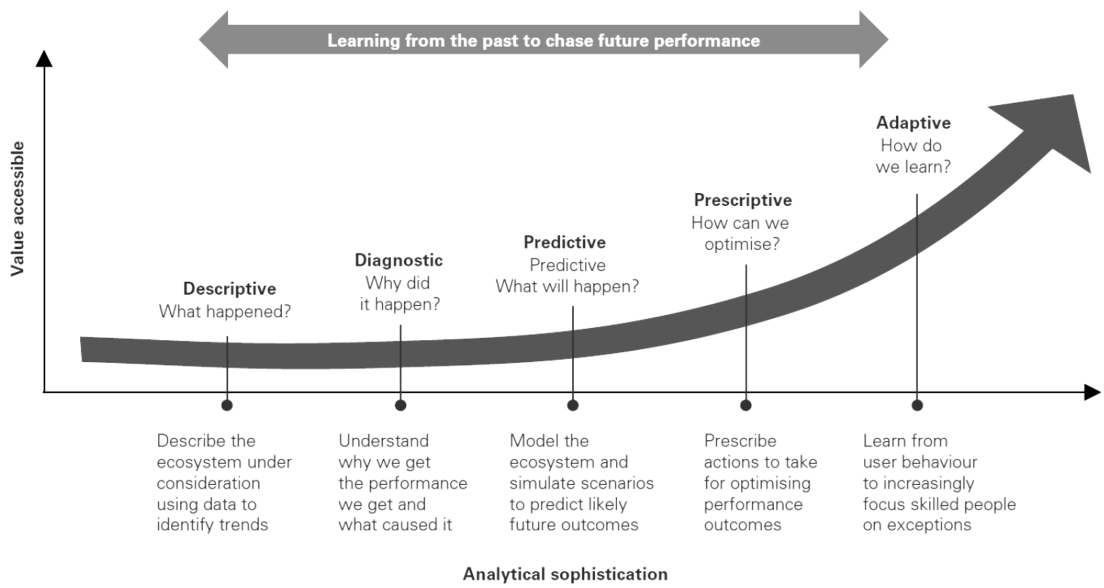

Figure 7: Levels of data analytics ((KPMG, 2017; Delen & Zolbanin, 2018))

Data mining tools are a fundamental category of data analytics. In particular, data mining techniques can perform better than other analytical tools when datasets are complex and have a high volume. Therefore, they can be a good option for modeling PUI sources in our research. In data mining, the raw data are preprocessed; dimensions are reduced and important variables among the big data set are selected. Afterward, the other categories of analytical methods can be applied to data for further analysis.

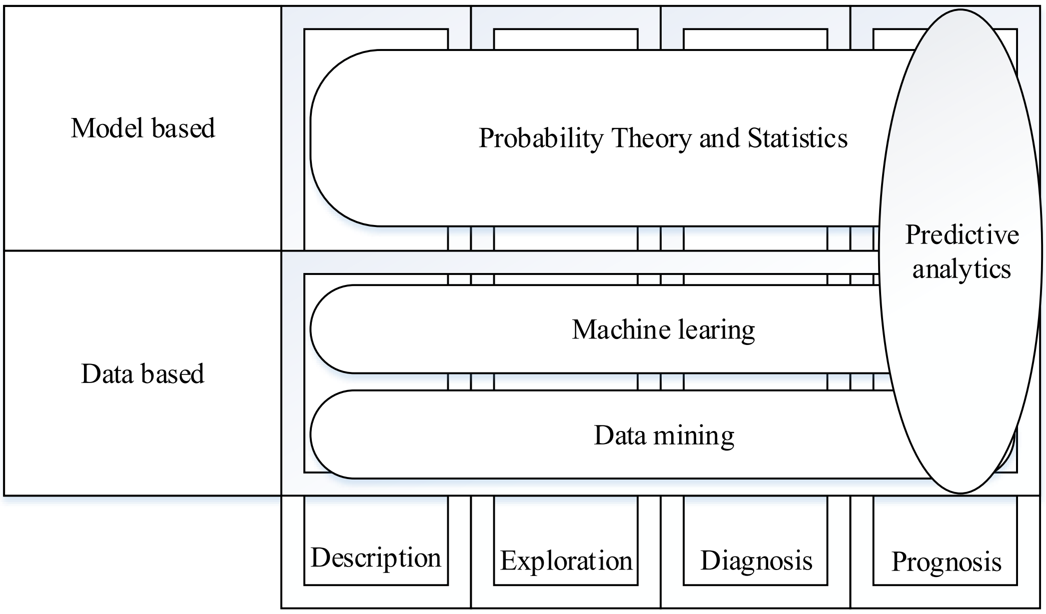

Figure 8: Sciences and disciplines of data analytics (From (Freitag, et al., 2015))

Figure 8 shows the position of data mining (data analytics) in terms of the complexity of the analysis and the type of model it can build. As illustrated in this figure, data mining can perform descriptive and explorative, diagnostic and predictive (prognosis) analyses of tasks. Moreover, it builds the relationship between parameters (only based on data), in contrary to model-based methods, which use expert knowledge or a kind of *a-priori* information for constructing the relationship between variables in a dataset.

Moreover, the four classes of analytics namely, description, exploration, diagnosis, and prognosis are presented in Figure 8. Chapters 5 and 6 uses these classifications and propose a useful set of techniques that can help each MOL stakeholder.

Based on research by Chen et al. (Chen, et al., 2012) the methods of data mining and machine learning can be grouped as shown in Table 4. This table shows currently used as well as emerging methods. In chapter 6, regression-based methods, neural networks, and random forests (Table 4) are applied for analyzing PUI on products such as, electric vehicles, wind turbines and motorboats.

Table 4: Fundamental and emerging approaches to data mining and machine learning in [big] data analytics (adapted from (Chen, et al., 2012))

Fundamental approaches	**Emerging approaches**
• Relational Database Management System • Data Warehousing • Extract, Transform, Load • Online Analytical Processing • Clustering • Regression • Classification • Association Analysis • Anomaly Detection • Neural Networks • Genetic Algorithms • Multivariate Statistical Analysis • Optimization • Heuristic Search	• Statistical Machine Learning • Sequential and Temporal Mining • Spatial Mining • Random Forest • Ensemble Learning • Mining High-Speed Data Streams and Sensor Data • Process Mining • Privacy-Preserving Data Mining • Network Mining • Web Mining • Parallel DBMS • Cloud Computing • Map Reduce

Other data analytic approaches exist, though. Dynamic systems, game theory models, control theory and agent-based simulation, are prominent examples (Kaisler, et al., 2014); but this research won't focus on those groups since nature of PUI data is more similar to the input data in machine learning and data mining methods. Therefore, these methods were selected to serve the purpose of processing data in this dissertation.

3.3.2 Changes in data analytical methods with digitalization

Some of the techniques for data analytics have been used traditionally in the product lifecycle. For example, visualization of the performance of machinery in order to measure key performance indicators or to use regression analysis for forecasting the demand for commodities that a customer is likely to buy. However, these techniques are nowadays improving both from the perspective of prediction accuracy as well as real-time analysis of data. Therefore, on the one hand, there is a need to study the effect of improvements of these tools on the application areas.

On the other hand, the application area of data analytics so far was limited to specific applications and not to a holistic lifecycle view. These separate application areas consist of measures and techniques for the financial, economic analysis, depreciation and modeling behavior of mechanical components. For example, these techniques are exploited to estimate the future value and health state of the assets. However, with the increased amount of data gathered from (smart) engineered products, new tools are also needed to investigate the integration of them in lifecycle data management. For more information, refer to Nabati et al. (Nabati, et al., 2017).

3.3.3 Current applications of data analytics in product lifecycle: State of the art

As mentioned in section 3.2, PUI has characteristics similar to big data. This section provides a review of the literature on the available works showing the current applications of data analytics in the product lifecycle as well as few practices, which were performed in CL-PLM. At the end, it lists the research gaps, which should be covered in the future concerning using data analytics for processing PUI in the product lifecycle.

Most of the research to this date has focused on using data analytics and PUI for improving individual processes from the lifecycle. Individual processes are maintenance, after sale management and user integration in design. Systems, which are developed for prognosing the lifetime of product parts or diagnosing the failures. The use of PUI in order to find the information regarding the product health state is among the major approaches, which currently exist and are built on PUI and data analytics techniques. Research in this regard has been conducted by contributors, such as (Hameed, et al., 2009; García Márquez, et al., 2012; Sheng, 2015; Takoutsing, et al., 2014; Lau, et al., 2012).

In respect of the application of data analytics in maintenance and PLM systems the following research works should be mentioned. Madenas (Madenas, 2014) studied the integration of maintenance data with the product data from PLM systems. They conducted a study on the automotive sector and evaluate it on military defense sector. They developed an architecture, which combines data from the Integrated Vehicle Health Monitoring (IVHM) System and technical maintenance reports with the PLM system. The new environment allows for merging and visualizing a broader range of data (product related- and process related-data). The architecture is driven by case studies. Muller et al. (Muller, et al., 2008) reviewed the concept of e-maintenance. The aspects of knowledge management, capitalization, and data analytics for proactive maintenance are discussed in detail in their paper. Gulledge, et al. (Gulledge, et al., 2010) investigated the possibility of integrating condition-based maintenance to the PLM systems. In a case study, they linked CBM with the PLM by the Oracle fusion middleware. They used interviews, investigated the documents to identify the business processes of CBM, and proposed a conceptual model (Figure 9) of how to link the CBM with the PLM systems. The approach to model the problem is a composite-application-design in that they connect the PLM and combine application concepts, which are then both linked to CBM. They designed and implemented a composite application to automate segments of the CBM to PLM business process.

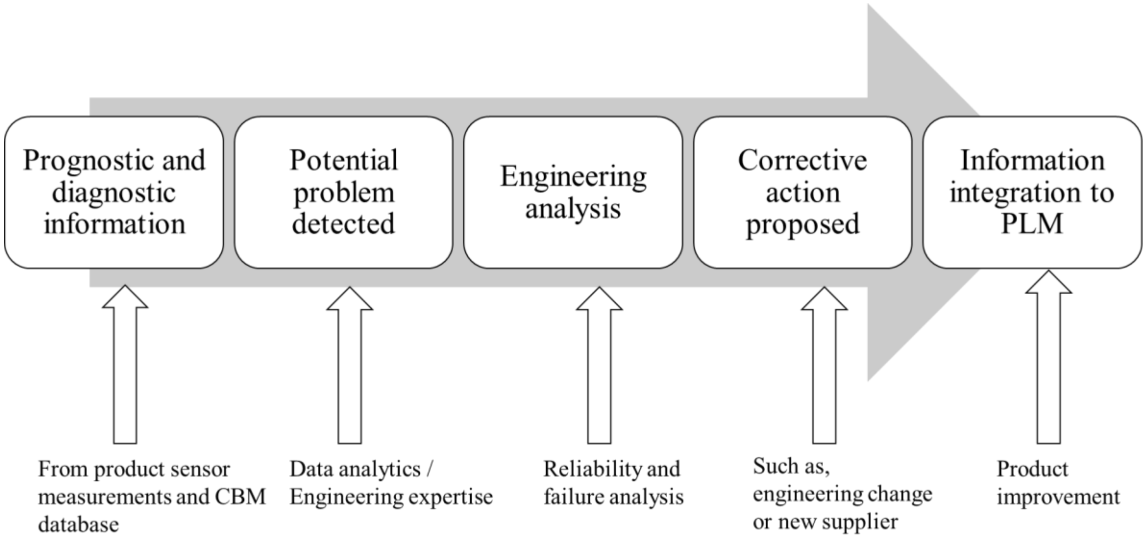

Figure 9: Processes and value chain for linking data of maintenance to PLM (Adopted from (Gulledge, et al., 2010))

Although these studies provide a good insight into the integration of maintenance into PLM, considering CL-PLM, however, the following aspects are not yet covered. They investigated, whether or not data from product or maintenance system (e.g., e-maintenance, CBM), can benefit any other applications or stakeholders in the product lifecycle, but they did not study the effect of smart engineered products on this process.

In the area of CL-PLM, researchers (Lachmayer, et al., 2014; Dienst, et al., 2011; Fathi, et al., 2007) applied data mining to find useful information from product MOL and integrate them to BOL for the aim of improving the design. Gopsill et al. (Gopsill, et al., 2011) provided a strategic view to the future of PLM. They mentioned perspective applications of data mining in PLM. Details of the studies are shown in Table 5. Lachmayer et al. (Lachmayer, et al., 2014) studied the application of smart products in the product development processes. They used data mining and biologically inspired evolution to improve the design of products from one generation to the next. The approach is to use smart components within the product. These intelligent components are called Gentelligent elements and can sense, collect, transfer lifecycle data. The data they transmitted are analyzed by data mining and the knowledge extracted is integrated into an evolutionary algorithm to improve the BOL activities. More relevant research works are cited in Table 5.

A study of the current applications of data analytics in product lifecycle shows that most PUI analysis studies have been carried out only in a small number of lifecycle processes and application areas. Apart from the mentioned studies, there are a very limited number of relevant works available. There is not much relevant research exists which brings together the holistic view of CL-PLM and aims to enhance lifecycle information flows, mainly to serve stakeholders. Therefore, in this dissertation, this issue is comprehensively investigated.

Table 5: Studies regarding the application of data analytics in CL-PLM

Paper name/ authors	Description	Studied aspect (technical approach)
(Rahej, et al., 2006)	Designing a data mining/ fusion-based architecture for maintenance	Data mining/ fusion
(Fathi, et al., 2007)	Condition monitoring of conveyor belt and integrating the results in BOL	Linking sensors and PEID to CM by Bayesian networks
(Dienst, et al., 2010)	Using Bayesian networks for knowledge acquisition for product usage data to improve product's design	Bayesian networks, which is a type of data mining technique
(Dienst, et al., 2011)	Analysis of use data in the hydraulic pump	Feedback assistant system, Bayesian network
(Lachmayer, et al., 2014)	Application of smart products and data analytics in improving the design of the next-generation product	Cyber-physical system
(Gopsill, et al., 2011)	Looking into the future trends in PLM, Knowledge discovery application in PLM, Emerging ICT technologies and their possible implications on PLM.	Understanding data and information across the lifecycle
(Dienst, et al., 2012)	Helping product developers in concept design from customer service and maintenance	Customized text mining from MOL to improve BOL

3.4 Challenges of applying data analytics for enhancing PUI in order to answer stakeholder needs

This chapter showed that, current tools and practices in PLM and CL-PLM systems need to be upgraded, when it comes to overcoming the challenges (section 2.6) regarding holistic management of product lifecycle.

Based on the discussion in this chapter and above-mentioned limitations of current tools (section 3.2), data analytics can be a suitable tool for handling PUI and reaching the objective of better lifecycle management. Moreover, based on the analysis of state of the art, exploitation of data analytics for holistic product lifecycle management has not been realized fully yet. This deficiency has to be addressed.

However, the use of data analytics for PUI in holistic product lifecycle management leads to several technical challenges. The technical obstacles to the realization of this application in practice are as follows. These challenges are adopted from state of the art including (Acatech, 2013; Schoenthaler, 2015; Kolodziejski & Szöllösi-Brenig, 2015).

- Determining who benefits from [big] data
- Defining strategy of processing data based on the field of application
- Obtaining skills and capabilities needed
- Integrating multiple data sources
- Infrastructure and architecture
- Risk and governance issues
- Funding for big data-related initiatives
- Security of data sharing and exchange
- Need for new business models (data-driven business model)
- Lack of knowledge and skill in performing advanced analytics
- Organizational culture
- Reliability of data

It is possible to group the challenges into different topics. The first one contains the challenges in using data analytical modeling techniques. These techniques are still not common practice particularly in most of the organization. As a solution, the awareness of using these models should be increased. There is a need to raise awareness of the potential of models in the broader engineering community and to equip engineers with methods and tools for using appropriate models.

Second, qualified experts can only carry out modeling and simulation. Although qualifications and education in the field of data science and cyber-physical system development have raised in recent years, the number of experts who know this domain is still limited. Using data science and make it part of the culture to use data-driven modeling on decision support in an organization still needs to be developed.

The third group of challenges deals with infrastructure. There is a need for a comprehensive foundation, that can provide not only secure connections and reliable data but also can handle higher volumes of data with more top quality than what exists today, which can be used in a real-time and safe way, with higher communication capabilities (Acatech, 2013).

Regarding the importance of challenges, surveys are available, which show the frequency of problems that the organization face in applying big data technology and analytics (Sivarajah, et al., 2017; Gartner, 2013; Colas, et al., 2014). Based on a study by Gartner (Gartner, 2013) a major challenge in the field of big data is, to get value out of [big] data (Figure 10). To this end, the beneficiaries, who can use the results of data analysis, should be identified.

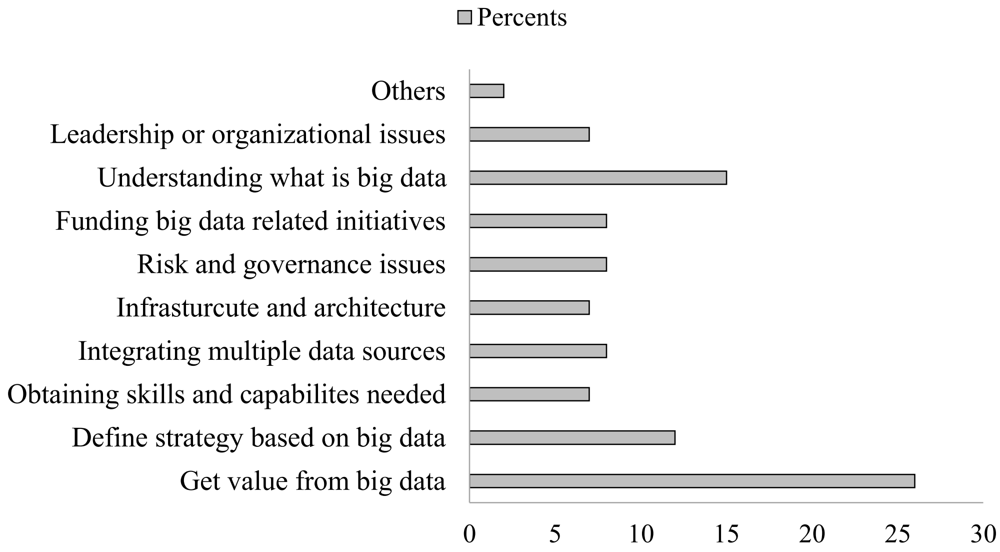

Figure 10: Challenges of using big data in organizations (based on a survey by (Gartner, 2013))

According to the data depicted in Figure 10, around 26% of the organizations face this challenge (get value out of data). The figure reveals it as the most crucial challenge. Therefore, for utilizing PUI data, one must understand who needs and can benefit from these data. Based on this need and gaps in research identified in this chapter, this research focuses on the gap to identify beneficiaries of PUI sources. Thus, chapter 4 investigates and identifies the stakeholders, who can benefit from PUI. Moreover, their information needs are discussed, and big data analytics is applied to provide them with the data they need.

In summary, the following can be stated based on the challenges described in chapter 2 and the identified research gap in this chapter. Referring to chapter 2 (section 2.6), utilizing emerging technologies and new information sources in the product lifecycle is a major challenge for organizations for getting value-added information from the lifecycle. It can be concluded from the challenges of using the product's operational information that providing added value to the information users in the product lifecycle is the most essential task in this context. This need should be addressed. Therefore, this dissertation concerns the identification of data and information needs of product stakeholders, who are the primary users or beneficiaries of the product. After identification of data and information needs, it uses data analytics techniques to process PUI and make the relevant information available to each stakeholder.

4 Concept development

Based on state of the art, as stated in chapter 2, the impact of digitalization on the stakeholders of the product lifecycle is continuously increasing. There are different types of stakeholders, who needs the data (or information) collected from operations of the product. As mentioned in chapter 2, companies have to develop better collaboration and to adopt new strategies, such as holistic management of the product lifecycle, to overcome technological adaptation issues of a competitive market, and environmental damage of products. In this way, using the currently available lifecycle data to gain inbsights from product lifecycle can be a cost-effective strategy. The advantage of using this strategy is that the organizations do not have to pay for the data. Furthermore, data from lifecycle, especially PUI, have the potential to provide new actionable knowledge for the organizations or provide more transparency in lifecycle processes.

To put holistic management of product lifecycle into practice, one of the solutions mentioned is to complete the concept of CL-PLM by providing access of stakeholder organizations to appropriate information sources from MOL. To achieve this goal, this chapter conducts the modeling and concept development for the provision of the right data (from PUI) to organizations. Therefore, this dissertation first identifies organizations who are beneficiaries of the product lifecycle (in MOL). Next, this research investigates the kind of information requirements they have from PUI sources. This chapter is organized as follows. Section 4.1 gives a summary of motivation and the need for developing the concept of “lifecycle stakeholders’ information needs”. Section 4.2 explains the overall methodology of research in this dissertation. Later, section 4.3 focuses on the methodology for the development of the concept. Section 4.4 starts with the first steps (as stated in 4.3) of developing the concept. Identifying MOL stakeholders based on two real-world studies are addressed in this section. Section 4.5 reports the discovered information requirements (needs) from PUI sources for identified stakeholders in section 4.4. Section 4.6 presents a summary of the concept. It gives a generic image of concept, which shows information needs and its implementation for linking MOL to CL-PLM. In the end, the use of data analytics tools to realize the information needs is discussed, which paves the way for chapter 5.

4.1 Rationale of concept [1]

This section presents the need and the reason for the concept of this dissertation which is called “lifecycle stakeholders’ information needs”. The arguments in this section address the importance of analyzing information needs of stakeholders (as

[1] Parts of the text in this chapter has been published in (Nabati et al., 2017)

mentioned in chapter 2) and the challenges of using PUI sources (as mentioned in chapter 3).

The importance of identifying MOL stakeholders and their data needs (as stated in section 2.5) is that they are vital elements of product MOL. Understanding them is necessary for completing a virtual representation of CL-PLM, which can facilitate the management of the entire lifecycle of an engineered product.

As stated in the previous chapters, one way to enable complete product lifecycle management is to facilitate the flow of information from MOL to BOL, or improve information flows inside MOL (refer to Figure 3 for a visual illustration of primary information flows). One possibility to boost information flows is to integrate and make good use of PUI sources. This research shows in section 3.2 that PUI cannot be exploited and embedded directly in the CL-PLM because PUI has a huge volume and consist of complex format and context. Therefore, using PUI data sources for serving information needs of product lifecycle organizations is only possible if the data are processed with right tools such as data mining (data analytics) and they are turned to output information with higher quality than raw data as well as less volume. Furthermore, it is necessary for this output information to have a relevant context to the needs of information user. Therefore, identifying and understanding the organizations that can benefit from PUI sources are of substantial importance.

Current advances in digitalization strengthen the potential to create value out of PUI. For example, if a wind turbine has erosion in the blade edges, this turbine could send the information that erosion has happened on the sides of a turbine blade. As a result of an analysis of information relating to erosion, operator of the turbine can receive the information, which could, for instance, reveal that under this level of decay on the blade, 20% less energy will be produced by a turbine in the next month. Thus, a decrease in the performance of the turbine becomes measurable. This piece of quantifiable information, which reveals a reduction in the performance of the turbine, can provide a practical insight for the turbine's operator. Having this information, the operator can plan for servicing the blades and contact the MRO service provider at the right time. Moreover, if this procedure is automatized (flow of information is automatized) and this information is also shared with the MRO provider and manufacturer, it is possible to plan simultaneously for maintenance process (real-time planning). That is, automatically ordering of needed material and spare parts as well as teaming up a maintenance group, which are trained for repairing the turbine blades. In this way the operator doesn't need to care for contacting service provider; MRO provider and manufacturer is automatically noticed about the need for maintenance service.

This example illustrates still an ideal state in many engineered products. These functions are either not fully realized in most of the current versions of engineered products, or they are not yet designed in a way that the product users can fully benefit from it. There is a need for more research and investigation. To this end, a concept is presented in this chapter that addresses the realization of these situations. Furthermore, chapter 6 deals with the potential of implementing scenarios as mentioned in the example above.

Another reason that contributes to rationality and necessity of the model in this dissertation is that in the past the control of product lifecycle as a collaborative activity was not of great importance and the focus of product lifecycle was only on the manufacturer. Now, by utilizing the IoT and other new technologies, there is a need for research on a linked view, which connects processes, stakeholders and products. In addition, it is vital to analyze the future of these processes after realization of digitization in the product lifecycle. In this respect, several areas in which PUI can contribute to improving product lifecycle processes or decision scenarios should be recognized.

To this end, based on the two major reasons mentioned above, which made the use of PUI more economical and practical, and extending the lifecycle management by considering the several stakeholders of lifecycle, and their varied needs, it makes sense to identify stakeholders, roles, their collaborative relations, as well as identify their future information needs from PUI sources. Therefore, the concept of "lifecycle stakeholders' data and information needs" meets this objective. In the following steps to realization of the concept are delineated.

4.2 Overall research methodology of the next chapters

Table 6: Overview of methods in each chapter

Chapters	Methods used	Description
Chapter 4	Inductive approach (qualitative- exploratory) Freeman's method	Concept development for understanding stakeholders in MOL and their data needs, two case studies
Chapter 5	Survey and experiment (qualitative approach)	Concept development for suitable data analytics per each type of data need
Chapter 6	Data mining (quantitative approach)	Implementation of concept on multiple use cases
Chapter 7	Operational validation (qualitative)	Validation of concept based on implementation on multiple areas
Chapter 7	Cross-validation (quantitative approach)	Validation of data mining models for each use case

Section 4.1 motivated the need of elaborating on the data from product MOL (PUI sources) and support the product stakeholders. To identify the right information

needs of product stakeholders, this dissertation suggests a concept of identifying stakeholders and the information needs, supporting the realization of information needs with data analytics tools and techniques and test the implementation of data analytics for getting value out of the PUI. This dissertation follows the commonly accepted flow of concept development, concept implementation and validation as the overall methodology. Table 6 shows a list of chapters and the flow for execution of this work.

The research approach for developing the concept in chapter 4 and 5 is the inductive approach. The inductive approach starts with detailed observations of the world, which moves towards more abstract generalizations and ideas; as stated by Neuman (Neuman, 2003). Similarly, in this dissertation, in order to apply the inductive approach, two studies on MOL of engineered products are analyzed thoroughly and in detail and based on the results of this analysis a general theory (or here called concept) of existing stakeholders in MOL of engineered products and their variety is proposed. The analysis takes into account multiple perspectives; such as, characteristics of CL-PLM stakeholders and their roles, available types of data in MOL of engineered products, as well as variety of processes in MOL of engineered products. Moreover, a method called Freeman's stakeholder analysis approach is used for analyzing the stakeholders and their data needs. More information in this regard is provided in 4.3.1.

Some additional characteristics of inductive research that used in this dissertation are listed in the following (Table 7). In this dissertation is conducted as an exploratory approach, namely, to explore and understand mechanisms in MOL, stakeholders and information needs and exchange. The concept is created in a bottom-up direction, as, the two cases and detail of processes, stakeholders and information needs are analyzed. Based on the identified commonalities, the concept is created. The focus is to understand stakeholders' types, their roles, behaviors, data and information required from the engineered products. Furthermore, the data intensity of this research is very high, similar to typical inductive research works. In our case, the high intensity of data is due to several groups of stakeholders, diverse roles, and several information needs (Table 7).

Table 7: Attributes of inductive research (from (Dudovskiy, 2018))

Attributes of inductive research	Description
Direction	Bottom-up
Focus	Understanding dynamics, behavior, processes, practices and their alterations in future
Data intensity	High

4.3 Methodology for the concept development and implementation

To build the concept of stakeholders’ data and information needs several parts exist. The methodology (parts) of concept development and implementation is illustrated in Figure 11.

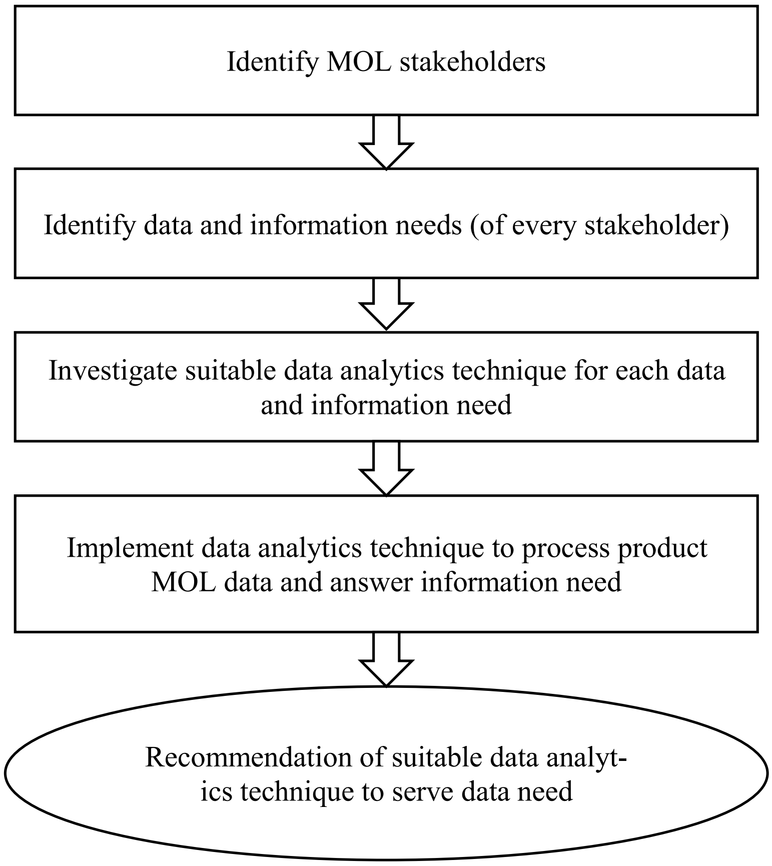

Figure 11: Methodology of concept development

First, the characteristics of stakeholders in CL-PLM who are affected by MOL data sources should be identified. Section 4.4 focuses on this part. For identification of stakeholders in MOL of engineered products, two studies are performed. Two appropriate industries for technical products are selected to identify stakeholders. One study is carried out on MOL of wind turbines while the other addresses MOL of passenger airplanes. More information about methods used in stakeholder identification is provided in 4.3.1 and 4.3.2.

In the next part (second step in Figure 11), major data and information needs of identified stakeholders based on the decisions they make and characteristics of PUI is defined. The results of this study is reported in 4.5. Moreover, in this step the information needs, which can be supported by data analytics, are identified.

For further completion of the concept, suitable data analytics technique which can be applied to facilitate reaching the information need is proposed (part three of Figure 11). Chapter 5 is dedicated to this analysis. Next, chapter 6 shows the implementation of the concept through three use cases. Based on the results of chapter 5 and 6 recommendations on suitable data analytics technique to serve data and information needs of stakeholders are provided. The concept and implementation scenarios are validated in chapter 7. Next subsection describes the methods for stakeholder identification.

4.3.1 Approach for analysis and identification of stakeholders

In order to identify and analyze lifecycle stakeholders in the CL-PLM, it is essential to apply a suitable method of stakeholder analysis. Review and comparison of different approaches for stakeholder identification can be found in (Nabati et al. 2017) and it is not reported here. Based on the review of possible approaches, Freeman's method can be the best matching method for the identification of stakeholders in MOL. The importance and suitability of Freeman's method for identifying stakeholders of MOL rise from the following facts. First, this technique can be applied to identify larger groups of stakeholders. MOL is a long-term phase and several groups of stakeholders at this stage deal with the product and its services. Second, the suitability of Freeman's method for modeling stakeholders in the product lifecycle has already been proven (Wuest, et al., 2014; Bischof, 2012; Nilsson & Fagerström, 2006).

Freeman's theory can be explained as "a structured approach, which identifies the stakeholders and sorts them according to their characteristics such as their impact on a specific action, interest, position, and power on a project or task" (Makan, et al., 2015). Zell (Zell, 2007) formulated the concepts in Freeman's stakeholder analysis in four steps; stakeholder identification, concern, interests and power assessment, stakeholder behavior identification and action planning (Zell, 2007). The steps are illustrated in Figure 12. Hereafter we call these steps as "Freemans 4-step method". In this research, we used these four steps to analyze MOL stakeholders because they can be applied easily in practice.

Using Freeman's 4-step method, it is possible to provide a comprehensive and in-depth view of the potential stakeholders of MOL. For more information regarding the description of each step, the reader is referred to (Bischof, 2012). For information regarding the application of these four steps in this dissertation, please see next subsection, 4.3.2.

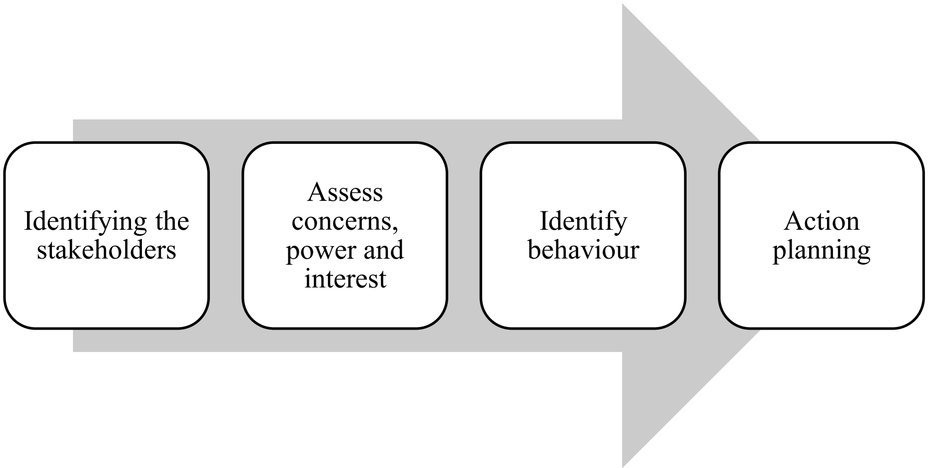

Figure 12: Stakeholder analysis Freeman’s 4-step method (retrieved from (Bischof, 2012))

4.3.2 Applying Freeman’s method for stakeholder modeling

Figure 13 shows the application of Freeman’s method for identifying and analyzing the stakeholders of MOL. Identifying and analyzing the stakeholders of MOL who can benefit from MOL (PUI) data are named as “stakeholder modeling” in this dissertation. As shown in subsection 4.3.1, Freeman’s 4-step method can be suitable for identifying MOL stakeholders. Therefore, this subsection provides more information about the adoption of Freeman’s method for applicability on product lifecycle (specifically, for the analysis of stakeholders of MOL of engineered products).

Identifying the stakeholders is the first step of Freeman’s 4-step method (see Figure 12). As mentioned before, this dissertation conducts two case studies, on commercial airplanes and wind turbines, for identifying the MOL stakeholders (see Figure 13). In each case study, several processes, such as maintenance and operation, as well as occupations of stakeholders and their roles are carefully studied. The result of this study is the identification of stakeholders involved in the process together with functions they perform.

After identifying the stakeholder groups for each industrial case, the stakeholders are grouped into major MOL stakeholder based on characteristics of CL-PLM stakeholders and their roles as well as available types of data they use from MOL of these products. Survey and interviews were used to group stakeholders and to learn about their roles in MOL of each product. Moreover, interviews from different stakeholders involved in MOL of commercial airplanes and wind turbines are conducted to get a better understanding of their roles in the product lifecycle. During the interviews, the interviewee (stakeholders) were asked about: (a) daily task they perform; (b) their jobs and responsibilities; (c) the information related to products, which they need for completing the tasks. All these tasks were carried out for identifying stakeholders (step one of Freeman’s method). Step two of Freeman’s 4-step method (Freeman, 2010) is applied here to prioritize the stakeholder groups (see Figure 13).

In this respect, the stakeholders are prioritized based on their power, interest or influence on the product. Detail results of applying step two of Freeman's method is reported in the appendix B (section 10.2). Moreover, the interested reader can refer to (Wuest, 2014, p. 113) for more information. A similar method is applied in this reference for prioritizing stakeholders.

A comparative study is done after acquiring the list of major stakeholder groups for each industrial study (see Figure 13). The comparative study examines the gained results of two case studies. This analysis revealed commonalities between stakeholders of two products. Therefore, similar stakeholders are grouped together. Consequently, the potential MOL stakeholders, who can benefit from product MOL data and information, were determined based on this comparison (step 3 of Freeman's method, Figure 13). The relation of MOL stakeholders, which was gained in step 3, is depicted with system analysis diagrams at the end.

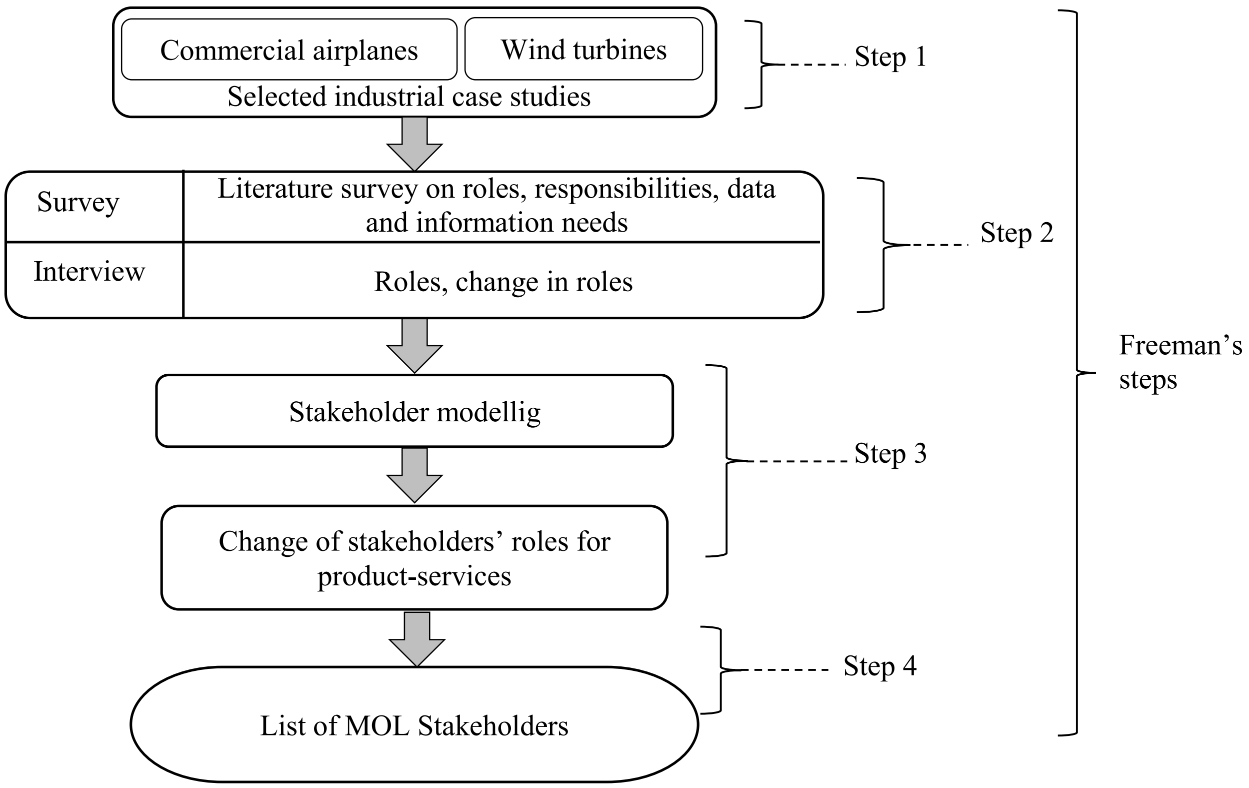

Figure 13: Application of Freeman's method for MOL stakeholder and information need analysis (referring to Figure 12)

To summarize, section 4.3 presented the methodology of research. Next sections start the development of the concept (part identification of product stakeholders) by analysing the characteristics of product lifecycle stakeholders.

4.4 Identifying MOL stakeholders

The industrial case studies are performed by looking into (examining) underlying MOL processes of wind turbines and commercial airplanes. To this end, this section describes the MOL schema for wind turbines and commercial airplanes. By analyzing the MOL processes and asking the people who work in these areas, different groups of stakeholders are identified at the end of each case study. The information is collected from state of the art and interviews with the experts.

4.4.1 Characteristics of stakeholders in product lifecycle and data needs from product MOL[2]

In order to develop a concept of identifying the stakeholders and their requirements from PUI, one has to learn about the characteristic of stakeholders in the lifecycle. Who can be considered as a stakeholder in the product lifecycle, which aspects of their roles have an impact on the lifecycle, as well as which characteristics do they have, or which features will change in future?

This subsection identifies the characteristics of lifecycle stakeholders and reports the factors, which affect their data needs. So far, product stakeholders' identification is well studied for product development and design process of CL-PLM. This is evident from an analysis of the existing literature (see Table 8).

Table 8 presents a broad spectrum of CL-PLM stakeholders. Stakeholders are classified in the table according to phases of product lifecycle and aspects of understudy in state of the art. Applied processes of the product lifecycle are product design, operation and maintenance, and reverse logistics.

Identifying stakeholders in the product lifecycle is not so easy and has several aspects. In Table 8 some of the most critical aspects are presented, which are explored by reviewing the state of the art papers. These aspects consist of: (1) internal and external stakeholder groups, (2) variety of users and various types of needs, (3) conflicting requirements, (4) analysis of power and influence of stakeholders, classifying by power or influence, and finally (5) smart products for CL-PLM stakeholders.

[2] Parts of the text in this chapter has been published in (Nabati et al., 2017)

Table 8: Research on stakeholder identification in the lifecycle of a product

Reference	Stakeholders of the Product lifecycle	lifecycle process/ phase	Studied aspect
(Majava & Haapasalo, 2015)	Software product management, research and development (R&D), management, marketing, sales, product team, standardization and regulatory bodies, subcontractors, competitors, suppliers and technology vendors.	Product design (BOL)	Conflicting requirements; Internal and external stakeholders; Identification based on product /component
(Nilsson & Fagerström, 2006)	• External stakeholders: users/ customers, distributors, governments, suppliers, communities, laws and regulations • Internal stakeholders: management, marketing experts, designers, purchasing, manufacturing, assembly and sales.	Product design (BOL)	Various users and stakeholders; Internal and external groups
(Anitha & Prabhu, 2012)	Customers, end-users, product management, sales and marketing, regulatory bodies, competitors and product development team	Product design (BOL)	
(Greiner, et al., 2015)	Waste transporter, waste management authority, waste disposal centre, bank/ investor, operator, operating company, police, CMS-analyser, third-party maintenance provider, certificate body/ auditor, logistic company, marine warranty surveyor, subcontractor of manufacturer, recruitment agencies, rescue coordination centre, training and educational institutes, transportation companies, transmission system operator, traffic management centre, insurance, weather forecasting, wind assessment institute, customs, suppliers of spare part.	Operation and maintenance (MOL)	Stakeholder priority analysis and classification based on influence and concerns on operation and maintenance process
(Wuest, et al., 2014)	OEM (designer, manufacturer), user/ owner, supplier, platform, boatyard owner, maintenance, insurance, the retailer	Product design improvement (BOL)	Power, influence, and interest analysis; Smart products
(Hribernik, et al., 2011)	Product owner, product user, executive officer of plant, operator, worker at plant, vendor of plant operator, executive officer of logistics company, worker at logistics company, manufacturer, distributor, municipality, legislator	Reverse logistics (EOL)	Identification based on product and component
(Air transport action group, 2011)	Government, tourism, airlines and aircraft operators, airports, manufacturers, airline suppliers, sales channels, cross-industry representative groups, aircraft financing, unions, education and training, media, business community, non-governmental organizations	Transport operation (MOL)	

Literature review discusses the criteria for identifying stakeholders as follows. Firstly, according to Nilsson & Fagerström (Nilsson & Fagerström, 2006), the internal stakeholder group includes stakeholders who work in the organization. This group is like employees in manufacturing companies. They hold various positions including the product manager, product designer, marketing manager, purchase, and sales manager. In contrast to the internal group, the external group comprises of stakeholders outside the organization. Such stakeholders include competitors, regulatory bodies, customers and suppliers (Majava & Haapasalo, 2015; Nilsson & Fagerström, 2006).

Regarding the frequency, with which the stakeholder groups and these requirements have been addressed, state of the art shows that manufacturer, supply chain, designer and product users have the highest frequency of investigation.

Some studies, e.g. (Sprague, et al., 1991; Korpi & Ala-Risku, 2008; Schuh, et al., 2008) addressed the supply chain and the needs of stakeholders for cooperation. Researchers (Anitha & Prabhu, 2012; Majava & Haapasalo, 2015) studied product design stakeholders. Another group of stakeholders that has gained good attention so far is the product users. Their wishes and needs are determining factors in the life of a product. In addition, the way they use the product affects product lifetime. Several studies were carried out to find the requirements of product users. Customer requirements under the concept of PLM have been mainly studied in the area of product design and development. See (Hadaya & Marchildon, 2012; Lee, et al., 2008; Schulte, 2008) for more information.

Although product users constitute a significant group of beneficiaries from MOL, however, when managing the product lifecycle from a holistic point of view, such as that offered in CL-PLM, other organizations like, maintenance providers and product recyclers also influence the product lifecycle. These groups of stakeholders of MOL are unknown and their needs, wishes, and behavior have not been studied yet.

Secondly, several users deal with the product lifecycle, especially when MOL of a particular product is very long. This is common practice for engineered products such as passenger airplanes, which may be operated by airlines in different countries for many years.

Thirdly and besides various users and stakeholders, another aspect concerning stakeholders' analysis is the conflicting requirements. For example, product's operation can be beneficial and provide more jobs for society; at the same time, it can cause pollution and environmental concerns. In this respect, regulatory bodies raise conflicting requirements with the interests of the public. Fourthly, analysis of power, interest, and influence of the stakeholders is necessary for finding the requirements of the stakeholders. They together refer to an extent to which a particular stakeholder has the power and authority to change product or product situation. This type of analysis is reported in Greiner et al. (Greiner, et al., 2015) for the product's operation phase. Nilsson & Fagerström (Nilsson & Fagerström, 2006) add that stakeholders act according to their interest. They use their power to influence the product in the direction they desire. Therefore, to manage stakeholders, special attention has to be paid to the stakeholders who have more power or are influenced more (Bryson, 2003; Sharp, et al., 1999; Nilsson & Fagerström, 2006).

Fourthly, stakeholder identification in the literature has been carried out based on identifying the actors involved in processes, projects or users of product components (Hribernik, et al., 2011; Majava & Haapasalo, 2015; Ford, et al., 2015). In this case, before identifying the concerned stakeholders, the processes, in which the stakeholders have been involved, should be identified. Examples of this type of process identification have been reported in Hribernik et al. (Hribernik, et al., 2011). In addition, stakeholder identification has also been reported for a specific project (Schmeer, 1999; Collinge, 2011). Examples are constructing a hospital or power plant. Finally, some researchers consider the stakeholders as the users of subsystems or functions of a product. For example, (Anitha & Prabhu, 2012) identified the relevant beneficiaries for developing software.

Finally, the last and fifth aspect discussed in the literature concerning stakeholder identification concerns the smart products and increases in dynamics of stakeholder data needs. In this case, the stakeholders who can benefit from a smart product or data produced by the smart products or virtual models of a product has been studied. Characteristics of this group include: They are supported by various information systems; they should be trained to know, how to use information systems, virtual twins and be able to interpret output they get from IT-based systems. Moreover, they should have multidisciplinary expertise.

Based on the synthesis provided in this section, it can be established that data needs of stakeholders from PUI data are a function of (is affected from) the roles of stakeholders, the processes they are involved in, increase in the smartness of products and their interests and influences.

The analysis of literature showed that a holistic framework or model, which shows all the stakeholders of MOL, is not yet available. Most previous research considered the identification of stakeholder requirements from BOL and EOL.

In MOL, apart from product users, requirements of other MOL stakeholders have not yet been identified and well-studied. Therefore, this dissertation focuses on stakeholders of MOL, conducts stakeholder analysis and identifies the requirements of different MOL stakeholder groups. This makes it possible to model the information requirements of stakeholders in their cooperation with each other. The issue of identifying collaborative information needs is very important in practice. However, as already mentioned, it has not been thoroughly investigated so far.

4.4.2 Study on MOL for off-shore wind turbines

MOL for off-shore wind turbines can be considered from the time the wind turbines are installed in the sea. Stakeholders involved in the installation processes are turbine manufacturers, shipping, logistics and electric power companies, as well as authorities. The pre-assembly process is partly done by the manufacturer on land (port area) mainly for components such as nacelle and hub. Afterward, the components are loaded on the jack-up vessel and transported to the sea by a shipping and logistics company. Upon arrival at a targeted site, the installation team from the manufacturer in cooperation with logistics provider performs the offshore assembly process. For the installation, a foundation is fixed in the seabed followed by tower, nacelle, hub, and blades. After a wind turbine is installed, the next task is to conduct a pilot start-up.

The pilot start-up runs different tests on the windmills and energy network. The tests are meant to ensure the performance, as well as the safety of the wind park. The tests can be partly performed by a certified body. Tests comprise of examining the availability, power curve, electrical system, acoustic and noise level. Other stakeholder groups, who are involved can be manufacturer and operator. The manufacturer (or owner’s representative) has to certify the product before handing it over to the operator, i.e., Factory Acceptance Tests (FAT). Afterward, the wind park is put into operation to generate electricity while the function and performance of the wind turbine are monitored on a timely basis. This is done to ensure that turbines can achieve maximum power generation.

The operation phase consists of generating electricity, planning and monitoring the equipment. During this time, several stakeholders are in direct and indirect contact with the product; for example, operator, government, NGOs, provider of services and investors.

During the first five years of wind turbine’s operation, Operator Company has contracted with the manufacturer for the accomplishment of warranty service. Therefore, the manufacturer performs maintenance in coordination with the operating company. In addition to manufacturer, several other stakeholders are involved in the maintenance and repair process, such as Logistics Company, customs, and shipping authorities (because the turbine is located in the sea). Similar to other engineered products, third-party Maintenance Repair Overhaul (MRO) provider and supplier of spare parts can also be other actors involved in maintenance. MRO provider or manufacturer does the overhaul on a timely basis when a need arises (Klinke & Klarmann, 2014; Nabati, et al., 2017). Figure 14 shows the major MOL process of offshore wind.

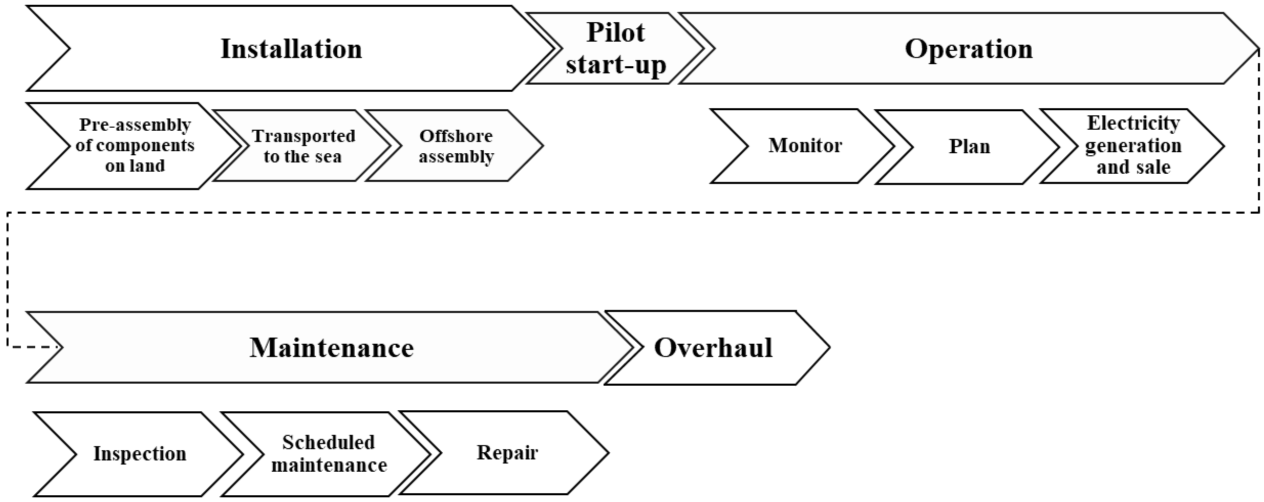

Figure 14: MOL processes of offshore wind turbines

From the perspective of sensor-based smart products, one can perform several types of measurements on an off-shore windmill.

These measurements consist of the following functional and environmental parameters; (1) temperature of the rotor, oil, and windings; (2) acceleration of the tower sway, turbine shroud, and gearbox monitoring; (3) vibration; (4) speed of the wind, rotor and blades; (5) fluid property of gearbox and oil; (6) position of power levelling, and (7) air pressure blades.

With respect to these parameters, data sensed can be useful for decision makers in remote monitoring the conditions of equipment. The data help in early detection of faults and accordingly plan for maintenance.

Other MOL data source for wind turbines consists of followings:

- Databases for business and in the companies such as ERP, CRM, SCM, TPS (Include historical data from the processes)
- GIS and positioning (position of equipment, vehicles, helicopters and ships, tracking)
- Weather condition
- Application and services data (on mobile devices and web)
- Maintenance plans
- Historical reports of failure and maintenance (product service logs)
- Demand for spare part

Nevertheless, the data if aggregated with other organizational information sources can help to enhance lifecycle activities, such as production scheduling and supply of spare parts.

Back to the aim of this dissertation, the following presents the analysis of stakeholders of MOL for windmills, in the above-mentioned scenario.

Resulting from the analysis of the aforementioned wind turbines' MOL processes and interviews with experts in this field, this study identifies eighteen stakeholder groups (Table 9). To illustrate a few cases; the OEM has the role of providing warranty service accomplished by involving internal stakeholders such as marketing, after-sales service, and production department. The table also lists the internal stakeholders for every stakeholder group. For instance, OEM has internal stakeholders such as product designer, marketing and aftersales. These internal stakeholders need information from MOL of the wind turbine to perform their tasks and make decisions (run processes).

In addition to OEM, the operating company is another MOL stakeholder of wind turbines. Its role is to run the wind farm and perform the daily operations. They monitor the electricity production and the function of the wind turbines. The operating company's internal stakeholders include energy trading manager, electricity transmission operator, maintenance scheduler and control room operators.

Table 9: Stakeholders of MOL for wind turbines

Stakeholder group	**Role**	**Typical internal stakeholders in this group**
Manufacturer (OEM)	Warranty service, component supply, assembly off-shore	Product designer, marketing, aftersales service, production department
Operating company (maintenance and service provider)	Day to day operation of the wind park, electricity generation, service and maintenance	Energy trading manager, financial manager, electricity transmission system operators, CMS analyst (Wind Energy Hamburg, 2016), scheduler, control room operator, technicians (Oelker, et al., 2015)
Government/ legal agencies	Setting rules, monitoring, financing (Schaar & sherry, 2010)	Waterways and shipping directorate (Klinke & Klarmann, 2014), ministry of environment and energy
Logistic company	Provide the transport mode to the off-shore site	Transport provider to offshore, transport providers inland
Utility and grid operation	Power generation, transfer, and distribution	Electricity trading, grind operation, utility
Suppliers of components and subsystem	Distribution and providing the material and spare parts	Mechanical/ hydraulic/ electrical/ tower/ foundation suppliers, industrial marketing

Stakeholder group	Role	Typical internal stakeholders in this group
Suppliers of raw material, composites, and material processing	Providing raw material for making the spare part components, material used during maintenance and service	Retailer, delivery, mines, casting companies
Wind turbine installation companies	Setup and assembly inland and on the sea	Logistics provider (transport vessel, pickup, cranes), assembly provider, helicopter, port facility and technology
Storage	Storage of components or spare parts, storage and transfer of electricity	Port yard, warehouse management (Wind Energy Hamburg, 2016)
Planning, consulting and project development	Consulting, coordination in areas of inspection, maintenance or operation	Engineering office, consulting, surveyors, data analytical system developers, offshore exploration and service, offshore accommodation (Wind Energy Hamburg, 2016)
Insurance company	Safeguard the setup of the wind park. Especially installation of cables, montage of the components and logistics to the sea for the personnel and the wind turbine components	Energy insurance bodies, property insurance section, personnel insurance section, finance insurance section, business interruption insurance (Allianz, 2016)
NGOs	Voting for start/stop of an offshore wind park construction	Wind energy associations, environmental associations, volunteers
Certification body and classification society	Inspection and certification-fit for purpose and safe to use design evaluation, site condition evaluation structure analysis, risk analysis, collision analysis, property assessment and maintenance audit	Technical inspection associations such as DNV GL, TÜV (The renewable energy hub, 2016) and members of Renewable Energy Assurance Ltd
Education and research	The companies who do research and provide training regarding wind energy	Universities, education and research institutes
Organizations	Companies who do business with the wind sector, or offer services related to wind turbines (wind industry)	Weather forecasting company, job finding agency (Human resource), media, associations
Investor	Bond or putting money to start or operate the wind farm	Financial institutions, banks, government (Schaar & Sherry, 2010)
Competitors	Organizations, which offer other products for producing electric energy	Fossil fuel, water force or nuclear power plants
Electricity users	The users of the electricity produced by the wind turbines	Industries, the general public

In summary, in this study, stakeholders are grouped based on their common roles and responsibility that they have. Although other stakeholders may exist, those identified in Table 9 are the major responsible ones. Sometimes the owner is counted as a stakeholder. However, in this analysis, the owner is not considered as an independent stakeholder. This is because the owner can be the investor or even operator, based on the financial agreement on the build and operation of the turbine.

4.4.3 Study on MOL of commercial airplane

Although airplanes differ from wind turbines, there are similarities between them in terms of MOL stakeholders. For this reason, this subsection describes MOL of commercial airplane processes to identify involved stakeholders. The description is partly adapted from previous literature (Schaar & Sherry, 2010; Air transport action group, 2011). MOL for commercial airplanes can be started when the aircraft goes for testing. Customer visits the manufacturer to audit the airplane and find out the results of testing. By passing tests, the airplane is recognized as airworthy. Next, the airplane is registered for the flag and route, which is going to be assigned to by the relevant authorities. Operation of the plane in the selected direction begins by the time the official procedures are finished. Figure 15 shows the major processes in MOL of the airplane.

The manufacturer may perform the maintenance repair process, especially within a guarantee period. Different airplane components are serviced after a specific duration of the flight (per hour flight). Likewise, the annual maintenance service and inspection are conducted. Although sudden failures may be fixed at the airport, the more serious failures are communicated to the airline maintenance team or third-party MRO Company. The overhaul process is performed for major components such as the engine when their performance is decreased or to prolong the time of the component. For such cases, the manufacturer conducts main overhauls. In some cases, the overhaul is performed by the airline, if the airline is big enough and has this capability. In the case that the airplane is sold to another user at the end of its useful lifetime, engine regeneration, replacement, and main components replacement are carried out. Furthermore, technical inspections are performed frequently during the operation of the plane. Mainly, technical inspection of safety systems and the components are performed before each flight. In this test, the functional health of the components of the airplane is checked and registered in a checklist while the cabin is thoroughly controlled by the crew before the flight. Although airplanes are considered to be mostly in the air when it comes to operation process, a main and comprehensive action is performed on the ground (ground handling process) (IATA, 2017; Ashford, et al., 1996). Aircraft ground handling process consists of passenger services, cabin service, catering and field operation service (Ashford, et al., 1996). The flight scheduling, control tower operation and airport ground activity planning are very information- intensive tasks among the ground activities. In terms of commercial airplane services, there are many entities involved in selling the services. Entities like the travel agencies, logistics companies, and tourism agencies provide services about commercial airplanes. Figure 15 depicts major processes in MOL of a commercial airplane.

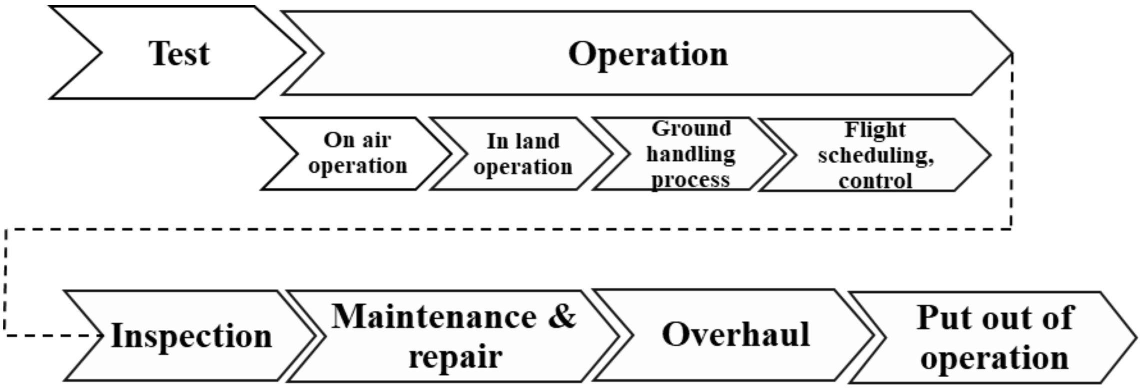

Figure 15: MOL processes of a commercial airplane

As commercial airplanes are considerable smart products, they are equipped with several sensors and monitoring technologies. Therefore, they send terabytes of data per flight. According to Marr (Marr, 2015), for example, the new models of Airbus A380, are approximately embedded with 10,000 sensors on each wing. As a result, there are many data sources (Marr, 2015). Some of the data sources from these products are listed below:

- Environmental and operation condition data (weather, temperature, humidity, air pressure)
- Airplane technical data
    - Air quality and pressure inside cabin, altitude, electric power, speed (Rolls-Royce, 2016)
    - Engine parameters (vibration, shaft pressure, and temperature, shaft pressure ratio, vibration stator vanes position, turbine cooling air temperature, shaft speed)
    - Wings monitoring (damage detection sensors (PennWell Corporation, 2016), stabilizer actuators, position, and transducers)
    - Icing and de-icing
    - Fuel consumption data
- Aircraft flight control system signals (Active Sensors Company, 2016)
- Security and crash alarms
- Cockpit voice recorder (First Market Controls, 2016)
- Flight data

With the aid of interview and literature survey, the mentioned MOL processes of the commercial airplane were analyzed. As a result of this analysis, sixteen groups of stakeholders were found. These groups are listed in Table 10. In this table, for instance, the government as the stakeholder has to define rules and regulations for airplane operations and use; finance the production or operation of the airplane, and sometimes operate the airplanes. The latter is especially common when the government is a shareholder of the airline organization. The internal stakeholders inside the government comprised of the ministry of transport, customs, border security, and regional/ national authorities.

Regarding the ownership of commercial airplanes, different types of ownership exist. It is common that many stakeholders own them. For example, ownership of an airplane can be in the form of a holding in which different companies own to share. In this form, a major share belongs to the main stakeholder (the holding company) which can be a parent airline company or group. Members of the holding company include airlines, government, military, finance partners or cargo transporters (airlines). Alongside the ownership are varying roles, which each stakeholder group undertakes. Another type of ownership occurs when individuals or private companies become owners (Iatrou & Oretti, 2016). However, in this research, the simplest case, regarding ownership, is considered.

From the perspective of contracts, different types of contracts are made in the airline industry. Particularly, the contracts concern financial issues for manufacturing or operation of airplanes.

Table 10: Stakeholders of airplane in MOL

Stakeholders	Roles	Typical internal stakeholders inside this group
Parent company (holding)	Groups of companies that receive the profit/loss, operating expenses from the airplane and its operation	Passenger and cargo airlines, bondholders, individual companies, corporations
Aircraft financing	Finance the purchase or operation of airplanes	Major aircraft finance groups, commercial aircraft sales and leasing companies, export credit agencies, lessors (Air transport action group, 2011)
Operator (Airline)	Services for operation of airplane and transportation of the passenger and cargo	Maintenance team, passenger flight sales, cargo sales, flight team, flight technical team, land operation team
MRO companies Maintenance provider	Modification, maintenance, repair and overhaul of airplane (ST Aero, 2015)	Technicians, flight engineer, maintenance supervisors, component specialists (Lufthansa Technik, 2016), technical and pilot training
Airline suppliers	Provide services to airlines for use and reoperation of airplanes	Flight services, fuel suppliers, major ground handlers, finance, maintenance, repair, and overhaul, systems solutions, major catering companies (Air transport action group, 2011)
Certificate body	Tests for quality, flight inspection standards, aircraft safety. Audit, certification	Preparation auditor, quality assurance inspector, aviation safety inspector, inspection manager
Government and Authorities	Rules and regulations, finance, operation (Schaar & Sherry, 2010)	Ministry of transport and infrastructure, registry for flag and route, airport authorities, customs, border security, regional/ national authorities (Air transport action group, 2011)
Sales channels	Business regarding use the airplane	Retail travel agents, major corporate travel agencies, major online travel outlets

Stakeholders	Roles	Typical internal stakeholders inside this group
Manufacturer (OEM)	Designer and producer. Pre-operation tests, shipment, aftersales service	Designer, maintenance and service group, production line, component and material suppliers
Organizations	Offer service or use the services from airplane	Tourism, credit rating agencies, media, employment agencies, private commercial companies
Education and training	Research and training providers/ users	Universities, flight training, research institutions (Air transport action group, 2011)
Associations	Promote collaboration and profession across different member organizations	National associations for airlines, airport associations, passenger right organizations, component suppliers, manufacturing associations, national air transport associations, cross-industry representative groups (Flight safety foundation, national air transport associations, Aircraft fleet recycling association)
User	The groups who use the airplane for transportation service	Passengers, organizations, freight forwarders (Air transport action group, 2011)
Competitor	Other transport modes	Railway, road logistics, shipping
Airport organization	Building and operating airport, including land operation and control tower management	Land operation team, tower control, flight engineer
NGOs	Groups taking care of environment and social issues	Interest groups (Schaar & Sherry, 2010), sustainable development groups, national and regional groups (Air transport action group, 2011)

In terms of financing, most common schemes for financing commercial aircrafts are secured lending, operation leasing and finance leasing (Airlines, 2016; Pulvino, 1998; Lessard, 2009). These contracts are written between investors and operators. Some other contracts are signed for airplane's usage. The availability-based contracts are the most common type in respect for airplanes usage. Rolls-Royce, for instance, leases the engines of airplanes under availability-based contracts (Rolls-Royce Corp., 2016).

From methodological aspects, step 1 of Freeman's method (as described in subsection 4.3.2) is completed with the aid of these two case studies. That is, general stakeholders have been identified.

4.4.4 Results: Potential stakeholders in MOL based on study of the wind turbine and airplane

After identifying the major stakeholder groups for each of the industrial case studies, a comparative analysis of the results was performed. More details of applying Freeman's four-step method are reported in Appendix B. In addition, the interested reader can refer to Wuest (Wuest, 2014) for a similar case of stakeholder identification.

This subsection presents the findings, potential stakeholders of MOL and their roles. The findings are presented in two models. These models address the relationship among stakeholders' groups and the products.

As shown in the previous section, Table 9 and Table 10 contain many stakeholders who differ in terms of the roles they play. However, some of them have common roles and responsibilities. Thus, the main roles in MOL of the wind turbine and airplane are synthesized from Table 9 and Table 10. They are presented in Figure 16.

Concerning the results of Table 9 and Table 10, it is possible to classify the stakeholders into three categories. Stakeholders who are in direct contact with the products belong to the first category. Therefore, they are considered more powerful or influential. This category comprises of groups such as OEM, operator, investor, and maintenance provider. They are displayed in Figure 16. Imagine one member, i.e., the maintenance provider, who can be involved in the service, inspection, repair and overhaul processes. For this reason, he/she has a direct influence on the product and considered as one of the main stakeholder groups of MOL processes.

The second category consists of actors such as the suppliers, competitors, NGOs and organizations who have less influence on the product or data of products compared to members of the first category (Figure 16). The remaining stakeholders who are less powerful or influential are included in other major stakeholder groups (Figure 16). In commercial airplanes, for instance, this group consists of education and research institutes, sales channels and associations (Table 10) They are among the less influential stakeholders for an airplane (in comparison to others such as an airline). Therefore, they are merged and grouped under the stakeholder group "organizations".

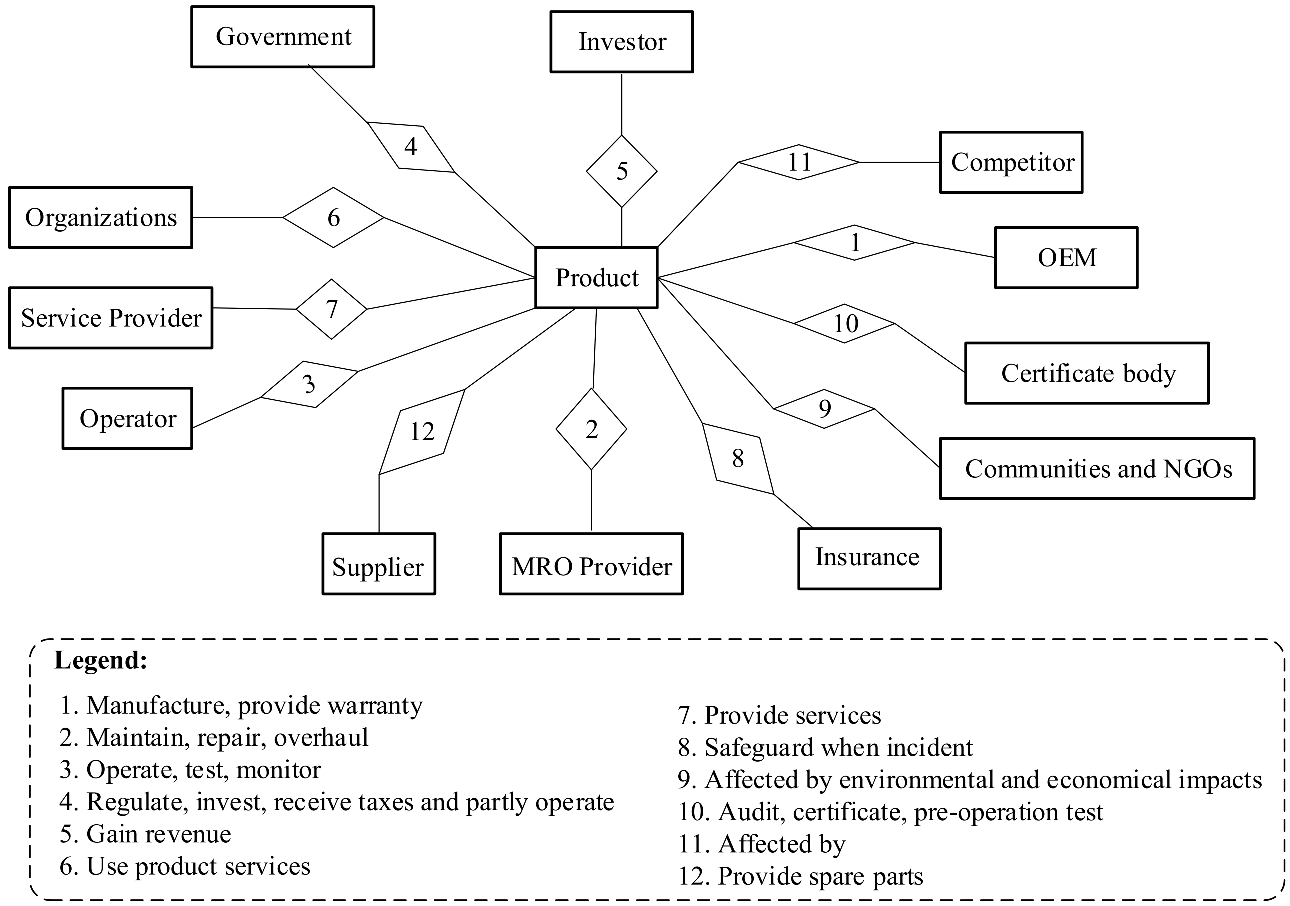

Figure 16: Major stakeholder groups of MOL and relationship with the product

This classification is done by considering the requirements of the second step in Freeman's Theory. The second step is power and influence assessment. For this assessment, we checked which groups are in direct contact with the product. Due to that, stakeholder identification in MOL aims to understand who can benefit from product-related data. Consequently, the stakeholder groups who are in direct contact use the data very often, which comes directly from the smart product.

The relationships between groups of stakeholders and products are illustrated in Figure 17. As shown in Figure 17, OEM has the role of providing warranty service, coordinates the availability of spare parts and does an overhaul (some cases). The operator performs the operation of the product, planning, scheduling, and monitoring for utilization of product as well as sells the service gained from the product (such as electricity). In addition, the operator performs maintenance in some cases. The manufacturer can be involved in the setup and activation process. Sometimes an operator carries out this process by himself.

The second achievement of this section is to model the collaboration among the stakeholders themselves (see Figure 17). This modeling is an informative way to discover relationships among different stakeholders of MOL. These relationships are necessary to elicit various types of needed data for stakeholders when they cooperate and share information. In this respect, each relationship can reveal the type of potential data to be collected and communicated.

Such typical relationships shown in Figure 17 depict the exchange of information, knowledge, products, and provision of supportive tasks or instructions. Below, section 4.5 defines the type of data, regarding the product that each stakeholder needs for decision-making, and that has been defined based on relationships presented in Figure 16 and Figure 17. It should be mentioned that these relations are a proposal based on studied cases and they can be extended in future for other engineered products.

4.4.5 Changes in roles in future (shifting to Use-oriented PSS)

The previous section showed twelve major stakeholder groups who exist in product MOL and can benefit from PUI. Figure 17 shows the relation and collaborative data sharing between the stakeholders. However, typical relationships are shown there. On the contrary, knowing this collaborative information sharing can help in case of automatizing the processes. An example of a collaborative information requirement as depicted in Figure 17 is as follows. The OEM provides after sales contract to the operator, at the same time the spare part for implementing aftersales service is delivered by suppliers and the maintenance process is performed by MRO provider. The coordination between supplier and MRO provider is done by OEM. Later this dissertation shows the changes in this collaborative information sharing in section 6.3, when automation and using PUI are taken into account.

As mentioned in section 2.1, one of the changes in engineered products is to consider products in combination with relevant services (PSS). Realizing PSS affects the relations among stakeholders of product MOL. Therefore, there is a need to consider the potential effects of PSS on collaborative data sharing (information needs) among stakeholders. To this end, this section presents discussions about the change of roles of stakeholders that results from new sources of data (PUI) and respective developments in PSS.

Roles are dynamic in most of the real-world situations. For this reason, it is unsurprising to find that manufacturers (and service providers) prefer to integrate products with services due to various reasons. Firstly, a sole product does not provide enough economic value and profit to the manufacturers. Given this situation, manufacturers have observed that more customer value and more profit can be gained by integrating products with services. Secondly, different types of product-availability-based contracts have led MOL stakeholders to be involved in designing contracts and predicting asset availability. In this context, IoT and Industry 4.0 technologies have received attention in designing Product-Services as they support the integration of products and their related services together. Thirdly, the manufacturer can only build smarter products provided that he has enough knowledge about necessary services to be integrated into the product.

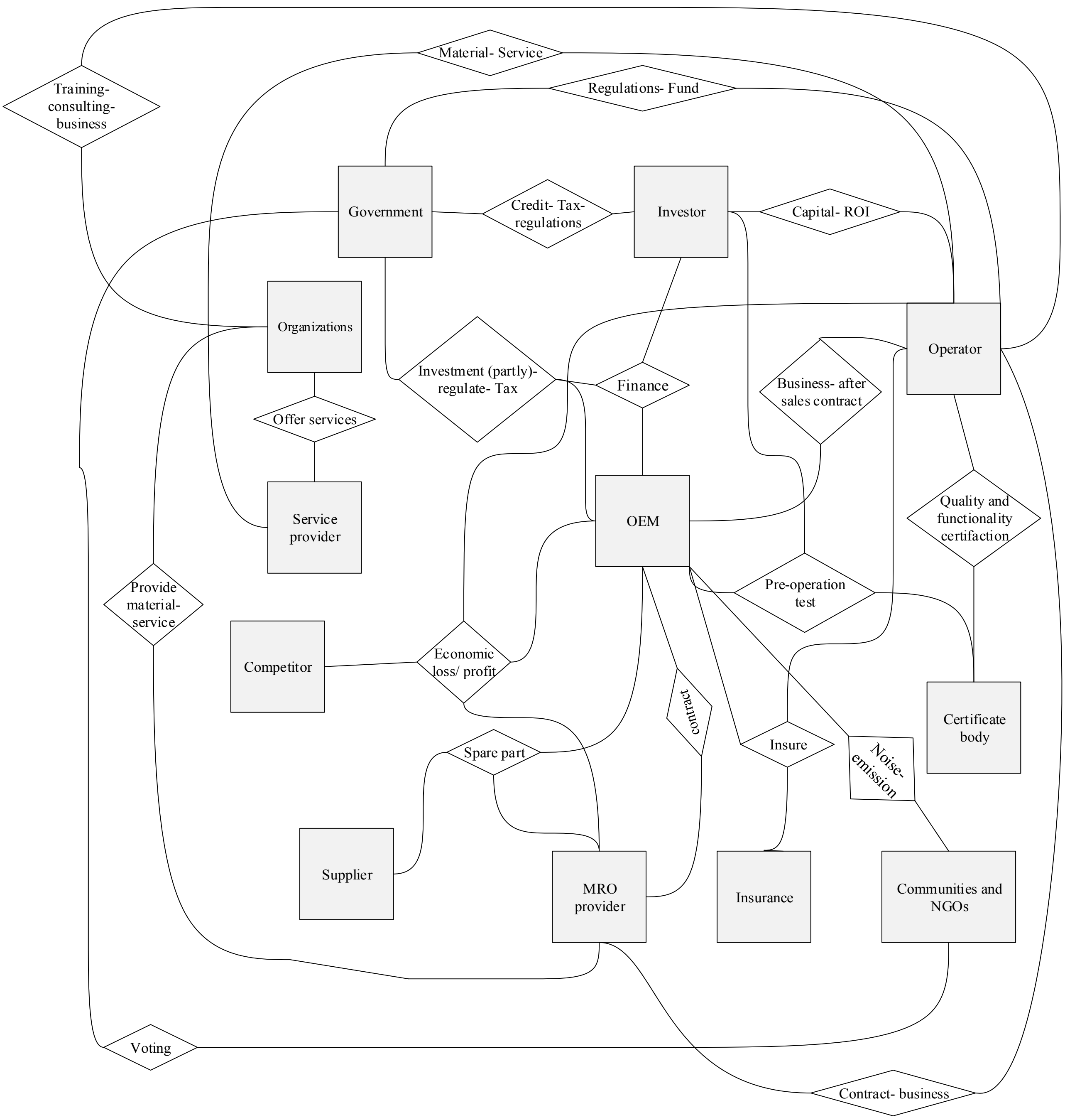

Figure 17: MOL relationships among stakeholders

For example, if the product offers the capability of predictive maintenance the manufacturer has to learn how to understand and manage this type of maintenance strategy. Thus, they have to know, which kinds of sensors to install on the device and what to measure.

Overall, these three phenomena have caused some changes in decision scenarios and data needs among stakeholders of the product lifecycle. Subsequently, stakeholders of product MOL has also been affected. The following paragraphs discuss the potential change in roles and collaborative data sharing among stakeholders when products are offered as use-oriented PSS.

Under Use-oriented PSS, a product in its MOL remains under the ownership of OEM rather than the operator, user or investor (Tukker & Tischner, 2006). Thus, some roles of stakeholders shift from classical to contract-based positions. Changes in the role occur (but unlimited) to the investor, OEM, MRO provider, government, certification body and service providers. The main difference happens due to broadening of commands/influences by OEM. By applying this type of product operating model (use-oriented PSS) the OEM or an entity, controlled by the OEM, has to take care of the operation of the system. Therefore, OEM faces many roles. The OEM can mainly be the manufacturer, maintenance provider, spare parts supplier and service provider. For instance, the OEM should care for providing spare parts (while without a PSS, product operators or MRO may be responsible for managing this task). Similarly, spare part providers, MRO company and organizations who provide other product related services can now contract OEM; for examples, those involved in logistics and transportation services or decommission the products and recycling.

Concerning the certification body, required audits formerly were performed at the site of operator/user. However, with changing roles, the certificate body has to change the scope of auditing for operating processes and parts such as performance tests and calibration. This is because such activities are no longer conducted by the operator, but by OEM who has taken the responsibility. Moreover, the insurance company has the option to work directly with OEM. Concerning insurance, this company collects some data from the product and its operation. Collected data are used only for assessing the risks and defining insurance fees. However, the data have the potential to be also used by OEM. In this case, these data can be used to give insight, from MOL condition of products to OEM. However, classically because of a confidentiality agreement between product owner (user) and insurance company, the insurer cannot use that data for commercial purposes. If the product data belong to the OEM and not the product user, it may be possible for the insurance and OEM to contract and share the data to improve product lifecycle. For regulatory bodies, new regulations about Product-Service offers should have to be formulated.

Taking together, among stakeholders of the product lifecycle, OEM may try to gain more governance on the entire product lifecycle. Considering this change, more opportunities in terms of contract-based services will emerge for other organizations so that they can collaborate with OEM in terms of contracting. Moreover, change in ownership rights can contribute to the availability of more data from MOL. Thus, more PUI source will become available. Availability of these sources paves the path for more service innovation.

New roles and jobs in product MOL. Considering the availability of PUI, there can be changes in the way current processes are being done. New practices can evolve. This part discusses potential new roles that can happen in the product MOL.

The stakeholder groups are likely to stay the same as identified in subsection 4.4.4. However, there could be changes in the roles of internal stakeholders. Internal stakeholders include employees inside organizations. The following shows some of the possible new roles.

Network Orchestrator. This role cares for coordinating the logistic network. In case of spare part supply, for example, the kind of transport mode, time, cost, the destination of spare parts would be checked. The difference between it and ordinary logistics provider is that Network Orchestrator decides the transport mode, for example, whether to send the pieces with a truck, by rail or by flight.

Digital chief officer (digital storyteller). The need has evolved to collaborate across marketing, sales, publishing development and in-house IT teams to foster digital thinking to other departments through a creative and distribution context.

User experience enhancer. User experience enhancer cares that the experiences their consumers make on their websites (through the user interface) are simple, intuitive and fun (Cadigan, 2017) at the same time, it can provide the organization with the data they need to get from users.

4.5 Data and information needs of stakeholders

Subsection 4.4.5 showed the identified stakeholders. This section investigates the data and information needs of stakeholders from PUI data.

The data and information needs that are required for each task in the use-oriented PSS are derived and presented. Product-Service offerings under use-oriented PSS scenario is considered in this section. The reason for studying use-oriented product services is that future product are more likely to be offered together with services. This consideration is backed up by the fact that use-oriented PSS can benefit from the digitalization trends in MOL.

Identification of data and information needs is firstly done based on typical data needs from the product (Figure 16). It is also supported by relationships between stakeholders (Figure 17). Therefore, potential data and information needs are clearly defined in Table 11. Later data and information needs from PUI are marked with an asterisk (Table 11). The table shows the stakeholders, the relations, and the information exchanged in each relationship. The column “Task” displays both the information exchanged between stakeholders (refer to Figure 17) and information needs from the product MOL (refer to Figure 16).

Table 11: Typical tasks, current and potential data and information needs of various stakeholders

Entity	Task	Typical and potential data and information needs
OEM	Receive finance	Interest rate- duration and terms of back payment- security of payment- costs of digitalization practices
	Manufacture (integrating feedback from MOL)*	**Product performance- functionality in MOL- status, behavior of products currently under use by other stakeholders**
	Provide after sale service (warranty)*	Sensor data for failure and state of use- the duration of use- location
	Receive material & service	Delivery data- quality- suitability to propose- price- potential disruptions- secure and authentic transaction
	Affect market	Profit- loss- direction of market- customer intelligence
	Receive spare parts	Delivery time- quality- quantity- meeting the required specifications
	Do maintenance contract*	Duration- scope- costs- selection among potential MRO providers- secure trading and knowledge sharing
	Receive insurance	Costs- service offers
	Receive pre-operation test	Ensure the functionality- customer approval- select the suitable certificate body
	Receive audit and certificate	Select the suitable certification body
Operator	**Receive aftersales service***	**Services, which should be done - service time and duration- estimated time for next service- predictive maintenance data (degradations, remaining life of components)**
	Operate, test and monitor	Demand- cost estimation- root cause- required skills-tooling- required spare parts and material- asset availability
	Obey the regulations	Regulation scope- new requirements relating digital trading and use of technology, international conventions, national regulations
	Receive consult /other services	Performance- costs- availability and quality of service
	Sell product (services)	Customer identification- market demand- contract design
	Return investment	Profitability of contract- contract design- rate of return
Certificate body	Perform pre-operation test	Sample data from processes-product function-specification
	Certify quality and functionality	Process info- performance of product/ employee/ organization
Government	Regulate*	Statistics- effects of product/service on society & economy
	Invest*	Potential benefits of project- ROR
	Receive taxes	Annual budget- required amount of income from taxes- payment capability of taxpayer (tax liability)
	Receive public voting/ communities requests	Voting rate and results- benefit/loss/ possibility of accomplishing the request
NGO and communities	Voting	Employment possibilities- potential damage to environment – support/ objection of public/ members
MRO	**Receive Spare parts**	**Inventory control for spare part, demand for spare part**
	Contract for service	Performance of product, history of service, history of spare parts

Entity	Task	Typical and potential data and information needs
Supplier of spare parts	**Provide spare parts***	**Frequency and amount of demand for spare part- delivery time**
Investor	Invest in production/ operation- grant loans*	Rate of return- security of return
	Gain revenue*	Rate of return- security of return- amount of return- **contract design**
Organizations	Offer services (training/ consulting/ business)	Demand- changes in market- competitor's strategies- regulations- price of purchased materials
	Use services of product	Scope of provided services- availability- quality- market for sell/use the service
Competitor	Affected by loss/ profit	Strategies of competitor organizations related to product- pricing- service offers
Service provider	Provide Service (to organizations)	Demand for service- potential of new services- potential customers- product performance- quality, availability & functionality
	Offer material/ service to OEM	Demand for material- amount and frequency forecasting

Note: The tasks marked with an asterisk (*) can currently be supported with new data sources using data analytics tools

As an example of data and information needs on the after-sales services, provided by the OEM, is illustrated here. To accomplish this service, the OEM needs/may need the following data (information): product failure; health status of the asset; operation hours of each subsystem, and the location of use. The OEM may need further information from the supplier, such as the terms of delivery, time of delivery, quality check report and price.

As previously discussed, new sources of information from smart products, when aggregated with currently available data sources, can provide more insights to stakeholders. Integrating such new data sources requires specialized analytical tools. As data analytics tools are being developed nowadays, they suit the present requirement. Prompted by this new requirement, the subsequent chapter, discusses in detail the data analytics and how major stakeholders can benefit from it. The discussion is extended to different data analytic techniques necessary to support decision makers (chapter 5).

Three of the data needs, which are marked with bold text in Table 11, are used in chapter 6. In chapter 6, use case scenarios are presented, which shows how the selected data needs can be met by using data analytic tools.

4.6 Summary and a generic picture of the concept

This chapter identified stakeholders of the product and their current and future information needs. The concept of stakeholder information needs shows the importance of considering stakeholders in the product lifecycle and their current and potential information needs from the product.

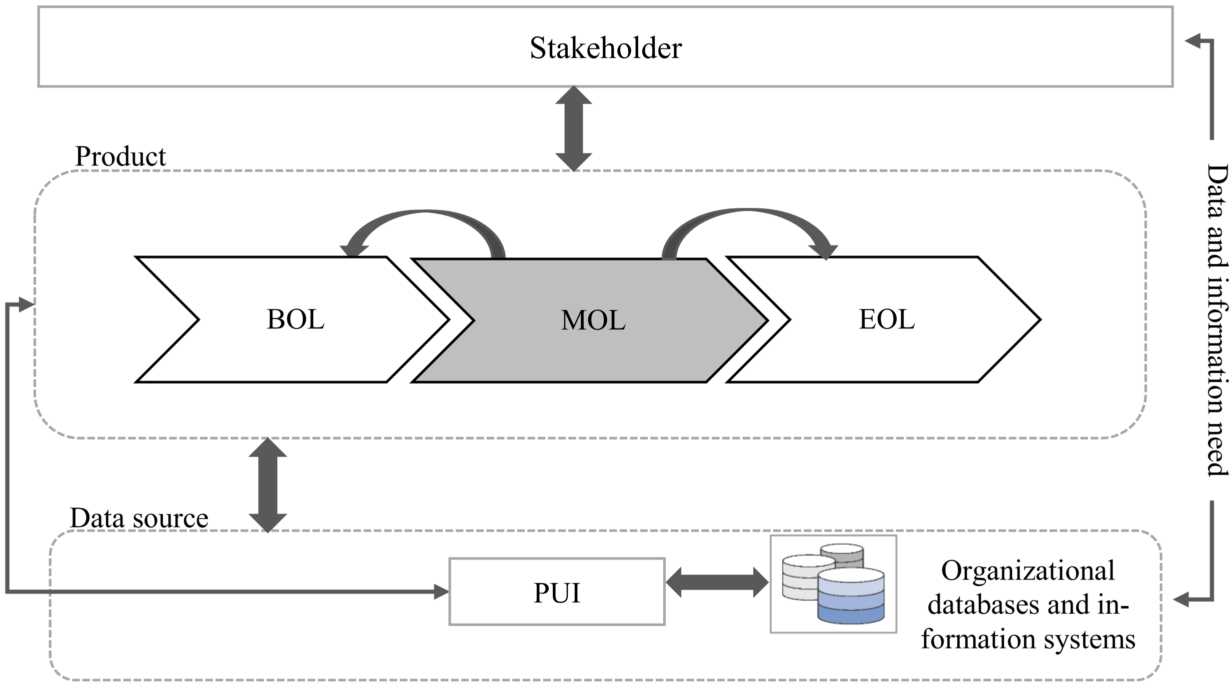

Figure 18: MOL stakeholders and product in operation, the interaction of elements in the concept of information and data needs with each other in CL-PLM

Figure 18 shows the constructs of "information and data need" concept. As shown, stakeholders are the major elements who define their data needs. This chapter presented the stakeholders and defined their data needs. In order to model data needs, they are turned into functional requirements and analyzed by data analytics. To be integrated to CL-PLM systems, these requirements should later be modeled with XML or other lifecycle-compatible programming languages. Based on the required variables in the data analytics and application scenario a new software application can be developed for providing stakeholders with an easy tool to access the data they need from the product. The data for analysis come from product and relevant information systems. Interaction of stakeholders with the service application or directly with the product affects the products and is shown with an arrow in the figure. The feedback from this application is back again to the system and applied. Later the feedback from the application can be shared with the lifecycle phases and used for future applications.

The data and information needs of stakeholders can be seen to have similarities with data and information requirements, which is addressed in the field of software engineering during requirement elicitation. However, this dissertation has chosen a management perspective rather than software engineering perspective. This perspective is chosen, because holistic management of product lifecycle and management of stakeholders was the focus on this research. Therefore, the term “data and information needs” has been adopted in this research instead of “data and information requirements” to consider this different perspective.

Chapter 6 discusses the application of data analytics to meet the information needs of stakeholders. To show the relevance of this discussion, it is enough to say that the concept in this dissertation considers a mechanism to process the data. This mechanism uses big data analytics from a technical point of view, which is a new application in this area. Using big data techniques to handle the PUI data can moreover contribute to a dynamic- information- process. Chapter 5 fills the investigation gap in this area.

5 Towards applying data analytics to meet the needs of stakeholders

As we have seen in the previous chapter, stakeholders of product lifecycles have been identified, and their current and potential information needs have been discovered. Because of the large volume and complexity of MOL data, stakeholders can benefit from these data only if it is processed. Therefore, techniques for data analytics such as data mining and machine learning are required. In this chapter, the application of data analytics is investigated to support the information needs of stakeholders. Areas are identified and classified in which data analytics techniques are applicable. Applicability here means that a level of transparency should be achieved and that meaningful information should be brought from product MOL data.

5.1 Criteria for selection of methods of data analytics for CL-PLM

Subsection 3.3.1 showed some standard classifications of data analytics techniques. However, in order to apply different techniques of data analytics in the domain of CL-PLM, there is a need to take into account several issues. Figure 19 shows some of the parameters (criteria) that should be taken into account. Taking these parameters into account can enable better classification of data analytics for CL-PLM, based on characteristics of information needs. These criteria are as follows. (a) Tools should be able to handle PUI data in terms of complexity and high volume. (b) The classification of tools should support decision-making in terms of the roles of stakeholders and their decisions in MOL. (c) Classification of tools should be sorted based on the purpose of analytics. For instance, whether something is going to be predicted or classified, or if the relationship between variables is modeled or not. (d) Tools should be sorted based on the format of input PUI dataset (e.g., tools for analyzing text data are different from those for analyzing sensor data). Another characteristic is (e) if the decision addresses a single product, several connected products and the operation between them in CL-PLM, or the business/collaboration that is caused by the use of products and their services. Based on these parameters (criteria), suitable analytics should be prescribed, which can be implemented in CL-PLM based on the application for the decision-making of different stakeholders.

Different algorithms of data analytics are analyzed to check if they can meet the mentioned characteristics. The research methods for this analysis are survey and experiment. Surveys are done utilizing available research papers and technical publications regarding applications of data analytics. Experiments are conducted to test the applicability and suitability of data analytics techniques in various application domains of the product lifecycle.

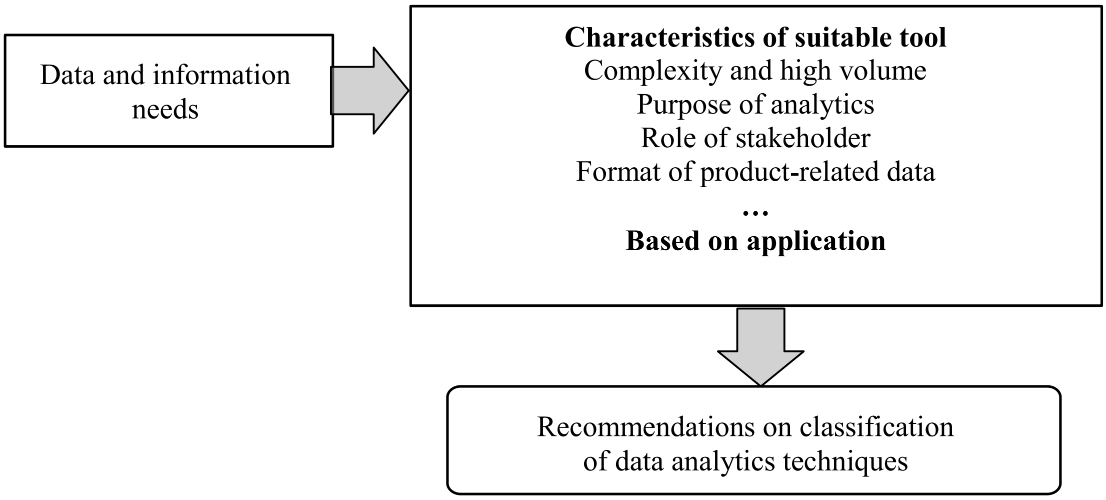

Figure 19: Investigating suitable data analytics technique for information need

5.2 Results of survey: Suitable data analytics techniques for CL-PLM

Table 12 gives an overview of approaches and applications for data analytics in state of the art, which can be applied in CL-PLM setting. The presented list of papers is selected after an extensive review of the literature. In the following short description of each approach is provided.

Vera-baquero et al. (Vera-baquero, et al., 2014) presented three categories of analytical methods. This classification is taking the business data and processes into account. Historical data analytics, business activity monitoring and predictive analytics are the three categories. Some examples of algorithms in each group are provided in Table 12. These methods are useful in decision support system development. From this perspective, they can be used for realizing MOL stakeholders' information needs. However, they do not fully cover the criteria mentioned in 5.1.

Hahmann et al. (Hahmann, et al., 2011) provided a classification based on the algorithmic and computational variations of the data analytical methods (Table 12). More explanation of each class is available in the data mining and big data analytics textbooks such as (Kaufmann, 2007). This classification, (Hahmann, et al., 2011), covers the criteria for "the purpose of analytics" (Figure 19). However, it does not take into account other criteria of information needs of MOL stakeholders.

Gandomi & Haider (Gandomi & Haider, 2015) divided the big data analytics applications based on the data format and analysis purpose. For instance, video analysis, text analysis, etc. Meer & Dasgupta (Meer & Dasgupta, 2012) introduced the analytical applications, which can be used in financial decision-making. These classifications are very useful when considering sorting data analytics in CL-PLM based on the purpose of analytics (e.g. (Hahmann, et al., 2011)) as well as based on the format of input PUI dataset (or MOL data). However, they do not consider the application (decision) scenario. That is, to classify information needs based on stakeholder and characteristics of decisions, these groupings are not feasible. Because each stakeholder may use several analytics methods from all of these groups; e.g., the operator uses association analysis as well as predictive analysis at the same time for gaining information about the functional performance of the product.

General Electrics categorized suitable analytics for dealing with data from a wind turbine (see Table 12) (General Electrics Corporation, 2013). This classification meets the mentioned criteria of section 5.1 more than the other available classification. However, a modification is still required. Thus, this dissertation extends the proposed categories for classification of data analytics and maps them in the context of CL-PLM applications. The procedure of this mapping is be presented later in the next section.

Table 12: Classification of data analytic approaches for improving the decision-making (techniques adopted from the category of data mining)

Reference name	**Classification of big data analytic approaches based on application**
(Vera-baquero, et al., 2014)	Historical data analysis for analysis of throughput rates, study process behaviors, detect undesirable behavioral patterns Business activity monitoring (identify bottlenecks and delays in or near real time, alerts for expected performance) Predictive analytics (build hypothetical scenarios, perform what-if type simulations, return event streams in simulation mode)
(Hahmann, et al., 2011)	Classification, association rules, prediction, clustering
(Fernández, et al., 2014)	Remote sensing, multimedia applications, biometric applications, biological and biomedical applications, pattern recognition applications
(Gandomi & Haider, 2015)	Text analytics, audio analysis, video analysis, social media analytics, predictive analytics
(Meer & Dasgupta, 2012)	Algorithmic capabilities, predictive capabilities, descriptive capabilities, reporting and MIS capabilities
General electrics (General Electrics Corporation, 2013)	Asset optimization, operation optimization, business optimization

5.3 Classification of methods of data analytics for CL-PLM

As the application area in this dissertation is data needs for stakeholders of CL-PLM, the proper classification of data analytics should take the characteristics of PUI and roles of stakeholders, and their decisions in CL-PLM into account (as shown in 5.1).

Adopting from General Electrics Corporation (General Electrics Corporation, 2013) and characteristic mentioned above of decisions in CL-PLM, this dissertation proposes three levels of analytics for supporting stakeholders and decision in the CL-PLM. These levels are asset analytics, operation analytics and business analytics. Figure 20 shows these categories of analytics, which can support stakeholders and applications in CL-PLM. Given that data from product usages (PUI) are available, these three levels can help stakeholders integrate these new data sources in their decision-making.

Asset analytics presents the tools and techniques, which support stakeholders in decision- making, concerning the product and its functionality. Decisions about repairing or replacing a component in the product, optimizing the performance of the product and improving the design of the product are in this group. The second range of tools helps stakeholders in planning and optimizing the processes of the lifecycle and operations. The decision regarding cost optimization, product configuration for achieving minimum operating costs and time, determining the best layout of a wind farm, minimizing energy consumption in a production line on the plant, scheduling maintenance, transportation analytics and sales analytics are categorized in this group. The third group of analytics relates to business analytics. These groups of tools support the collaboration between the lifecycle stakeholders and decisions regarding the business and strategic tasks in the product lifecycle. Examples of this group include contract optimization for developing product services, improving the electricity sell from wind turbines by analyzing the market, customer segmentation and supply chain analytics (Figure 20).

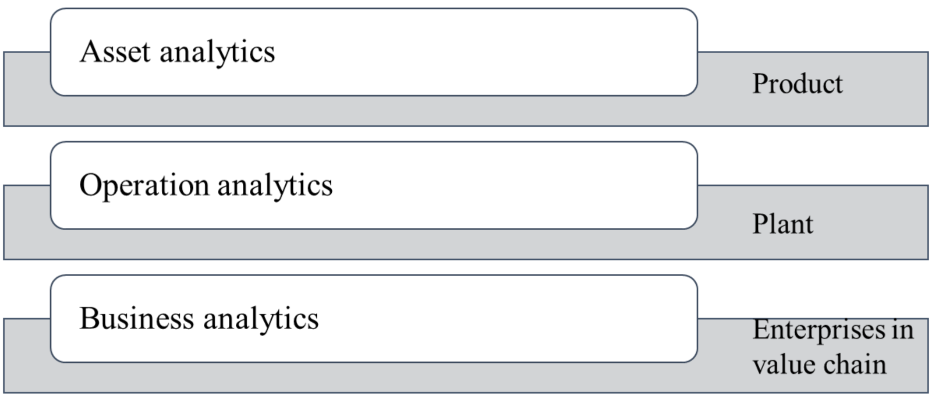

Figure 20: Proposed levels of analytics in the CL-PLM (for decisions in CL-PLM)

Based on these categories, the techniques suitable for each of identified stakeholder (based on information needs of section 4.5) are suggested in this chapter, as shown below.

Table 13 shows different methods and algorithms that each of the three categories (asset analytics, operation analytics and business analytics) can contain. Analytical methods and application areas in Table 13 are partly adapted from Chen et al. (Chen, et al., 2012).

Table 13: Application areas of data analytics in CL-PLM (types of analytics) (adopted partly from (Chen, et al., 2012))

	Asset analytics	**Operation analytics**	**Business analytics**
Application (technical approach)	Lifetime estimation, Component condition monitoring, Product performance monitoring (Alert system- track & trace).	Demand forecasting, Scheduling, Inventory planning, Distribution Scheduling, Distribution risk, disruption and safety monitoring.	Economic analysis, E-commerce, Market analysis, E-government, Security, Network monitoring, Financial forecasting, Citizen engagement, E-Poll, Application based on search engines and data management, Customer segmentation.
Analytics	Recommender system by association rules, Information integration, Semantic services & anthologies, Neural networks, Reliability modeling, Degradation modeling, Regression modeling, Probability modeling, Pattern recognition.	Preserving privacy by data mining, Spatio-temporal analysis, Multivariate analysis, Clustering, Time series modeling, Mathematical optimization, Regression modeling, Neural networks, Support vector machines, Pattern recognition, Alert systems.	Recommender Systems, Social media mining and monitoring, Crowd-sourcing, Social and virtual games, Anomaly detection, Text and web mining, time series modeling, Clustering, Decision tree modeling, Pattern recognition, Visualization, Alert systems.

The following example illustrates the point: (Zhang, et al., 2017) studied the application area of component condition monitoring for blades and rotors in an axial compressor. A relevant decision problem in CL-PLM can be to monitor and control the blades and rotor preformation based on PUI sources of axial compressor. The PUI can contain information about the performance of the rotor or blades. The CL-

PLM application area of lifetime estimation and product performance monitoring (see Table 13) can be improved by having these PUI sources. Moreover, analytics such as Neural Networks, Reliability modeling can be used to process data and provide a better understanding of the performance of each component for each related MOL stakeholder.

5.4 Suggestions for suitable data analytics for each stakeholder[3]

As shown in section 5.3, this dissertation classifies data analytics for applications in CL-PLM to asset analytics, process analytics, and business analytics. Building on these findings (section 5.1 to 5.3), this section proposes potential appropriate data analytics techniques for each stakeholder. Moreover, data analytics techniques for each stakeholder are expressed in different decision-making applications.

The research method for this section compromises experiments and analyzing state of the art. The proposed data analytics techniques (see Table 13) are matched with MOL stakeholders by considering different CL-PLM areas in which a stakeholder has a role (see Figure 20).

Table 14 shows the methods of data analytics to support stakeholders in MOL. Applications and suitability of data analytics techniques for information needs of MOL stakeholders are determined partly by experimental testing of techniques. Another method for identifying suitable data analytics is the analysis of scientific papers. To illustrate this table, some stakeholders are selected here. The product operator, OEM (manufacturer) and investor are chosen to demonstrate the potential needs of the data analytics (Table 14). Product operator can benefit from applications of big data analytics through forecasting and monitoring the product condition. To reach the goal, new sources of data have to be integrated with the current condition monitoring systems to increase the power of prediction. For example, sensor data from the wind turbine's operation with the weather data, integrated with the current condition monitoring systems can be used to predict the failure of components. For technical details and similar examples refer to (General Electric Company, 2013). Likewise, data-driven approaches can help operators to reach the best product availability (Yin et al., 2014; Si et al., 2011). Nevertheless, it is crucial to monitor the health state of the components. To achieve this, methods of anomaly detection can be applied to detect the irregular behavior of components. Finally, the degradation analysis can be used to identify wear and corrosion in the parts. Later in chapter 6, the suggested analytics for operator, OEM and MRO provider are implemented through prototypical scenarios. These stakeholders are marked with bold text in Table 14. Moreover, the analytics, which is implemented in chapter 6, is illustrated with bold text.

[3] Parts of the text in this section has been published in (Nabati et al., 2017)

Table 14: Proposed applications of data analytics to support stakeholders

Stakeholder	Asset analytics	Operation analytics	Business analytics
Operator	**Applications of big data for availability forecasting** (Anninni, 2014), reliability analysis (Pelham, et al., 2015; Kauschke, et al., 2016), estimating the remaining life of components (Kammoun & Rezg, 2018)	Route development (Akerkar, 2014)	
OEM	Supplier performance evaluation with track-and-trace product during its usage (Pátkai & McFarlane, 2006), **component degradation analysis**	Optimization of production process, risk, and disruption controlling through social media mining	
Authorities			Applications based on hearing, Social media analysis, crowdsourcing, opinion mining (Anninni, 2014), monitoring the long-time effects on the environment
Investor	Simulation for depreciation analysis, obsolescence management (Jennings, et al., 2016)		Economic analysis, sustainability analysis
Spare part supplier		Optimization of lead time for the supply of the spare parts, demand forecasting (Tracht, et al., 2013)	Risk and disruption controlling through social media
Logistic provider		Applications based on forecasting, alert systems for bad weather condition	
NGO and communities			Understanding the opinion of the general public from social media, alert system for new employment opportunities
Organizations		Vendor performance analysis and monitoring, cluster analysis for customizing services with customer segmentation, improve service quality with performance evaluation by integrating PUI data	
Service providers		Understanding value-added services by mining opinions of users	Enhanced demand forecasting for service by integrating product status (PUI) and customer data (defining trading and purchase strategies

Stakeholder	Asset analytics	Operation analytics	Business analytics
			by demand forecast information)
MRO provider	Reliability analysis, estimating the remaining life of components, lifetime assessment for each component	**Forecasting the order time of spare parts** (Taigel, et al., 2018)	Data-driven business model
Certification body	Degradation analysis	Condition-based maintenance (condition monitoring)	

For the product operator forecasting the demand is the most critical task. Methods of time series analysis through integration of the environmental data (smart product data) can be applied to predict the demand.

Similarly, for the OEM, the methods for monitoring the availability and performance of the product and its health state are useful for aftersales services. New MOL data sources can provide OEM with information about the actual condition in which the product is being used. This can enable OEM to offer more customized after sales services to the customers. OEM also needs to monitor the supply chain. For instance, the coordination of the provision of raw material should be monitored. For this purpose, new methods of data analytics such as using social media analysis and text mining to monitor supply chain (e.g., disruptions and performance) have gained popularity in recent years. Another operational decision scenario for OEM is finding the quality problems of product or material. In this respect, exploring the root causes of problems in material can be facilitated by analyzing quality-related data in orders. Moreover, advances in the RFID technology gave OEM the chance to track and trace the product. When data of RFID are linked with the process data, it is possible to solve quality problems and find the root causes.

Investors can gain insight from MOL data sources, which allows them to design better contracts. For instance, economic analysis and forecasting of the rate of return on investment can be better done when the performance of the product is predictable. In the case of wind turbines, for instance, the geographical location of the wind park influences the fluctuation of electricity production. Concerning the availability of winds and strength of wind speed, they should know, which areas are more appropriate for investment considering geographical, environmental (sustainability) conditions. Being aware of this type of information can help investors to choose the most suitable projects to invest in, furthermore, as mentioned before the product can be offered under a lease agreement. In this case, if the ownership of it remains with the investor, he/she should consider depreciation management and obsolescence management. Both these issues can be supported by data analytics.

5.4.1 Suggestion for internal stakeholders regarding data analytics applications

It is possible to suggest relevant data analytics approaches for the internal stakeholders of each stakeholder group. Some of the internal stakeholders have been introduced in previous chapters (Table 9 and Table 10). To illustrate more and give examples, Table 15 shows selected internal stakeholders. For instance, internal stakeholders within the "operator" group are maintenance manager, control room operator and technicians. Due to a large number of internal stakeholders, only a limited number are studied here. The table presents relevant analytics for supporting the information needs of each internal stakeholder.

Recommended analytics for each internal stakeholder depends on its role, decision scenario, and type of available PUI sources (similar to 5.1). For example, one of the potential applications of analytics for product designers (in OEM group) often involves design prototypes (samples) that use simulations for modeling the dynamics. The simulation output data are not usually used for any purpose. If the output simulated models are further analyzed by a data analytics approach during the time, or after some versions of design change, it is possible to get further insight into the design parts or the correlated parameters. For example, wind turbine blades are one of the most critical components of the wind turbine. Blades are subjected to large variations in stress while operating. Throughout the year temperature variations and severe weather conditions like rain and ice also affect the operation of blades (Dassult Systems, 2015; Wind Energy Hamburg, 2016). Simulation models are developed to predict the behavior of the turbines under different environmental conditions. Usually, these models show the behavior of one component in a unit product. By applying data analytics, it is possible to see the wear and tear of parts during the time. In addition, it is possible to integrate this information with predictive maintenance system to receive suggestions about service times.

As stated in Table 15, visualization can be beneficial for technicians (an internal stakeholder of the operator). In some cases, even with a straightforward display, a solution to a problem in a product's components can be found. For instance, by analyzing the heat map images of stress on the components of a product, during the time, it could be possible to explore that, there is a correlation between two components, which previously was unknown.

The three categories of analytics for CL-PLM, namely, asset analytics, process analytics and business analytics (Figure 20) are also useful for internal stakeholder groups. For example, maintenance administrative manager (Table 15) can benefit from the results of business analytics. Likewise, while sales and marketing staff can benefit from process analytics, maintenance technicians can benefit from asset analytics.

Table 15: Types of big data analytics to support the stakeholder (internal stakeholders)

Stakeholders	Internal stakeholder	Analytics to support the stakeholder
Operator	Maintenance administrative manager	Predictive maintenance **(Butler, 2012)**. Applications of data analytics for forecasting for demand costs etc.
	Control room operator	Multivariate statistical analysis for monitoring the behavior of several products together (e.g. all turbines in a farm)
	Technicians	Real-time visualization, enhanced dashboard interfaces of components and measures
	Product designer	Simulation for design for maintainability with modeling the behavior of different component modules of turbine and their interdependencies **(Dassult Systems, 2015)**, enhanced design of smart products through PUI data integration, visualization and computing
OEM	Sales and Marketing	Recommender systems for suggestions on products, clustering, and classification of potential and current users, e.g. for electricity sales **(SunGard solutions, 2013)**, market segmentation
	Product EOL collector and recycler	Tracking with web-based data, estimating the life of parts while being reused, track and trace part, supplier performance

In summary, different stakeholders can benefit from new sources of PUI/ MOL data. A suitable and necessary tool is data analytics. Examples have been presented in this section for stakeholders and the appropriate analytics, which can support their roles (decision-making tasks). In section 5.6, the procedure for the implementation of data analytics techniques for supporting MOL stakeholder by analyzing PUI is described.

5.5 Steps of implementation of data analytics

So far, stakeholder groups who can benefit from product MOL and their information needs were identified and data analytics suitable to realize information needs defined. Next steps are to know, how it is possible to implement mentioned data analytics techniques in practice and get the right information that can respond to the information needs of beneficiaries? To answer this question, first, the generally accepted procedure for implementing data analytics projects are explained below. In addition, later in chapter 6, different implementation scenarios are presentment.

CRISP-DM is a well-known standard for data mining. It demonstrates the standard procedure to carry out a data analytics project (Shearer, 2000). Figure 21 shows the steps of applying data analytics in CL-PLM. These steps are adopted from CRISP-DM and modified for being applicable to CL-PLM environment.

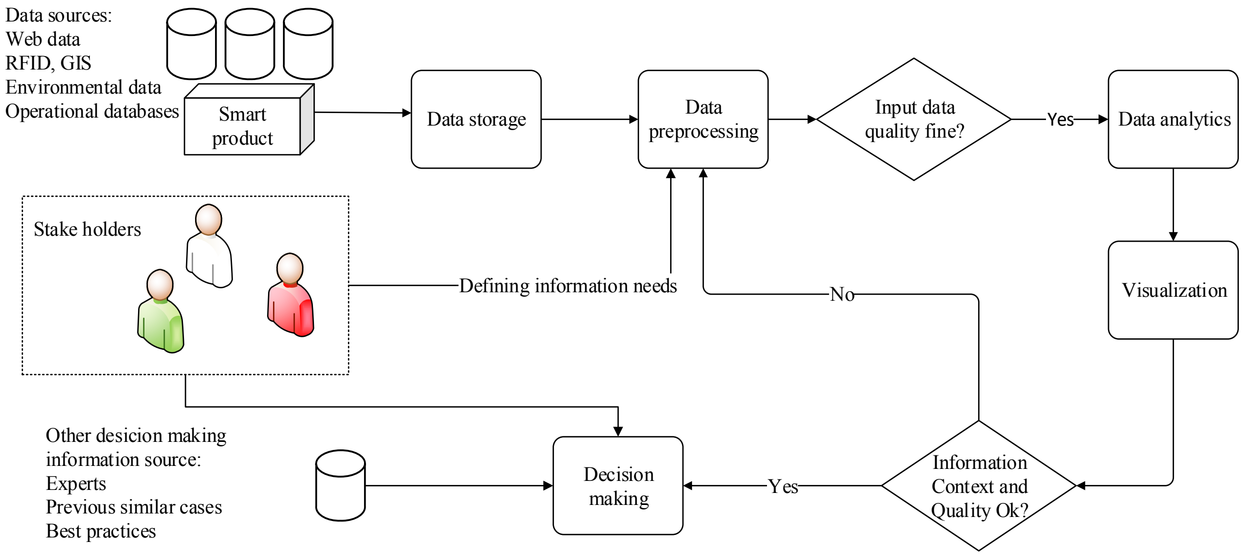

Figure 21: Steps of applying data analytics for decision-making in CL-PLM

Figure 21 shows the procedure of using data analytics for the conversion of PUI and providing MOL stakeholders with their required information. The smart engineered product collects the data from the product in operation, via sensors. The data from other sources such as organizational databases are also identified and acquired. Afterward, all or a part of the data are stored. Based on the information needs of stakeholders and the decision circumstance, data selection tasks start by considering attributes of variables (type of measurements) in PUI sources, which are stored. We can call the dataset, which is formed at this stage as "input dataset". After relevant input dataset (from PUI) have been selected and its features specified, parallel processing and distribution are done. This stage depends on the volume of data. If the volume of the dataset is not very large, this step can be skipped. Next, the input dataset should also be pre-processed. The goal of pre-processing is to increase the value and accuracy of the input dataset. During preprocessing missing values are handled and normalized, until it is ready for use in the analytics. Data analytics contains the models for three categories of asset, process and business analytics (see 5.3). It can perform predictions, classifications and extractions or finding the associations between the variables of the input dataset. The results of analysis later are delivered to the stakeholders through the user interfaces (e.g., dashboard). At this point, the user assesses the quality of analysis results, adapts it to the logical thinking process and considers it as an alternative to decision-making, performing other analytics (if necessary) or rerunning the same model. With the help of gained knowledge, it is possible to specify the other decision alternatives.

5.6 Discussion and potential applications of data analytics in CL-PLM

In summary, this chapter provides an understanding of data analytics techniques and classes, which can be used for answering the information needs of stakeholders. Different stakeholders are mentioned, and functional requirements are defined. By functional requirements, it means here, the techniques of analytics and examples of relevant decision scenarios, which stakeholders face. Particularly when stakeholders work with PUI, these tools are necessary for them to obtain their desired information.

Big data analytics also has the potential to be used in CL-PLM for monitoring the collaborations between the stakeholders, particularly for ensuring the secure exchange of information. Therefore, fraud detection and monitoring the security of the network with data mining methods are another potential application area of data analytics in CL-PLM. To this end, data access rights, unauthorized access and the flow of information can be monitored for every stakeholder. Publications such as (Moustafa, et al., 2017) address fraud detection, intrusion detection with big data analytics (Moustafa, et al., 2017). As mention by Moustafa et al. (Moustafa, et al., 2017), “Intrusion detection methodologies have been developed using approaches involving data mining and machine learning, artificial intelligence, knowledge-based and statistical models”. Although the following applications are available the possibility of applying them to monitor stakeholders’ collaboration in CL-PLM requires further research.

Regarding the mentioned operational analytics (section 5.3), several novel applications can be devised and their applicability to CL-PLM be tested. For example, the operator can benefit from operational analytics to find the percentage of customers with the highest affinity to sign a long-term contract (Ahlemeyer-Stubbe & Coleman, 2014). He can also use the applications to find and describe groups of customers with homogeneous usage. Moreover, understanding operational features and general business forecasting can help the operator to achieve better performance.

In summary, this chapter introduces a classification of data analytics for different applications of it in the CL-PLM. It also investigates the kind of techniques, which can be used for the realization of each information need of stakeholders. The contribution of this chapter can be a scheme, that specifies the major areas and application domains of data analytics. The people who want to apply data analytics in the product lifecycle can use this as a guideline. This study was tried to be comprehensive. However, other cases and applications in future can be investigated and possibly added to this template. The concept proposed in chapter 4 and 5 is the first step in showing the use of data analytics in CL-PLM. It shows the use of product data and information and applications of data analytics to improve the information requirements. It can be extended in the future.

The next chapter shows the application of the concept of “MOL stakeholders’ data and information need”. The application domain of this concept is vast. Therefore, the implementation scenarios, in chapter 6, are selected in a way to be rather focused in order to reach a more transparent and effective vision of realization of a specific information need of a stakeholder from PUI sources. Thus, the details of the implementation of data analytics in practice are shown in every scenario.

6 Concept implementation

This chapter shows the real-world application of the previously developed concept of " MOL stakeholders' data and information needs" (Chapter 4 and 5) through three scenarios. Each scenario represents a decision-making case, for which the beneficiary requires product information.

Based on the suggested data analytics classes of chapter 5, applications of data analytics techniques in selected areas of CL-PLM are shown. Real-world data from the product use (PUI) are tested in the proposed methods. At this stage, the identified information needs (from section 4.5) are also examined. First, it is evaluated, whether the information needs of the decision maker have been correctly identified. Second, it is presented, how the information requirements are modeled with the use of data analytics. Third, the type of output information that stakeholders can expect to receive, in term of format and context, is explored.

The implementation scenarios are selected from three different products, namely wind turbines, leisure boats and electric vehicles. The reason for choosing three implementation scenarios is dictated by the necessity of testing the developed concept on different engineered products. Moreover, in each implementation scenario, for each product, the information need, and stakeholder is varied from the other. Also, data and information requirement of different stakeholder, as well as various data analytics techniques are addressed. This diversity enables us to check the accuracy and completeness of the concept better.

In this chapter, at the beginning of each section, the scenario is briefly introduced. Subsequently, the application of previously introduced approaches (chapter 5) on the data needs of stakeholder is tested. Finally, the results of processing PUI in the scenario are shown and compared.

6.1 Scenario 1: Contract optimization for the operator of a wind park

In the field of wind energy, selling the electricity of wind farm happens under different conditions. A part of electricity that is traded through energy exchange such as EEX (European Energy Exchange). The other part is sold based on contracts. While by the former way energy is traded every 10 minutes, the latter is long-term and contracts usually take between 5-20 years. These contracts are well known as the Power Purchase Agreement (PPA) or over the counter (OTC). OTC and PPA have differences from the financial perspective; however, from the perspective of the effect of using PUI for improving these contracts, they can be considered the same.

PPA is defined as a contract signed between an electricity generator and a power purchaser. In the area of wind energy, the electricity generator is usually a wind farm operator. The power purchaser can be a large power buyer/trader company. In some cases, power purchaser is the government. Moreover, the investor is another stakeholder involved. An investor can be a commercial bank which funds the construction of a wind farm or provides loans for operating the wind farm.

Designing effective contracts is very important for stakeholders involved in PPA. Under PPA, operator sells for a predefined period a certain amount of energy (electricity) produced by a wind park to the buyer (energy trader). In case of a stop of the electricity transmission, the operator has to pay the penalty to the buyer's organization. Therefore, it is evident that the availability of wind turbines for producing electricity is crucial.

The energy and payment price should be calculated as they are defined in the contract between the operator and the energy trader. In the case of overproduction, unless specified by contract, the energy trader does not make additional payments to the operator. In some case, the buyer can sell the excess amount back to the operator. In other cases, the operator cannot claim extra-transmitted energy. Therefore, the operator encounters a loss.

In this case, real-time PUI can help the operator to manage the contracts better. By gaining the data from wind turbines, it is possible to predict the output electricity of the wind farm and to monitor this output, so that the effect of variations in output electricity on contracts are as small as possible.

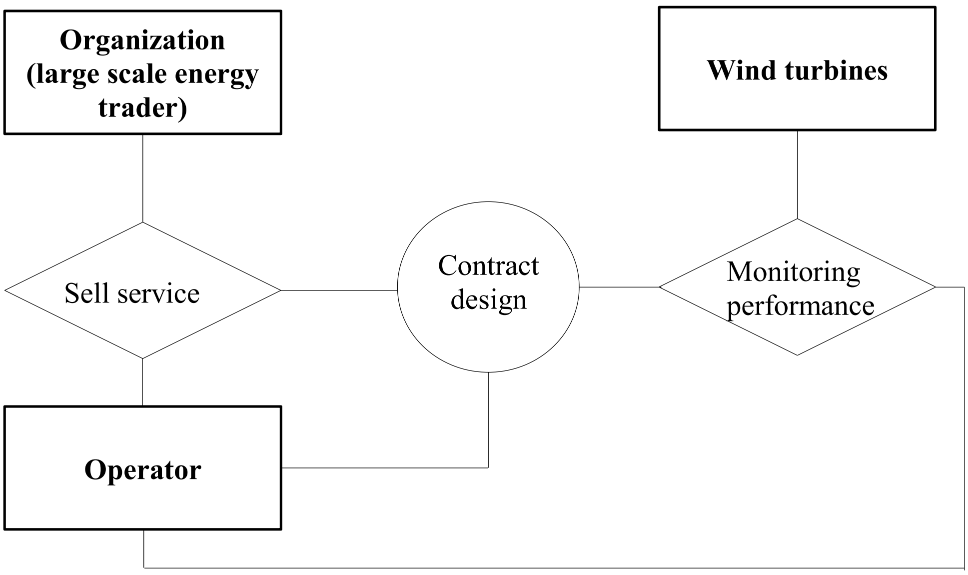

Figure 22: Information needs for transaction between operator and energy trader, with wind turbines PUI

Figure 22 shows the cooperation between the operator and energy buyer company. Energy trader has here the role of the organization who buys services of the product

(4.4.4). By monitoring the performance of wind turbines and measuring the variations in electricity production, contracts can be improved. To this end, this figure shows a contract management system, which uses the PUI. Other input information sources are the condition of funding and requirements for the rate of return for investment. The contract management system optimizes the design of contracts based on these parameters. This system can support operator and investor to provide their information needs.

In this scenario, two sets of PUI are used to improve contracts. The first dataset is measurements of operation of a wind park during of 7 years, as well as the environmental data. Analysis of this dataset aims to find out the times of the year, in which overproduction happens. Second, the dataset is data of operation and maintenance of a wind park, as well as environmental condition. The latter has been added to incorporate the available capacity of wind park for power production and therefore better estimate the amount of electricity that can be sold. Subsections 6.1.1 to 6.1.3 addresses the first dataset. Subsection 6.1.4 presents the second dataset. Table 16 provides a summary of information for scenario 1.

Table 16: Summary specifications of scenario 1

Role of stakeholder	Product operator- Product service buyer (energy trader)
Data need	Contract design (based on product performance)
Characteristics of available PUI	Electricity production, weather data, maintenance data
Type of analytics	Business analytics
Aim of analytics (application)	Data description (visualization)- Prediction
Algorithms	Visualization -CART- RF (see 6.1.1 for more information)
Reasons for selection of CART & RF methods	Because of showing results similar to decision-making logic (tree shape), ease of interpretation, and robustness of model construction.

6.1.1 Description of data, data analytics techniques and model accuracy

In this subsection, we use real-time data from a wind park. The data are available for free at http://www.sotaventogalicia.com/en/real-time-data/historical.

First, the data are used in order to understand the effect of the wind park's location on the overproduction of electricity. To this end, it is tested, which location-dependent factors such as wind speed, wind direction and time of year (month) influence the fluctuation of power generation.

For the aim of contract design, intervals between measurements do not need to be very close (e.g., every minute). Nevertheless, analysis can be more accurate if it covers a more extended period (e.g., several years). Thus, data collected daily over seven years (2010-2017) have been considered. Figure 23 shows the layout of the wind farm and characteristics of information available from the provider of the website. Moreover, Table 17 shows a data sample.

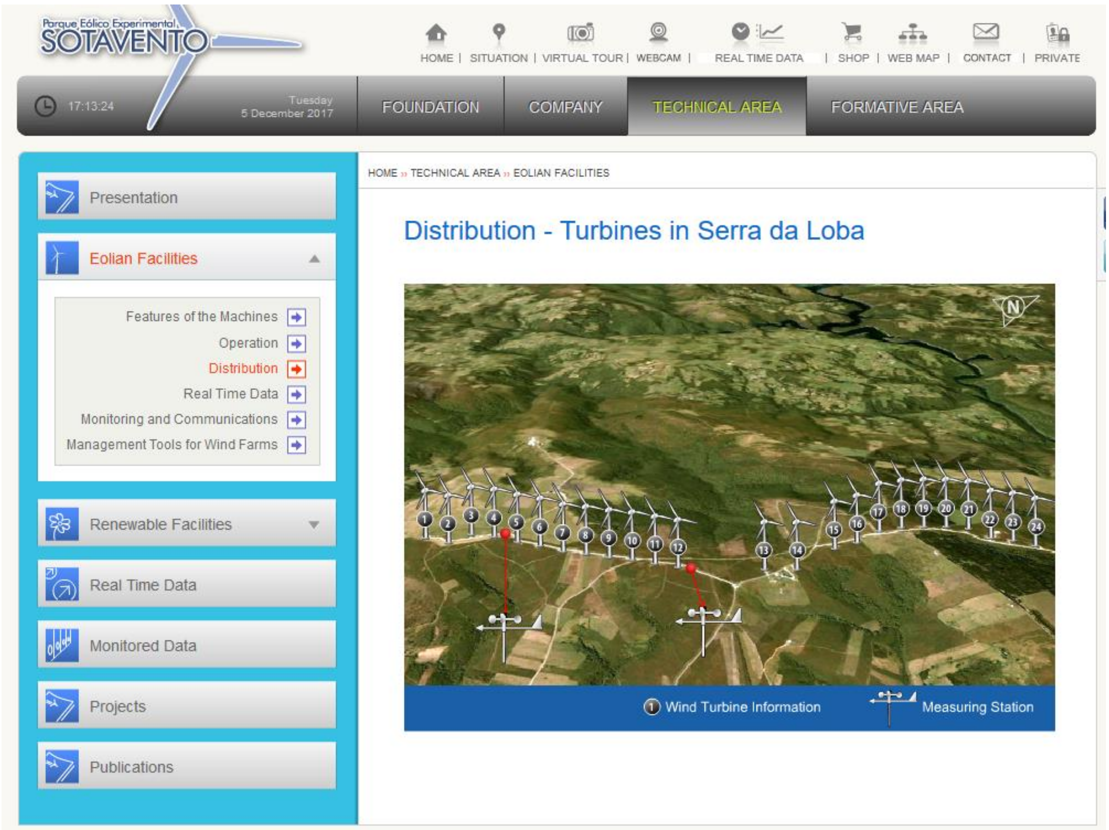

Figure 23: Real-time PUI data of a wind farm

Table 17: Exempt from data for the prediction of wind output

Date	Speed (m/s)	Direction (°)	Energy (kWh)
11/1/2017	5.07	143	63,211.84
11/2/2017	4.39	141	41,867.65
11/3/2017	3.35	138	10,926.37
11/4/2017	4.18	343	31,067.67
11/5/2017	4.04	316	18,708.46
11/6/2017	3.78	59	13,252.06
11/7/2017	3.7	281	21,563.23
11/8/2017	4.18	344	24,356.06
11/9/2017	3.12	355	3,285.56
11/10/2017	3.41	343	8,975.04
11/11/2017	2.57	306	165.57

Classification And Regression Tree (CART). Decision tree is a data analytics technique that uses a tree-shaped model for a decision and their implications. This technique can be used as a supportive tool for decision-making. CART is a technique of data analytics that can perform this issue. It can model the "non-linear relationship among variables. Moreover, it is suitable for taking high-order interactions, and missing values into account. The advantage of CART models is that they are simple to understand and results are easily interpretable" (De'ath & Fabricius, 2000). Respectively, it can show the variables, which affect electricity generation and the condition, under which the energy generation changes.

From the algorithmic point of view, tree-based models explain the variation of a response variable by repeatedly splitting the data into more homogeneous groups, using combinations of explanatory variables that may be categorical and/or numeric.

Based on the work of De'ath & Fabricius (De'ath & Fabricius, 2000) tree models can be used for interactive exploration and for description and prediction of patterns and processes. Advantages of trees include: (1) The flexibility to handle a broad range of response types, including numeric, categorical, ratings, and survival data; (2) invariance to monotonic transformations of the explanatory variables; (3) ease and robustness of construction; (4) ease of interpretation; and (5) the ability to handle missing values in both response and explanatory variables (De'ath & Fabricius, 2000).

For technical information, regarding trees, please refer to Breiman (Breiman, et al., 1984). In this dissertation, we apply CART for regression analysis to understand the variations of electric power production based on the month of the year, wind speed and wind direction. Moreover, we use CART as a classification tree in order to classify the wind speed (and its effect on electricity production).

Random Forests (RF). Random forests approach is another data analytics technique, which is build based on statistics. It works by building several tree models. In the case of regression, random forests calculate the average of these trees and find the output value for the prediction model. As this technique runs the prediction for several times, the answers are robust and result with less variance (comparing regression analysis models, analysis of variance, linear discriminant analysis, or survival models). Technical information regarding this method is available on Hastie et al. (Hastie, et al., 2009). In this scenario, we apply these two data analytics techniques for analyzing wind farm data.

Model accuracy measures. This dissertation uses two criteria for calculating the accuracy of data analytical models. The formulas for error calculation of these two measures, RMSE and absolute error, are provided in the following.

RMSE formula: MSE stands for mean square error. RMSE is the square root of MSE. MSE measures the distance between the real value $f(x_i)$ and the estimated value $\hat{f}(x)$. In addition, N is total number of observations (records), upon which the model is built.

$$RMSE = \sqrt{MSE}$$

$$MSE = \sum_{i=1}^{N} \frac{(\hat{f}(x) - f(x_i))^2}{N} \tag{8}$$

Absolute Error: Absolute error measures the average absolute difference between real value $f(x_i)$ and the estimated value $\hat{f}(x)$ for N observations.

$$\frac{\sum_{i=1}^{N} \left|\hat{f}(x) - f(x_i)\right|}{N} \tag{9}$$

6.1.2 Operators gain a better insight through data visualization

This subsection first visualizes the data. Data visualization can summarize the data. Therefore, it is an effective mean of providing MOL stakeholders with the information they require. The dataset contains wind speed and direction. The visualization is done on the first dataset (mentioned in 6.1.1).

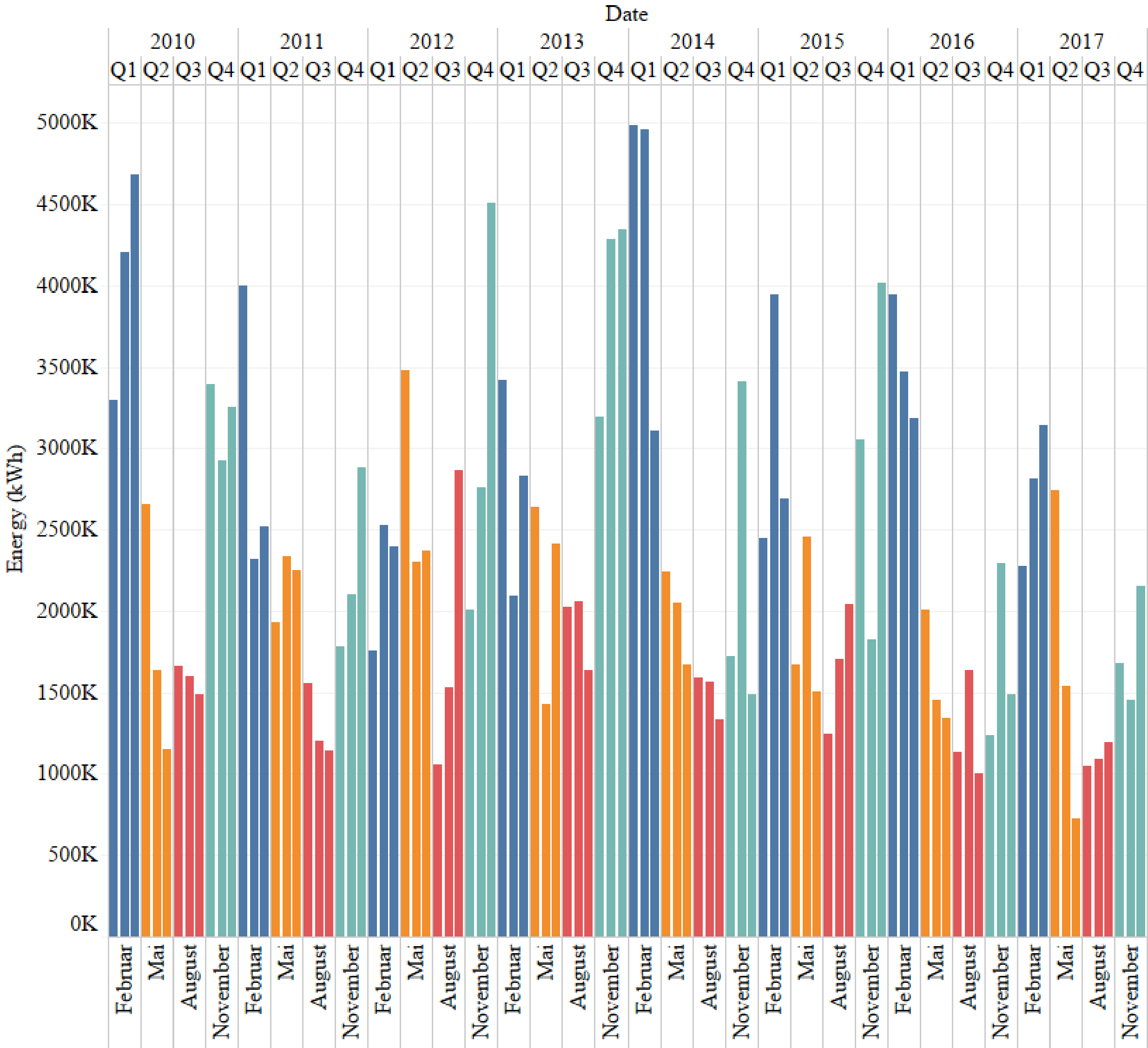

Figure 24: Variation of electricity production during years

Visualizing the energy output is an excellent means of helping the contractors to examine, in which months the production of electricity is higher and when it is less. Moreover, it can help to decide if a specific pattern of energy production can be chosen during the different months of the year.

Figure 24 shows the output electricity from the wind farm for seven years. Wind data are collected per day, but to find information, which operators need for the contracts, data are aggregated per year, quarter of season (Q1 to Q4) and month. The wind farm operator can infer from this plot that the maximum energy can be generated in spring (January until March). Moreover, in 75% of the cases, the energy produced in December shows to be higher than other winter months.

The energy production is minimal during the summer months. As illustrated in the figure, July and August are among the months with the least productivity.

Based on this information, the operator can design contracts by considering that the yield is lower for summer months. He may commit less energy purchase during these months. On the contrary, the operator is now aware that overproduction of energy happens from December to March. Therefore, a new strategy can be, accept energy supply for more centers during these months, temporarily, and sell the extra produced energy.

Nevertheless, the operator can get an initial insight regarding contract optimization from the visualization of PUI (such as Figure 24). To get more exact information, other models of data analytics can be implemented. In the following two techniques of data analytics are applied (see subsection 6.1.1) to find more potential patterns regarding variations of electricity.

6.1.3 Implementation of CART and RF for identifying factors affecting the variation of electricity

This section investigates data from 2010 to 2017 from the wind park (the first dataset, see subsection 6.1.1). They should facilitate the operator's understanding of the factors that produce changes in electricity production from his wind turbines. This section applies CART and RF for analyzing the dataset. The implementation follows the procedure described in 5.5.

Variables on the dataset are as follows: wind speed, direction, year, month, day, the season of energy production and the actual amount of energy produced. Analysis aims to determine, which of these variables have a more significant effect on the electricity produced by wind turbines. In addition, the analysis shows factors, which enable the operator to understand the pattern of variation in energy generation. For applying the CART model (as described in section 6.1.1), R software has been used. The analyzed data are from daily measurements. Data have been divided into train and test dataset. This division helps to build a better prediction model.

Figure 25 shows the results of the analysis with CART technique. The results reveal that among the mentioned variables wind speed has the most influence on the generation of electricity at this location. With the increase in speed, the amount of produced electricity has also increased.

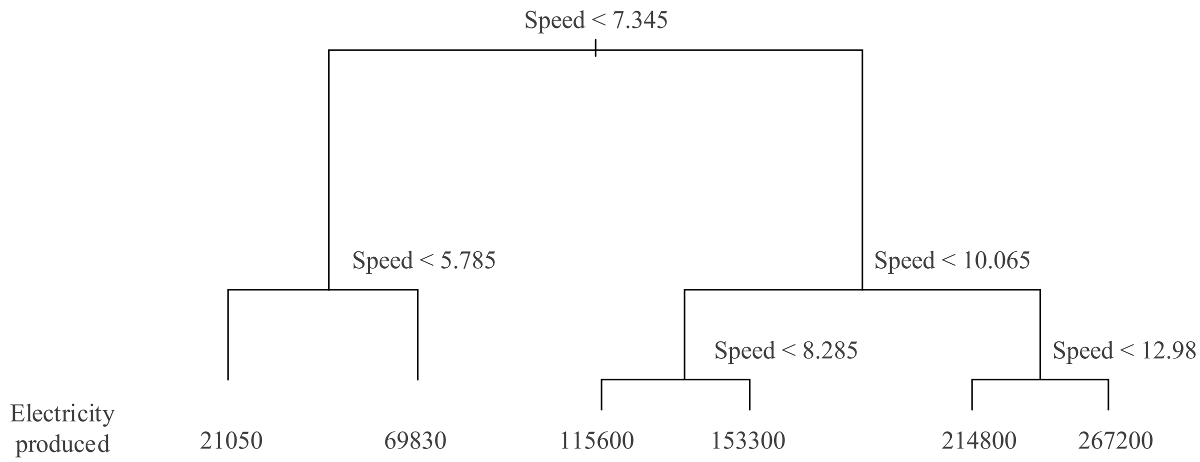

Figure 25: Tree model shows that variation of output electricity in year depends on the speed of wind

Results of CART model is also reported in Table 18. In this figure, "split" shows the condition on or partitioning of data. It is decided by the CART algorithm. The letter "n" means the number of observations (here days), which meet the split condition. Also "yval" shows the estimated output electricity. The results indicate, that in most of the days (1511 out of 2177 days), speed is less than 7.435 m/s. In these cases, the average produced electricity is 37740 kWh per day. Moreover, only in 237 days, the speed was higher than 10.065 m/s. In these days, the average produced electricity is 228700 kWh per day (node 7), which means surplus electricity (more electricity than demand) has been produced.

Table 18: Results of regression tree

Node) split, n, yval	*** denotes terminal node**
1) Root, 2177,77740	
2) Speed < 7.435, 1511, 37740	
4) Speed < 5.785, 994, 21050 *	
5) Speed > 5.785, 517, 69830 *	
3) Speed > 7.435, 666, 168500	
6) Speed < 10.065, 429, 135200	
12) Speed < 8.285, 206, 15600 *	
13) Speed > 8.285, 223, 153300 *	
7) Speed > 10.065, 237, 228700	
14) Speed < 12.98, 174, 214800 *	
15) Speed > 12.98, 63, 267200 *	

This information arises the next questions; a) what are the characteristics of days (or months) in which speed of the wind is higher than normal? b) Are there other factors that affect higher speed ranges? To answer these questions first insights can be extracted from Figure 24. As shown in the diagram, higher electricity production occurs between December and April. However, to better test this insight another decision tree is built. Therefore, the next decision tree aimed to investigate the following hypotheses.

Hypothesis 1: Factors such as time of year and direction affect the speed of wind location of this wind farm.
Hypothesis 2: Speed of wind is higher during months December to April at the location of this wind farm.

To find out if these hypotheses are true, the next tree model is built. Following paragraph reports the results of this tree model.

Classification tree for wind speed. Speed data have been divided into ranges; 0-5 m/s, 5-10 m/s, 10-15 m/s and 15-20 m/s. The purpose of this division is to classify the factors affecting higher wind speed (H1). These classes are named as “speed range” and are included as a new variable in the dataset. In the following, another tree model is built that takes into account the new variable “speed ranges”. This model also uses CART algorithm. However, it is set up to build a classification tree.

Table 19: Results of classification tree for wind speed

Node) split, n, yval (yprob)	* denotes terminal node
1) root, 2906, 5-10 m/s (0.306 0.584 0.098 0.009)	
2) Month: April, December, February, January, March, November, October, 1688, 5-10 m/s (0.256 0.584 0.142 0.016)	
4) Direction < 170.5, 811, 5-10 m/s (0.283 0.642 0.072 0.001) *	
5) Direction > 170.5, 877, 5-10 m/s (0.231 0.530 0.207 0.030)	
10) Direction < 286.5, 704, 5-10 m/s (0.191 0.522 0.247 0.038) *	
11) Direction > 286.5, 173, 5-10 m/s (0.393 0.560 0.046 0.000) *	
3) Month: August, July, June, May, September, 1218, 5-10 m/s (0.376 0.586 0.036 0.000) *	

Table 19 shows the results. In this table “yprob” presents the probability that observations can be categorized under each of four classes of “speed range”. Table 19 shows that in most of the days (1688 days out of 2906) between October and April the speed of the wind is 5-10 m/s. Only in 704 days, when the direction of the wind is between 170.5^{o} $<$wind direction $< 286.5^{o}$, the probability of stronger winds increases (Probability 0.247). By utilizing this tree model (Table 19), it is possible to identify the months and directions in and from which the wind blows higher. Namely, from December to April, the amount of produced energy can be higher (answer H2). In addition, with regard to H1, it is possible to state that the direction of the wind affects the wind speed. However, the year and daytime did not show any statistically significant effect on the speed.

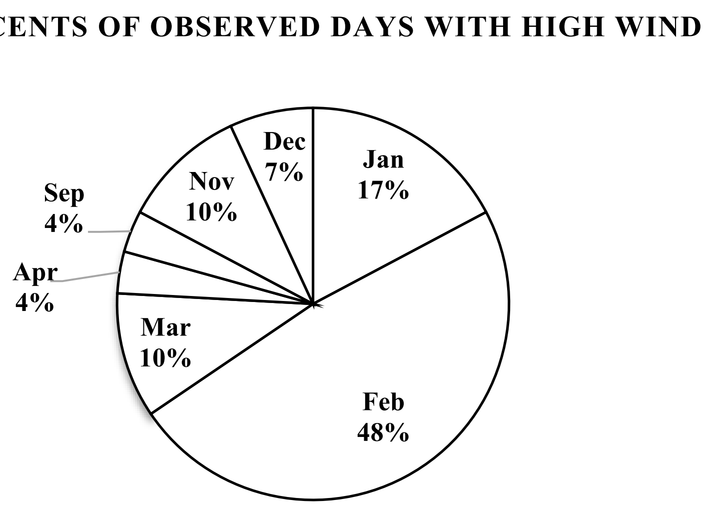

Figure 26: Months with a higher frequency of strong wind (speed 10-15 m/s)

Another analysis was performed to precisely select the months with higher wind speeds (and higher energy production). Based on Figure 26, high winds blow less in September and April (only 4%). Moreover, in October no high-speed wind (10-15 m/s) has been recorded for the past seven years in this location. Therefore, it can be inferred that the months, with surplus energy production are November, December, January, February, and March. With this information, the operator can adjust the contract. Evaluation of tree models is reported in section 7.1 and in Appendix F.

Model accuracy. For the CART regression tree, the RMSE (see section 6.1.1) between the real values and the predicted ones as well as absolute error are 42947.1957 and 10376.13, respectively.

These values are compared with the results of the random forest model. This comparison aims to assess models and select the more accurate ones. Next part explains the results of RF technique.

Applying RF. In order to test H1 with another method, Random Forest (RF) technique was selected. The aim is to provide the best test of H1 and H2 and to obtain the best possible algorithm for the operator of the wind farm. This section reports findings of RF model. In the end, it compares the results with the CART (Table 18).

Implementing RF to predict energy (regression). Results of applying RF are shown below.

Call:

```
randomForest(formula = Output.Energy ~ ., data = WT6, importance = TRUE)
```

Type of random forest: regression

Number of trees: 500

No. of variables tried at each split: 4

% Var explained: 89.64

Figure 27: Results of RF on electricity generation data

Results of the RF model show a higher accuracy rate in comparison with tree model. Figure 27 visualizes these results. RF model is built based on 500 trees. From Figure 27 it can be seen that 89.64% of variations in data have been captured and could be modeled. This rate shows that the model was successful in taking into account different characteristics of data. Therefore, it is statistically probable that it can provide an accurate prediction of output electricity. Evaluation of prediction power of this model is reported later in this subsection.

Figure 28 shows the variable importance. The RF model recognizes speed as the most critical variable in energy generation and fluctuation. The finding approves the results of CART tree (Figure 25). Based on the RF model, year, direction and month (i.e., June) affect electricity prediction (answer to H1).

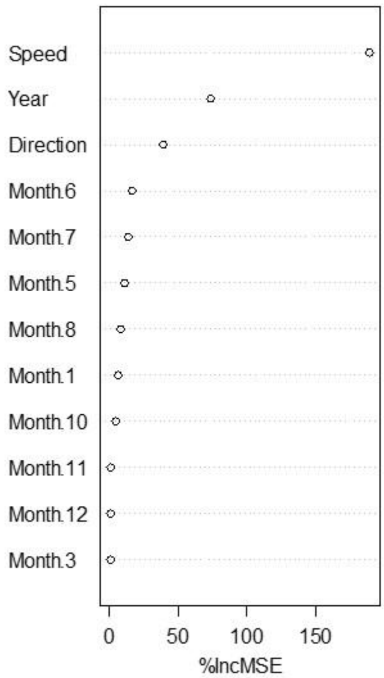

Figure 28: Variable importance recognized by RF model

Evaluation result for RF shows absolute error and RMSE as follows. Mean absolute error is 26039.882 and RMSE: 6087.036.

Compared to the tree model, the error rate for our random forest model is less. Therefore, it is recommended that operator use this model instead of CART regression model (Figure 25).

6.1.4 Incorporating maintenance data

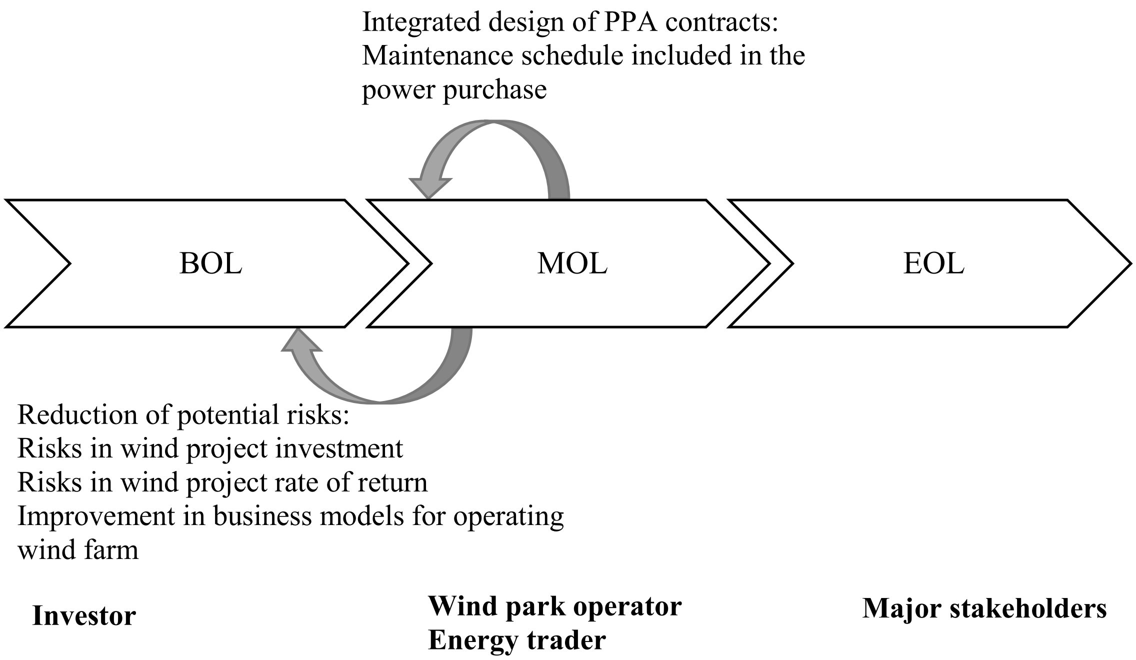

Figure 29: Improvement in contact design for related beneficiaries in different lifecycle stages

Figure 29 shows possible improvements in contract design for related beneficiaries. Here we focus on integrating the maintenance information into contract design. In this context, the wind park operator requires information about the amount of electricity that can be produced by considering weather and maintenance data.

The analysis in this section uses the second PUI dataset mentioned in 6.1.1. It shows the produced electricity of a wind farm from January 2016 to the end of August 2017. The dataset reflects measurements with 15 min intervals. The variables in the dataset include date, hour, speed in different heights, wind directions for different altitudes, output electricity, and available capacity. Maintenance schedules are reflected in terms of available capacity in the dataset[4].

Based on experiments with the models as shown in 7.3.2-7.3.3, the random forest model can perform well on the prediction of electricity. In the following subsection we test whether RF can perform well when maintenance data are included.

Applying random forest. An RF model with 1000 trees has been build. The procedure of model building is similar to 6.1.3. The model shows an acceptable prediction accuracy of 93.31 %. It can be concluded that the model has covered most of the variation in data. An error analysis yielded a mean absolute error of 10618.57 and an RMSE value of 6995.667.

Therefore, another information need of operator for contract design can be met through incorporating maintenance schedules and performing such predictions. For example, when we know in advance about a possible maintenance plan on 3% of turbines, the operator can take into account that energy production is at least 3% less than average supply value.

In summary, a real-world realization of MOL Stakeholders' data needs is investigated in section 6.1, from a technical perspective. The scenario addresses the need for wind farm operator to optimize electricity contracts. Product long-term performance (7-8 years), wind farm location-related data, as well as maintenance have been analyzed to provide insight into the better design of product-service-sell contracts.

6.2 Scenario 2: Improving OEM decisions for managing Electric Vehicle (EV) components

The second implementation scenario is conducted on EV. Although EVs can be produced as mass production, we include them in our study. One reason for choosing EV is that the trend of development of this product from the perspective of digitalization and developing new services based on PUI is similar to engineered products (section 2.1). Therefore, the concept of "MOL stakeholders data and information needs" (chapter 4) and supporting the information needs with data analytics tools (chapter 5) can also be tested and showed for improving performance of EV, similar to an engineered product. For a more accurate analysis, a study has also been conducted for the identification of MOL stakeholders of EVs. The results of this study are presented in Appendix D. Appendix D explains that the main MOL stakeholder groups, which were found in subsection 4.4.4 and ones that can benefit from PUI

[4] Dataset has been published by Bremen Big Data Challenge 2018 competition, https://bbdc.csl.uni-bremen.de/.

data, shows to be similar in both EV and engineered products (such as wind turbines and airplanes).

In the following scenario, a data need of OEM, as a major MOL stakeholder (Figure 16), is selected. The selected data need is "product performance", as presented in Table 11. Supporting these data need with data analytics techniques has been addressed. The techniques have been selected from the category of asset analytics (Section 5.1). Subsection 6.2.1 presents the scenario in detail and explains the necessary implementation steps for data processing and supporting the information need of this stakeholder.

6.2.1 Information needs of OEM: Pattern recognition for assessment of EV's battery lifetime

This subsection shows a scenario for identifying relevant information for OEM in the area of electric mobility by applying asset analytics (as described in chapter 5) on the PUI data of EVs.

Figure 30 shows a schema of the scenario. PUI is available from MOL of an EV. Stakeholders, such as OEM, suppliers and recyclers (shown in the figure) can benefit from these data.

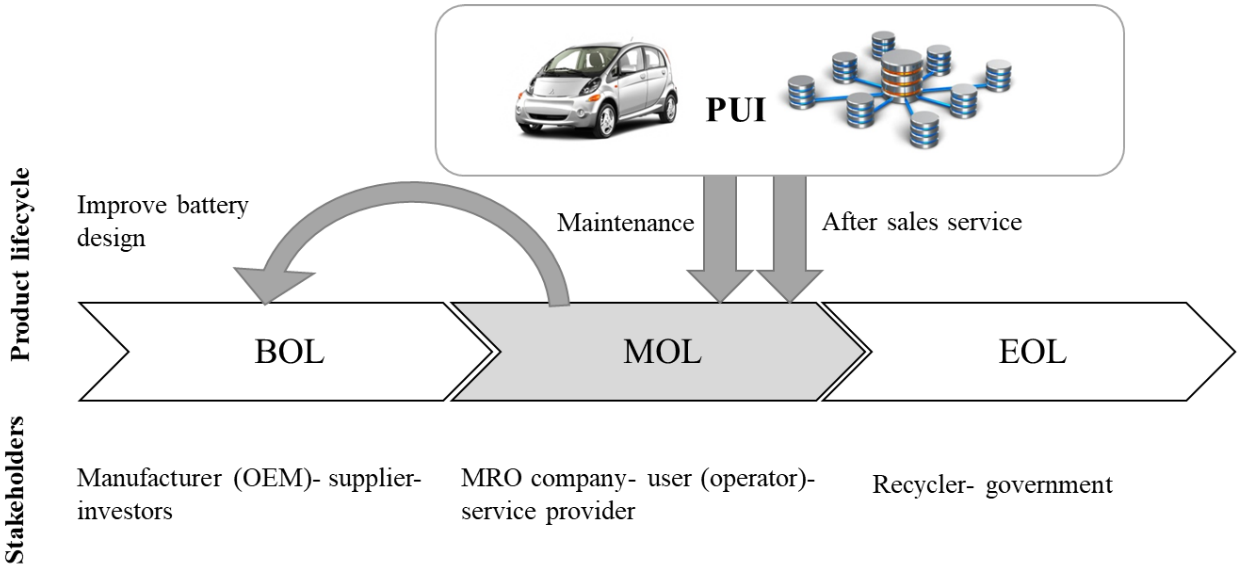

Figure 30: Sources of MOL data and information for scenario of OEM information needs

An EV manufacturer, hereafter called OEM, wants to improve lifecycle management for the produced vehicles. For this reason, it is crucial for the manufacturer to know the performance of products under different conditions of use and to get more information about it. Understanding product performance helps OEM develop better strategies for after-sales service and repair and maintenance. Because the battery in an EV is the storage area for electric power and the major source of energy for the movement of the EV, ensuring the correct operation of this component in EV is very important for OEM. To this end, asset analytics for analyzing product data to find useful information for the OEM has a high priority.

OEM, having information on the battery life of the car, can not only improve the after-sales service and maintenance but also can share this information with designers of EV to help them improve their design and battery life.

One of the technical information that the OEM needs to know is about the battery life of the EV. The battery life of the automobile varies according to the pattern of use and environmental conditions, under which the EV operates. Based on the results of previous research (Taefi, et al., 2016), the battery lifecycle varies with parameters such as air temperature, driving behavior, number of brakes, number of stops, and path topography. Over time and by repeated use of the battery, the amount of time that the battery keeps its charge decreases. Thus, an important piece of information from MOL of EV could be to monitor the performance of the battery and to know the actual rate of battery degradation. The higher the real battery discharge rate, the faster the battery is worn. Therefore, it is a sign that the battery needs to be replaced.

By designing sensor systems on the machine, the manufacturer has been able to gather up-to-date information about the performance of its vehicles, which are at the customer's disposal (use). However, information required for OEM from these sensor systems with data analytics is not fully exploited. Although several models of battery charge have been developed to calculate battery cycle and its worn-out pattern, the models that are based on PUI sources are still not fully developed; and therefore, not commonly used by OEMs to learn about battery performance. Among these data, there is information such as driver change, starts and stops, outside temperature, battery charge and charge frequency (Taefi, et al., 2016). Now, OEM wants to predict the lifecycle of the battery. For this purpose, in the first stage, the actual battery discharge rate under different usage and environmental conditions should be anticipated.

Table 20: Summary of specifications of scenario

Role of stakeholder	Product manufacturer (OEM)
Data need	Product performance
Characteristics of available PUI	Sensor data (in quantitative form)
Type of analytics	Asset analytics
Aim of analytics (application)	Prediction
Algorithms	SVM- NN
Reasons for choosing SVM & NN methods	Because of ability to handle high volume of data- several measurements- high prediction accuracy- robustness against missing data

The proposed methods in chapter 5 for asset analytics can be used for this purpose. Since the problem deals with prediction and depends on several variables, the following algorithms can be used; Support Vector Machines (SVM) and Neural Networks (NN) analysis (as described in chapter 3 & 5). Table 20 shows a summary of the technical specification for this test scenario.

6.2.2 Sources of information and dataset description

In this dissertation, data come from a CAN-BUS sensor network. Sensors in the CAN-BUS system can collect the data of the EV when it is in operation. The size of the dataset from CANBUS system is 436 MB which contains a large number of records (four log files with 1048573 rows, 250813 rows, 10645 rows and 1048576 rows).

Data on EV usage and weather. The EV understudy has an automatic gear system with motor power 49 KW and 16 kWh 330 V Lithium-ion battery. Captured data from CAN-BUS system are data related to date and time of drive, odometer, data about usage of electricity inside the EV (e.g., lights, air conditioning), speed, braking, route information and motor power. Moreover, data of weather condition, such as outside temperature, rain and wind speed are available.

In the following initially, SVM and NN are shortly introduced. As mentioned before, asset analytics algorithms for product performance forecasting and degradation analysis are used here.

Support Vector Regression. This technique is a well-known algorithm in machine learning and data mining domain. Support Vector Regression (SVR) is a version of Support Vector Machines (SVM), which is used for predictions. The mechanism of SVR is that it maps input vectors nonlinearly to a very high dimensional feature space (Wuest, 2014). SVR produces nonlinear boundaries by constructing a linear boundary in a large, transformed version of the feature space (Hastie, et al., 2009). By using this technique, it is possible to predict the values of one or more variable based on several other variables. Theoretical background of SVR is not explained here. The interested reader can refer to Hastie et al. (Hastie, et al., 2009) and Appendix C for more information.

Multi-layer Perceptron Neural Networks. Multi-layer Perceptron (MLP) is a class of feedforward artificial neural networks. This technique is well applicable to modeling a functional relationship (Günther & Fritsch, 2010), such as the relationship of electric battery discharge with the factors that affect it. As stated by (Günther & Fritsch, 2010) MLP consists of a directed graph. This graph is organized into at least three layers. Each layer has some nodes (neurons) and the layers are connected. Each neuron uses a nonlinear activation function. The output from each neuron receives a weight before it is transferred to the next layer. MLP utilizes a technique called backpropagation for training data (Rosenblatt, 1961). For technical information regarding MLP, please refer to (Günther & Fritsch, 2010; Bishop, 2006).

This research uses MLP with the backpropagation algorithm for training data in prediction of battery charge. The reason for selecting this technique is its robustness for modeling functional relationships. Note that in this research, before testing neural networks, a linear regression model was applied to the data. The results of the fitting linear model were poor. The results are available in Appendix E for the interested reader. This shows that the relation between battery charge and sensor measures does not have a linear form. Therefore, algorithms should be selected that can adequately model nonlinear relationships between data. Multi-layer perceptron is a suitable one for this aim. The next subsection, describes the modeling steps.

6.2.3 Calculation of Support vector regression and Neural Networks for predicting the battery charge.

In the following first procedure of modeling, the PUI is described and later the techniques of SVR and NN are illustrated.

Steps of modeling. The stages of modeling are similar to the procedure described in 5.5. However, for a clear overview of the problem more information is provided here. First, the log files were converted to tables and the measurements were ordered in columns. During preprocessing, data are aggregated together based on timestamp. The need to combine data is due to the fact that measurements of the sensor network in CANBUS system of the EV are performed at different time frequencies. Next, the relevant variables for this study are selected based on asking experts in the area of electric mobility. Moreover, statistical analysis for exploring dependencies among variables is performed. In addition, to enhance the richness of modeling, two more variables have been created; trip number and duration of the trip. The trip number is a variable, designed so that if the driver of EV has been changed the effect of this change can be recognizable from the model. After cleaning the data, the model was built on data. For developing the model, data were divided to train (2/3) and test (1/3) datasets.

SVR modeling. Next, the SVR model is built. Type of model used in SVR is eps-regression. eps-regression controls, the acceptable error rate of the model. The formulas of this error mechanism have been shown in equations (2) and (3); please refer to Appendix C. The model has several parameters. These parameters consist of (a) kernel type, (b) cost function, (c) epsilon and (d) gamma. The best combination of parameter values should be found in order to build a good prediction model. The process of finding the optimal parameters is called “parameter tuning”. In this dissertation, experimental results and optimization functions have been used to decide on the best values of parameters. The results of this parameter tuning are reported in Appendix C.

SVR model after tuning. After obtaining the best fitting values for parameters, (Appendix C), the SVR model was built. This model can support OEM for prediction of battery charge. Table 21 shows the results of prediction based on SVR model and results of prediction versus actual values. The error of SVR model (goodness of fit) are: RMSE=0.2909 and Absolute Error= 0.1079.

Table 21: Results of battery charge prediction with SVR on a data sample

SVR predictions	Actual Values
31.052	31.833
38.746	40.833
61.299	61
69.441	67.167
31.886	36.833
73.58	74
37.772	38.571
85.621	83
50.215	50.75
54.967	54.833
80.724	79.643

Predicting battery charge with neural networks. Analyzing a model with neural networks also follows steps similar to those of the SVR model, as mentioned early in this subsection. Model of NN was built by trying different weights on synapse and various number and combination of layers (hidden layers). Figure 31 shows the final network with input variables and weights. This network has three hidden layers. The prediction results for neural networks training model is reported in Table 22. The results are compared with actual values of battery charge. The absolute error of this model is 0.170 and the RMSE equals to 0.149.

Comparing the results of both SVR and NN shows an acceptable level of accuracy in model building. The evaluation of the accuracy of the NN and SVR model is performed in chapter 7. For this evaluation, 10-fold cross-validation is used to evaluate the best for each model. Based on the comparison of errors, SVR algorithm had a better prediction accuracy for this scenario. Therefore, building a model based on SVR can be recommended to OEM.

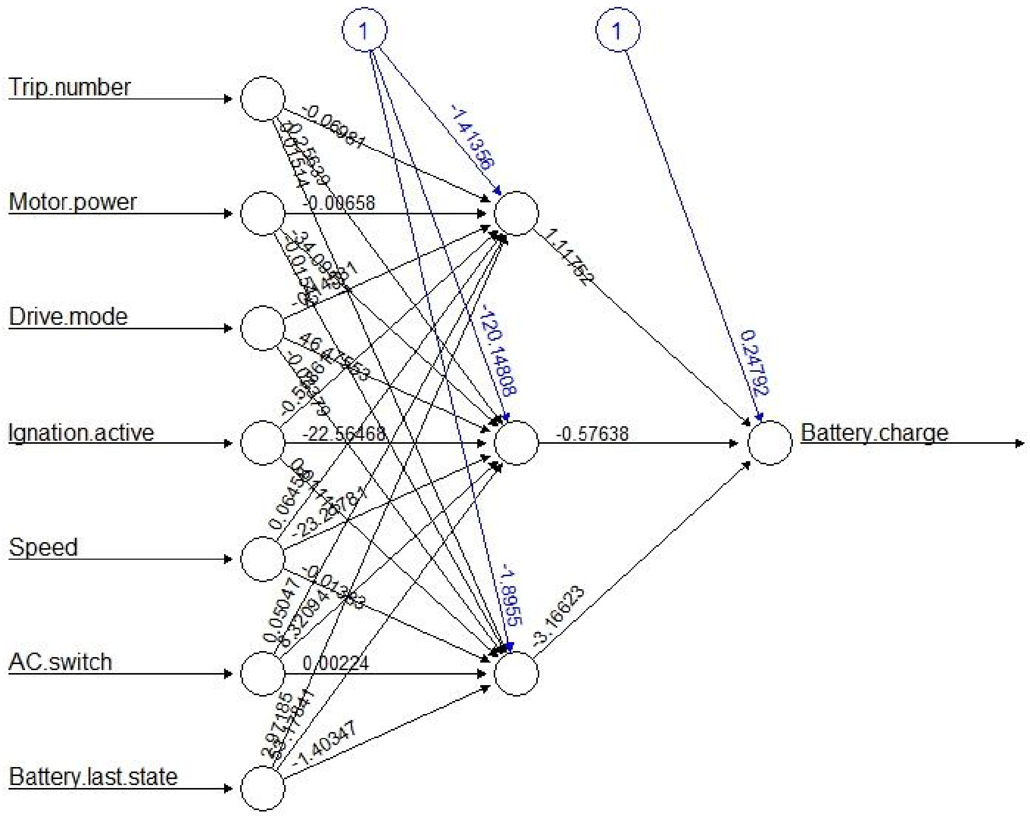

Figure 31: Neural networks implementation on EV data

Table 22: Results of NN prediction

Predicted Values (NN)	Actual Values
94.755	97.333
82.388	97.25
93.926	97
87.345	87.25
86.14	84.857
84.161	83.75
82.734	82.75
78.45	81.5
79.104	77.5
95.093	76.68
81.852	99
94.269	98.25
93.273	97
89.569	94.5
86.74	88.5
85.081	84.5

In summary, this section showed the model building with SVR and NN, based on PUI for predicting the performance of the battery. Table 23 shows the predicted values of both SVR and NN for the same observations (These are trained data). It is possible to compare the results from the two different prediction models. Moreover, the prediction models can be compared with actual values to assess the accuracy of data analytical models.

Table 23: Battery charge prediction results- real vs. NN and SVR

No.	Support Vector Regression prediction	Neural Networks prediction	actual Values
1	95.781	83.233	98
2	94.496	94.755	97.333
3	96.412	82.388	97.25
4	94.791	93.926	97
5	85.041	87.345	87.25
6	85.935	86.14	84.857
7	84.986	84.161	83.75
8	83.897	82.734	82.75
9	80.582	78.45	81.5
10	79.713	79.104	77.5
11	84.416	95.093	76.688
12	96.798	81.852	99
13	92.222	94.269	98.25
14	94.789	93.273	97
15	92.295	89.569	94.5

Taking the exhibited calculations into account, in this scenario, a battery charge prediction for EVs was carried out. The prediction models can later be completed and integrated into a virtual model for the status of battery charge (which can be run under virtual twin concept). With monitoring the decrease in the percentage of charge, which this battery can keep, decisions for maintenance and service can be made on a more real-time basis.

This model can also provide the basis for developing a new product-service application. The EV user can observe and directly use prediction of battery charge of the EV he is driving while OEM can be able to design more customized and more precise models for EV after sales service from forecasts of battery charge for thousands of machines. They can conclude the performance of batteries in its produced EVs by summarizing and making ensemble models of battery charge based on several devised performance states (behaviors).

What is the average warranted time of the battery? Between 8 to 10 years (or 100000 miles). Losing 10 percent of battery capacity in ten years is acceptable, but 40 percent is an issue. In our case, OEM currently does not offer a warranty for battery of EV. For our experimental EVs, the battery still can be used with 100 percent capacity. OEM can run the mentioned prediction models every year to monitor the capacity decrease in batteries. Moreover, as studied in this scenario, the behaviors of drivers such as the frequency of brake usage, duration of charge, weather condition can be tested by using our model.

6.3 Scenario 3: Improving MRO information needs for spare part planning

In this scenario, a leisure boat as another engineered product has been selected. Main stakeholders in MOL of a leisure boat have been identified in Wuest et al. (Wuest, et al., 2014). The finding is similar to the stakeholder groups of chapter 4. Therefore, from the perspective of stakeholder groups, it is possible to apply the concept of stakeholder data needs (as developed in chapter 4) to the leisure boats. Application of data analytics techniques to support information needs of MRO, as a stakeholder from MOL of leisure boats, is tested in this section.

Description of scenario. MRO needs coordination and information on maintenance procedure and provision of spare parts. MRO should exchange information with two or three other stakeholders to perform the maintenance task. The information needs (information exchanged) between MRO and operator encompasses the duration of maintenance, and services which should be performed, as well as spare parts, needed. Based on the finding of chapter 4, the following information exchange (Figure 32) should be carried out. Based on Figure 32, MRO should coordinate the time and quantity of spare part delivery with the supplier. Besides, MRO and operator should agree on a period during which the operation of the product can be stopped (in order to perform maintenance). Here we study the implementation of this information exchange with the help of data analytics.

If the product is functioning under use-oriented PSS, MRO should also coordinate with OEM to access and receive PUI; or the information that he requires. Otherwise, MRO can get PUI from the operator. In both cases, “operation analytics” techniques (section 5.1) can help to improve planning decisions for maintenance operations. Particularly these group of techniques can support improving planning and coordination tasks (see Figure 20 & Table 13). In the following, the case of improving MRO decisions to schedule leisure boat maintenance by forecasting the spare part consumption is examined. To this end, on-time information should be exchanged respecting the need for service of products.

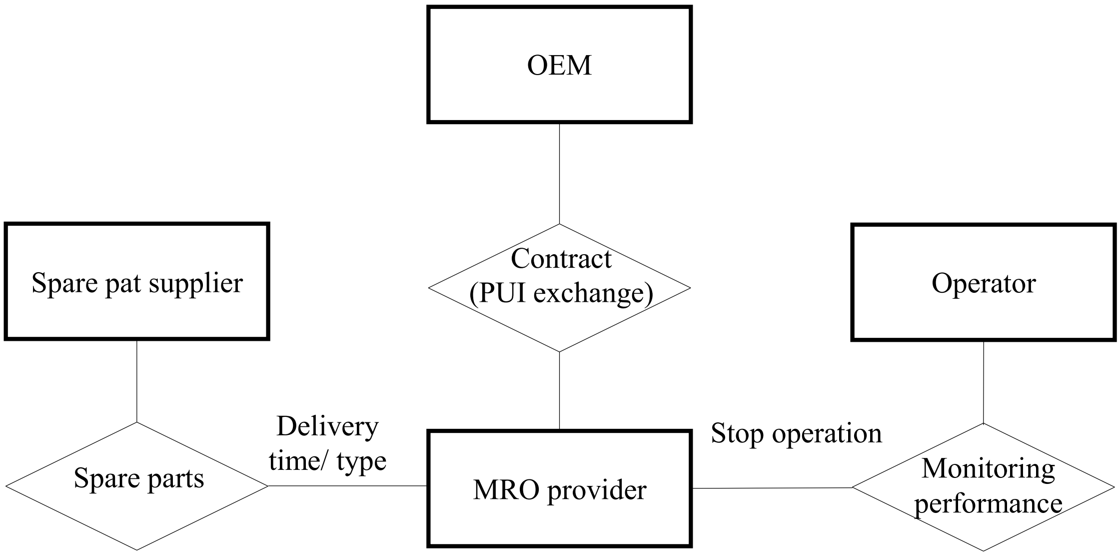

Figure 32: Information needs and exchange for MRO with supplier and operator during spare part planning

6.3.1 Planning spare part consumption for leisure boat

One of the MRO's information needs is planning the consumption of spare parts used in the maintenance and repair procedure. MRO must accurately anticipate the number of spare parts used; in order to prevent shortage during the operational work. Lack of spare parts can cause the workflow to stop. In addition, anticipating the demand for these parts is necessary to prevent incurring extra costs due to the storage of unnecessary spare parts as inventory. In this section, we examine another engineered product. This product is a leisure motorboat. Maintenance and repair of the boat are of great importance because, in the event of a downfall in the middle of the sea, the boat defects can jeopardize the life of passengers. The spare part is a major component that is used in the maintenance process of the boats.
Spare parts that MRO need can be categorized into two major types. Some spare parts that can be repaired and used again. These groups include boat rotatory components. Rotatory components include propeller and some electrical parts; such as relays and switches. Other groups of spare parts are consumable spare parts (Garg, 2013), such as fuel filters, oils, air, valves, and fuses. These parts are usually disposed after service. However, they are consumed more frequently and therefore; they are more on demand during maintenance task.

For the MRO Company, it is important that there is no shortage of spare parts in the repair process, especially in situations where a boat with failure is under service by MRO. In this respect, access to up-to-date information on boat performance (extracted from PUI) can be beneficial and practical. Here PUI and extracted knowledge from them is valuable. A series of boat PUI relates to the function of major components of the boat, for example, motor, propeller etc. In the event of a defect in a component, the spare part to be replaced can be determined by analysis of the PUI.

Namely, with access to data on boat operation and its analysis, when a boat failure occurs at the place of use (e.g. working at sea or during the boot-up), coordination between boat operator, MRO provider and supplier can be arranged. That is, as soon as the operator delivers the boat to the MRO, spare parts needed for repair are at the same time provided to the MRO.

Where the boat PUI is available to MRO, MRO realizes that a failure of the boat has occurred by receiving an alert. Therefore, the operator does not need to care for the maintenance of the boat. MRO orders the spare parts. In case that real-time data are shared with the spare part provider, MRO does not need to contact spare part provider. The spare part provider automatically receives a message regarding the amount and type of required spare part. The MRO task can begin at the earliest possible time for the operator.

In sum, in this scenario, we examine how MRO can forecast the demand for spare parts by having access to PUI and how it can provide other stakeholders particularly spare part supplier with the relevant information.

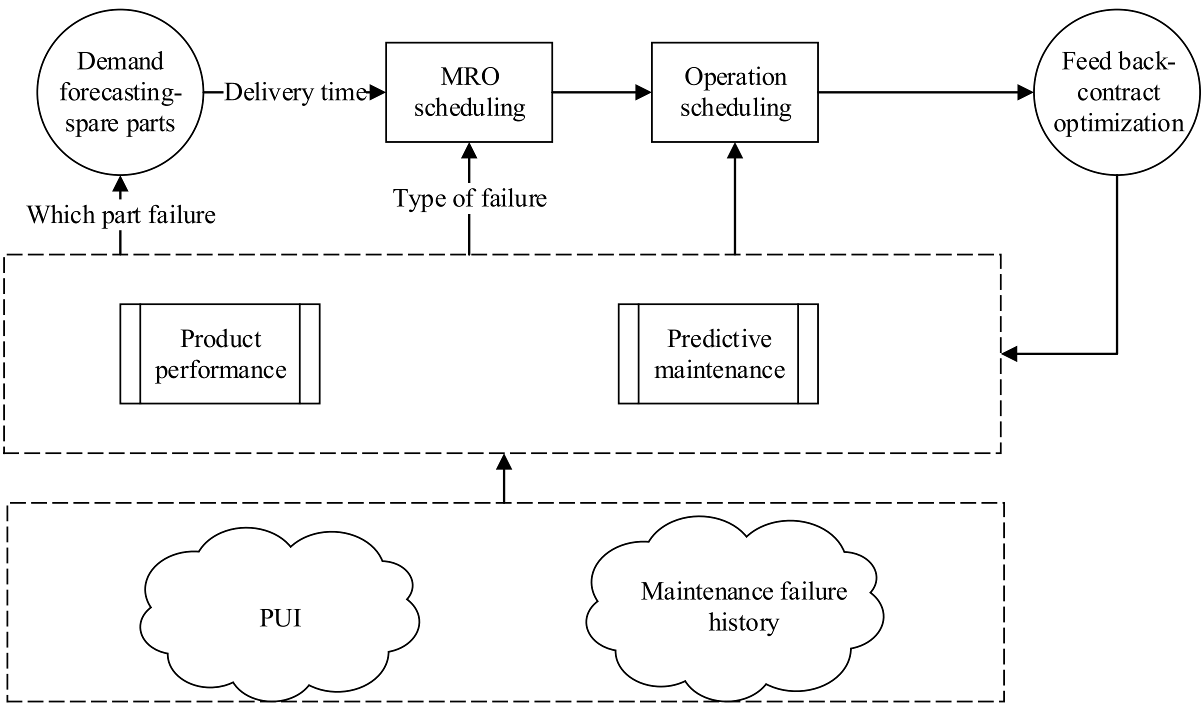

Figure 33: System under examination for scenario 3

Figure 33 shows an overview of the system under review. The goal is to predict the demand for spare parts, taking into account up-to-date information on product performance. Initially, the product is monitored to control product performance. The maintenance system predicts the probability and type of failure. Note that predictive maintenance systems were introduced in subsection 3.3.3 and are not the focus of this scenario.

Having at one's disposal the breakdown information, e.g., a defect type, and being informed in advance about the need for spare parts, the system can coordinate the spare parts for the MRO. If spare parts are not available in stock, an automatic order is placed to the spare part's supplier. Moreover, the information of maintenance service is sent to the boat operator. This information includes a warning indicating that the boat should be serviced. All the information exchanged between the MRO, operator, and OEM is eventually entered into the contract management system. This information is used to optimize future contracts between the operator and the MRO.

6.3.2 Optimization of the need for spare parts for motorboat repairs and maintenance

This section focuses only on spare part order and inventory management visualized in Figure 33. We examine PUI and application of data analytics to improve this module.

MRO checks the required spare part with the warehouse. He orders this piece. In order to plan and determine, e.g., how many pressure switches have to be ordered, MRO analyses the past data and observes that the use of this piece is randomized. The following figure shows the randomizing nature of consumption of this part during the last three months (Figure 34).

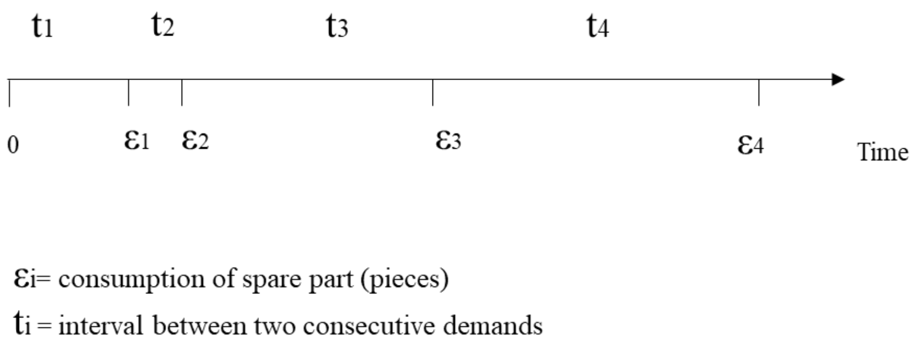

Figure 34: Random nature of spare part consumption (source (Garg, 2013))

In the next step, the forecast is presented in more detail using the techniques mentioned in Chapter 5. To do this, we propose here the time series prediction techniques and neural networks. According to previous studies, these techniques are suited to predict the demand for spare parts, and their use can have higher accuracy than other algorithms.

Table 24: Summary of specifications of scenario 3

Role of stakeholder	Maintenance Repair Overhaul provider/ Spare part supplier
Data need	Spare part demand
Characteristics of available PUI	Spare part specification and priority, location-related data, supplier data
Type of analytics	Process analytics
Aim of analytics (application)	Prediction
Algorithms	Time series - NN
Reasons for selection of Time series & NN methods	Because of satisfactory accuracy to demand forecast

6.3.3 Description of data

In total, the datasets contain 7673 records from the spare part inventory management system. The data show spare part consumption for ten months. During these ten months, the information such as the amount of spare part used for the maintenance task, type of spare part, customer ID and the supplier of spare parts were collected. Some attributes such as boat model and customer ID was excluded from this analysis because they didn't show any influence on the amount of spare part consumption. A new variable was added instead, which shows the criticality of a spare part. If the spare part is already available in the warehouse, the criticality is low. If the spare part is not available in the warehouse and previously a failure alarm is reported from a boat, the criticality and need for this part are marked with "Critical". If the part is available, but the inventory is less than average consumed per day, the criticality is Medium. If else, it is considered to have low criticality. This data can be a type of PUI and captured via RFID, PEID or sensors.

Note that this dataset contains hypothetical data. It means that data are simulated based on a real-world dataset, but it does not come directly from real-world practice. Because the issue of automating spare part prediction is still new, no experimental dataset could be found from previous research. Table 25 shows the variables and a part of the values in the dataset.

The goal is to better predict the type and amount of required spare part for the next coming months. To this end, two data analytics techniques are examined here. One is time series forecasting and the other is neural networks for time series.

Table 25: Sample of PUI regarding spare part consumption

Date	Supplier	Order (Part) criticality	Spare part name	Quantity
1/1/2014	Supplier 6	Medium	Coolant switch	5
1/1/2014	Supplier 6	Medium	Fuse	2
1/2/2014	Supplier 3	Medium	Thermostat	1
1/3/2014	Supplier 6	Critical	Cylinder liner	1
1/3/2014	Supplier 6	Low	Fuel filter	4
1/3/2014	Supplier 1	High	Gear filter	4
1/3/2014	Supplier 8	High	Piston ring	3
1/3/2014	Supplier 8	High	Cylinder head	1
1/3/2014	Supplier 3	Medium	Gear	7
1/3/2014	Supplier 8	Medium	Fuel filter	1
1/3/2014	Supplier 6	Low	Gear filter	2
1/3/2014	Supplier 7	Medium	Coolant fluid	4
1/3/2014	Supplier 8	Medium	Cylinder liner	2
1/3/2014	Supplier 6	Low	Oil filter	2
1/3/2014	Supplier 6	Low	Gear filter	2
1/3/2014	Supplier 8	Medium	Piston ring	2
1/3/2014	Supplier 6	Critical	Oil pressure switch	3
1/3/2014	Supplier 8	High	Engine oil	1
1/4/2014	Supplier 2	Medium	Gear filter	1
1/4/2014	Supplier 2	Medium	Oil filter	2
1/4/2014	Supplier 2	Medium	Sea-water pump gasket	3
1/4/2014	Supplier 3	Medium	Engine oil	2
1/4/2014	Supplier 3	Medium	Piston ring	2
1/4/2014	Supplier 2	Medium	Engine oil	1
1/4/2014	Supplier 1	Medium	Oil pressure switch	3
1/4/2014	Supplier 2	Medium	Gear oil	1
1/5/2014	Supplier 7	Medium	Cylinder liner	8
1/5/2014	Supplier 4	Medium	Engine oil	4
1/5/2014	Supplier 4	Medium	Engine oil	3
1/5/2014	Supplier 1	Low	Coolant switch	1
1/5/2014	Supplier 7	High	Piston ring	3
1/5/2014	Supplier 7	High	Oil pressure switch	3
1/5/2014	Supplier 7	High	Gear oil	2
1/6/2014	Supplier 4	High	Coolant switch	3
1/6/2014	Supplier 8	High	Piston ring	3
1/6/2014	Supplier 8	Medium	Piston ring	2

6.3.4 Applying time series analysis and neural networks

In this scenario, we chose time series modeling. The reason for this choice is that the data contains temporal measurements. It means that the spare part demand pattern is time-dependent. In the following, we test if models based on time series can accurately and adequately predict the need for spare parts of MRO when having PUI.

Time series forecasting. Time series is a group of forecasting methods, which is predominantly applied in demand forecasting. Based on the characteristics of data as described in the last section, the forecast model should not only take into account the variables such as supplier, type of spare part and its priority but also consider that date and seasons can affect the spare part demand. For this reason, in this section, we test the possible existence of seasonality effects and trends in the data. This analysis is done by Tableau software. For information regarding these models and their application in Tableau, please refer to (Brockwell & Davis, 2002; Tableau, 2018).

Neural networks for time series forecasting (NNETAR). The neural network model used in this study is a feed-forward neural network, which has one hidden layer. Moreover, for time series data (e.g., seasonal effect modeling), this neural network uses an ARIMA (autoregressive integrated moving-average) model. It means it uses the lagged data (autoregressive data with a time lag). In this dissertation, we run this type of model by using a function from R software. This function is called "nnetar()" and it is available in the package "forecast". For more technical information regarding this function, refer to (Hyndman, et al., 2018).

Data visualization. First, in order to have better insight about the type of spare parts which are consumed more, a grouping has been carried out. Figure 35 visualizes the consumption of spare parts, per type. It can be observed from this figure that liquids such as engine and gear oil or coolant fluid are the most consumed. Next, some parts such as piston ring, sealing washer, and gear filter are the most used parts during maintenance and service processes.

Figure 36 shows the variation of spare part consumption over time. It is evident that without accurate and robust predictive tools the fluctuations and uncertainties of this consumption cannot be modeled.

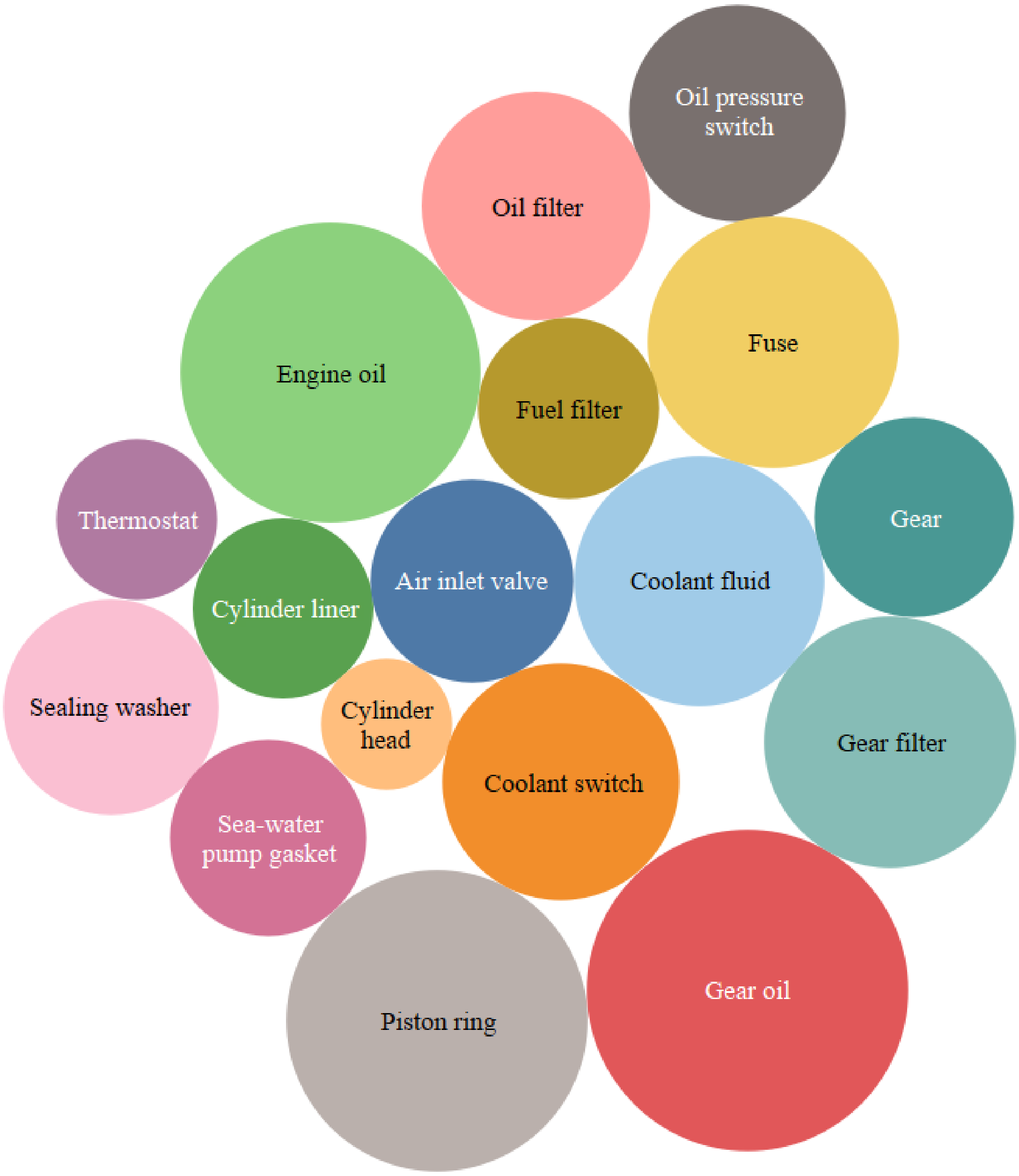

Figure 35: Spare part consumption per type

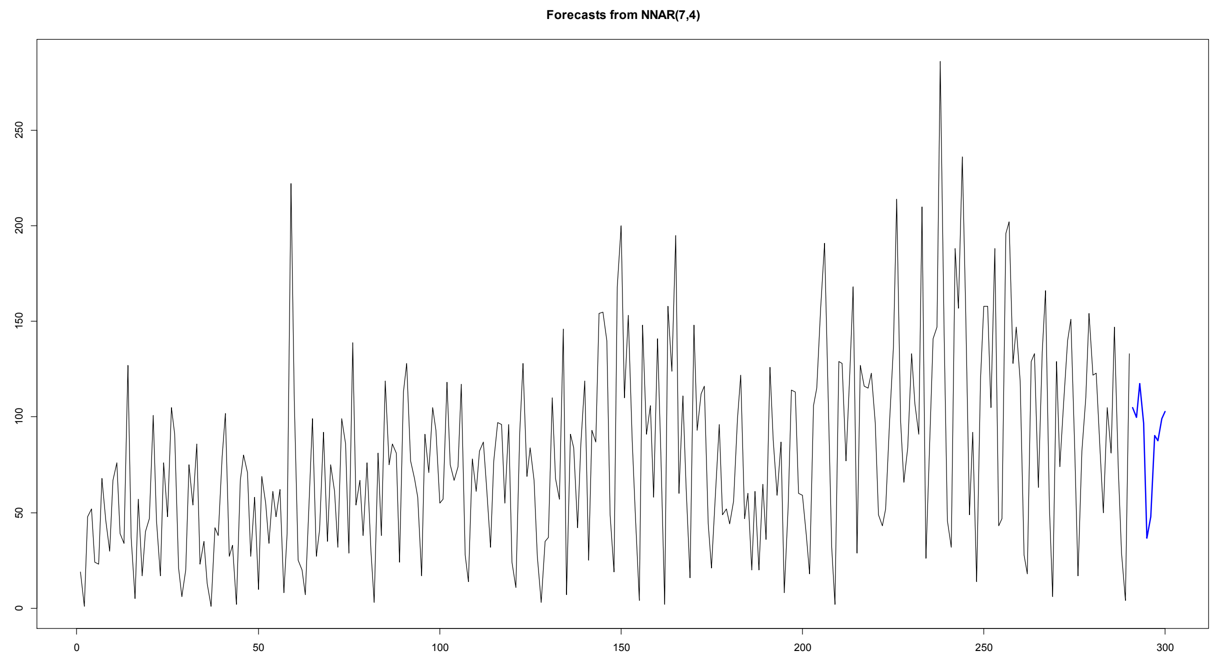

Figure 36: Variation of spare part demand in our scenario during the time

Time series forecast results. Figure 37 shows the results of forecasting for the next three months per spare part. MRO can use these results for better planning of the maintenance process. Moreover, the spare part supplier can benefit from the forecasts to manage its demand.

Spare part name	January 2014	February 2014	March 2014	April 2014	May 2014	June 2014	July 2014	August 2014	Septemb er 2014	October 2014	Novemb er 2014
Air inlet valve	68	57	59	72	95	199	88	125	137	145	154
Coolant fluid	73	117	128	178	189	207	126	226	214	223	233
Coolant switch	76	96	108	125	128	204	117	161	172	180	188
Cylinder head	22	23	31	26	57	89	23	47	58	61	65
Cylinder liner	35	51	69	83	81	134	32	70	95	99	102
Engine oil	201	98	190	256	233	296	161	288	275	285	295
Fuel filter	51	60	68	46	77	103	73	116	110	117	124
Fuse	123	74	133	138	142	178	116	285	240	259	278
Gear	74	45	75	47	94	149	65	140	131	139	146
Gear filter	74	66	153	128	130	238	108	221	203	214	225
Gear oil	173	152	236	228	244	303	269	340	344	365	386
Oil filter	48	60	93	121	161	184	131	208	212	230	248
Oil pressure switch	77	49	131	87	75	143	103	151	135	141	146
Piston ring	167	92	192	179	214	248	193	275	259	271	283
Sea-water pump gasket	79	39	66	103	82	101	89	158	144	155	167
Sealing washer	74	83	122	87	102	144	91	130	128	131	135
Thermostat	48	62	68	30	63	82	51	79	74	76	78

Figure 37: Actual and forecasted parts per month

Table 26 presents the predictions by the time series forecasting. Real values are also reported in this table. Therefore, it is possible to assess the power of the algorithm.

Table 26: Results of prediction by time series

Month	Spare part name	Quantity- actual	Quantity- predicted
Sep-14	Thermostat	95	74
Sep-14	Sealing washer	219	128
Sep-14	Sea-water pump gasket	152	144
Sep-14	Piston ring	404	259
Sep-14	Oil pressure switch	221	135
Sep-14	Oil filter	199	212
Sep-14	Gear oil	405	344
Sep-14	Gear filter	246	203
Sep-14	Gear	203	131
Oct-14	Piston ring	322	271
Oct-14	Oil pressure switch	175	141
Oct-14	Oil filter	222	230
Oct-14	Gear oil	427	365
Oct-14	Gear filter	208	214
Oct-14	Gear	159	139
Oct-14	Fuse	223	259
Oct-14	Fuel filter	107	117
Oct-14	Engine oil	282	285
Oct-14	Cylinder liner	97	99
Oct-14	Cylinder head	52	61
Oct-14	Coolant switch	195	180
Oct-14	Coolant fluid	171	223
Oct-14	Air inlet valve	161	145

the error of these predictions are RMSE= 57 and MAE= 14.13.

Calculation of NNETAR for predicting spare part consumption. In this subsection, another method for forecasting spare parts is tested. Moreover, the results are compared with the time series forecasts. As mentioned before, the second algorithm to be tested is neural networks for time series (NNETAR). For using this method R software and package “forecast” has been used. Table 27 shows the result of forecasting.

Table 27: Results of actual vs. predicted values for some of spare part types by NNETAR

Month	Spare part name	Quantity- actual	Quantity- predicted
Sep-14	Gear	203	139
Sep-14	Oil filter	199	178
Sep-14	Oil pressure switch	221	716
Sep-14	Piston ring	404	975
Sep-14	Sea-water pump gasket	152	144
Sep-14	Sealing washer	219	135
Sep-14	Thermostat	95	65
Oct-14	Air inlet valve	161	141
Oct-14	Coolant fluid	171	179
Oct-14	Coolant switch	195	179
Oct-14	Cylinder head	52	60
Oct-14	Cylinder liner	97	97
Oct-14	Engine oil	282	242
Oct-14	Fuel filter	107	102
Oct-14	Fuse	223	152
Oct-14	Gear filter	208	178
Oct-14	Gear oil	427	506
Oct-14	Gear	159	139
Oct-14	Oil filter	222	178
Oct-14	Oil pressure switch	175	716
Oct-14	Piston ring	322	975
Oct-14	Sea-water pump gasket	138	136
Oct-14	Sealing washer	169	135

Similar to other data models, the error associated with our fitting procedure has been calculated to assess the goodness of the fit. These errors make it possible to compare the results of this model (nnetar) with time series forecasting. The MAE equals to 101 and the RMSE is 34.5572. It can be seen that the errors of prediction for this model are higher than that of the time series model. Therefore, we selected the time series prediction (refer to Table 26) for this case. It should be noted that there was no need to standardize the data in scenario 2 (not necessarily required for the algorithms). This is one reason that the values of error are higher compared with those of scenario one and two. In section 7.1, error calculation is performed on normalized data of this scenario. In this context, it is possible to compare the error rate of other scenarios with the one currently discussed.

It can be concluded from this scenario that MRO can use time series modeling to forecast the spare part demand and improve planning maintenance by using PUI.

This type of control for spare part consumption also makes it possible to design customer specific maintenance intervals.

6.4 Conclusion drawn from the concept implementation

As presented, the approach in this dissertation provides the stakeholders with the information they need about the product usage and support them in decision-making. The scenarios in this chapter considered several stakeholders' information needs. OEM, MRO and operator were considered as the major stakeholders and their information need about optimizing the electricity sell (wind turbines), product performance (EVs), as well as demand of spare parts (leisure boats), have been addressed.

In conclusion, we can learn from these three scenarios that the application of the identified information requirements and solutions based on data analytics (chapter 4 and 5) are applicable in real life. The situations described in this chapter, contribute to this overall goal of identifying and meeting information needs along the product lifecycle, by showing the realization of a specific information need of stakeholder from PUI (concept in chapter 4) and supporting this understanding with the appropriate application of data analytics techniques (chapter 5). Next chapter validates the concept and data analytical models.

7 Validation and Discussion

In this chapter, validation for the outcome of chapter 4, proposed categorizations in chapter 5 and the evaluation of data analytics models from chapter 6 is presented. The validation results are also discussed.

7.1 Validation of the conceptual model

The goal of the concept in chapter 4 and 5 was to support the development of the concepts adjunct to the concept of CL-PLM, which can be used for organizations in nowadays real business environment and can help them in digitalization as well as better managing the data, processes and stakeholders of their product or service. For checking the validity of the concept and its usability both from the perspective of academia and industry, two approaches have been selected and applied. The first type of validation used to examine the concept is operational validation. Operational validation is chosen, because it can assess if the concept can produce correct results after it is put into operation. The second approach to validate the concept is, comparing it with the findings from the current state of the art research. In the following, these two approaches are discussed.

As the first validation approach, operational validation is used. Operational validation has been introduced in chapter4 (section 4.2). To validate our concept operationally, the identified stakeholders and their information needs were tested on two other engineered products. Namely, the concept of stakeholder's information needs in chapter 4 was developed based on two engineered products, airplane and wind turbines. For the validation, two other products EVs and leisure boats were considered. Appendix D shows the MOL stakeholders of EVs and compares them with the findings in chapter 4. The results show that for most of the stakeholders, the roles, major data and information needs and the possibility of data analytics on PUI are similar in both cases.

Moreover, to perform operational validation, the scenarios of chapter 6 are designed in a way that they consider not only different engineered products but also different stakeholders (operator, MRO and OEM), diverse data needs and various categories of analytics. Particularly, the use of asset analytics for the scenario of EV, operations analytics for the scenario of spare part management in leisure boats and business analytics for the scenario of wind turbine have been addressed (chapter 6). All the cases showed a satisfactory modeling result, which has the potential to support stakeholders also in the real-world application. It can be concluded that the concept was validated for use in proposed operational areas of the lifecycle of engineered products.

In addition, three experts have reviewed the conceptual model of MOL stakeholders, their relations, data and information needs (as presented in chapter 5). Expert's qualifications are as follows: First in the area of data analytics (6 years' work experience), product service systems (more than ten years' work experience) and closed-loop product lifecycle management (8 years of work experience). Moreover, the main part of the concept (from chapter 4) has been published the Journal of Product Lifecycle Management.

For the second validation approach, this dissertation discusses the validity of the concept by a critical discussion and analysis of the relevant methods regarding the use of data analytics from the stat of the art, as follows. The following questions regarding the concept are considered and their answer was investigated in state of the art: (a) Has the processing of PUI an impact on the decision-making of stakeholders at the product MOL? (b) If the answer to this question is positive, the question arises how the increase of data quality, can positively affect this decision-making, after applying appropriate data analytics tools?

The reason for formulating validation question into (a) and (b) is that the concept of MOL stakeholders' data and information needs is designed to support the stakeholders in decision-making. Moreover, this support should be enforced by bringing more added value to the decision makers by providing them with information about product use (PUI or product MOL data and information).

To answer (a), the following arguments should be made. The research works, which were found in the literature regarding the relation of data analytics and improvement of organizational decision-making, as well as the use of operational data of the engineered product, include three surveys and some journal articles. The field survey results of Ghasemaghaei et al. (Ghasemaghaei, et al., 2018) identified factors in improving the quality and effectiveness of decision-making using data analytics. In addition, they determined the effect of these factors in decision using a quantitative model. Their results from their study are shown in Figure 38.

The findings of Ghasemaghaei et al. (Ghasemaghaei, et al., 2018) show that data analytics can positively affect decision-making. Having data analytics competencies can actively improve decision-making (factor of 0.883). Data analytics competencies are data quality, analytical skills, tactical knowledge, and the level of complexity of the tool. According to Ghasemaghaei et al. (Ghasemaghaei, et al., 2018) all these factors affect the quality and effectiveness of stakeholders' decisions. In addition, the size of datasets affects the quality of decision and reduces uncertainty, but this does not necessarily increase the effectiveness of the decision. This research was conducted on 151 companies (see below).

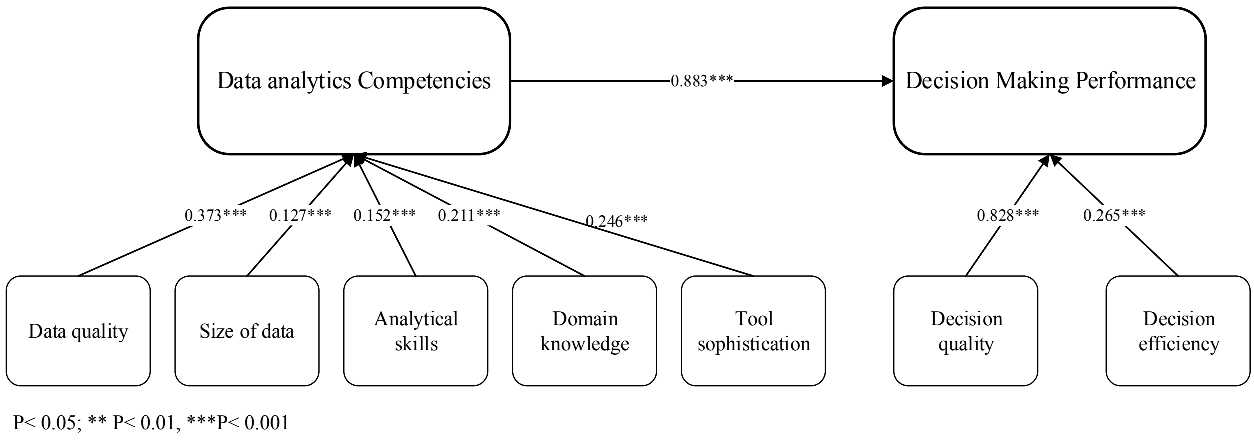

Figure 38: Factors affecting decision-making quality and effectiveness (from (Ghasemaghaei, et al., 2018))

According to the results of Ghasemaghaei et al. (Ghasemaghaei, et al., 2018), it can be concluded that, firstly, data analytics improves decision-making. Although this is a general statement, it can also be applied to a specific field such as decision-making of stakeholders in MOL of engineered products. Secondly, stakeholders can benefit from data-analytical methods. However, not only appropriate tools and algorithms should be used, but also the quality of the IT-based PUI (input data) should be adequately controlled. In addition, (based on Figure 38) the stakeholders should hire the people, trained in the application, as well as those who are familiar with the domain knowledge. In the case of the product lifecycle, this knowledge consists of information about the product, the processes of the product lifecycle or the business affairs of the product and its services.

Since this research has been conducted on a wide range of organizations in various industries and with products and services (151 companies), this thesis expands the results of this research to different stakeholder organizations in the product lifecycle. However, to investigate if the finding of Ghasemaghaei et al. (Ghasemaghaei, et al., 2018) is generalizable to the stakeholders of products, we used the results of another study of (KPMG, 2017; KPMG, 2016). The survey of KPMG (KPMG, 2016) has been conducted in various industries regarding the use of data analytics. The survey has been done on 704 company from several sectors. The participants were mainly heads of organizational departments. The stakeholders who participated in this survey were adopted from different areas of the value chain, such as logistics, production, marketing, finance and management.

The findings of the study show that organizations that work on the realization of engineered products (such as automotive industry, machinery, and plant engineering) are ahead of the rest of the industries (service, finance and process industry) in using data analytics. The study shows that 30% of organizations from automotive and 15% of the organization in the machinery industry, currently use advanced data analytics techniques and benefit from them (KPMG, 2016).

KPMG (KPMG, 2017) claims that organizations who used data analytics are satisfied with the obtained. Because it made monitoring (of processes, products etc.) possible. Other researchers have also admitted this finding, for example, Elbert & Scharf (Elbert & Scharf, 2018). They investigated the effect of digitalization on the decision-making of shipper and container company. Elbert & Scharf (Elbert & Scharf, 2018) found that the use of IT-based services on decision-making depends on the decision maker and the characteristics of IT-based service. If the IT-based service facilitates the processes of the stakeholder, it is likely that he/she uses the service. For example, shippers use to track and trace of containers.

Therefore, concluding from the studies mentioned above, it can be inferred that firstly using IT-based services (including services, which are build based on data analytics) can improve decision-making in the level of asset and processes management and monitoring. Similarly, using asset analytics and processes analytics (chapter 5) is helpful, because they can assist the stakeholders in getting more insight into engineered products, in addition to the processes related to product operation and planning in the lifecycle.

With these findings the answer to the question: “if the increase of data quality after applying appropriate data analytics tools have a positive effect on decision-making?” is positive (answer to validation question (b)). For stakeholders who deal with the product it can improve monitoring, and for stakeholders in charge of processes, it can support the operation.

In terms of decision-making based on product use information (validation question (a)), the results of KPMG (KPMG, 2016) survey confirm that part of the input data, which used in data analytics, is PUI. PUI has been considered as sensor data and customer data in this survey. The frequency of using sensor data in organizations is 60% and customer behavior data are used by 45% of the organizations. From these statistics, we can draw that around 50% of organizations use PUI.

Moreover, it can be inferred (KPMG, 2016) that only a part of the input data of a decision can be related to product and PUI. Therefore, determining the effectiveness of decision-making depends on several other factors and PUI is only a part of it. These factors, according to the field of decision-making, and decision-making based on data analytics (Ghasemaghaei, et al., 2018; Popovič, et al., 2012), can be data quality, organizational culture, domain knowledge and analytical skills.

According to the results of this thesis, (KPMG, 2016) and ongoing research on product usage data, it can be concluded that processing PUI can have a positive impact on decision-making. However, other factors affect decision-making and have a higher influence on decisions comparing PUI. According to the studies conducted in this thesis (chapter 6) and the results of other research, we can conclude that data analytics has the potential to support stakeholders in the lifecycle (answer to (a)).

7.2 Validation of data analytics models

For validating the quantitative part of this dissertation, namely, the implementation part, we used the Cross-Validation (CV) technique.

CV is a quantitative assessment technique, which evaluates the extent to which the results of a statistical model, can be adopted. Adoption means whether this statistical model is suitable to be used for other independent datasets or not. In simple words, CV is a way to choose the best data model and measure the accuracy of modeling. The best way to perform CV is tenfold CV.

The accuracy of tenfold CV is higher than a single model because the tenfold CV estimator has a lower variance (comparing single model). Thus, we apply this technique for evaluating data models. For information about the mechanism of CV refer to data mining and statistical analysis books such as (Bishop, 2006; Hastie, et al., 2009). In the following, CV results for all data models in chapter 6 are provided. The list of models is as follows.

- Cross-validation of regression tree for wind power generation model
- Cross-validation of RF for wind power generation model
- Cross-validation of classification tree for the wind speed ranges model
- Cross-validation of NN for the EV battery charge prediction model
- Cross-validation of SVM for the EV battery charge prediction model
- Cross-validation of NNETAR for the spare part planning model

In the following, the measurement error for each iteration of CV is presented. The number of iterations is 10. Because we use tenfold CV. For every iteration, two random datasets (test and train) are used. The datasets are generated in a way that they are statistically independent of each other. A new data model is created per iteration. The procedure of modeling for every iteration is similar to model biulding in chapter 6 (such as SVM, NN model). Note that the datasets are scaled to [0,1] values before running CV. The reason for scaling is to get values on the same scale. Equal-scale results can enable us to compare model performances.

7.2.1 Cross-validation of regression tree for wind power generation model

CART regression tree model was described in Table 18. The average error rate for CART tree model is 0.197 (Table 28). It shows that the model has acceptable accuracy.

Table 28: Cross-validation repetitions and mean error rate for CART regression tree

Cross-validation repetitions	Error of fitted model
Round 1	0.1809
Round 2	0.1887
Round 3	0.1801
Round 4	0.2280
Round 5	0.1639
Round 6	0.2520
Round 7	0.1799
Round 8	0.1856
Round 9	0.1755
Round 10	0.2408
Mean error rate	**0.1975**

7.2.2 Cross-validation of RF for wind power generation model

Using the RF model (Figure 27) can cause less modeling error than CART regression tree (Table 18). The average error rate of RF equals to 0.105 (Table 29), comparing to 0.197 (Table 28) for CART. Thus, the RF model shows better performance.

Table 29: Cross-validation repetitions and mean error for RF model

Cross-validation repetitions	Error of fitted model
Round 1	0.1428
Round 2	0.0822
Round 3	0.0902
Round 4	0.0678
Round 5	0.1003
Round 6	0.1013
Round 7	0.0951
Round 8	0.1464
Round 9	0.1058
Round 10	0.1177
Mean error rate	**0.1050**

7.2.3 Cross-validation of classification tree for the wind speed ranges model

Cross-validation is run on the classification tree for speed ranges (Table 19). The results of running the model 10 times by 10-fold classification are listed below:

Table 30: Cross-validation and error of classification tree

Cross-validation repetitions	Error of fitted model
Round 1	0.5498
Round 2	0.5517
Round 3	0.6220
Round 4	0.5862
Round 5	0.6014
Round 6	0.6069
Round 7	0.6186
Round 8	0.5414
Round 9	0.5808
Round 10	0.5739
Mean error rate	**0.5833**

Based on the results the error rate for the classification tree is 0.583 (Table 30). This rate is not very satisfactory. Therefore, using this model together with visualization results is advised (Figure 26).

7.2.4 Cross-validation of NN for the EV battery charge prediction model

The following shows the results of tenfold CV for NN model on EV battery charge prediction. Error measurement criterion is RMSE (explained in chapter 6). The error of fitted model, in tables, show RMSE for each iteration of CV (total ten iterations). The average of RMSE for ten iterations has been reported afterward.

The average RMSE of the prediction produced by the neural networks model is around 0.1735 (Table 31). Therefore, the performance of the model seems to be satisfying. Because a lower error (near 0) is favorable for an efficient model, errors less than 0.5 are generally considered statistically acceptable and can be a sign of a good model.

Table 31: Cross-validation repetitions and mean error rate for neural networks

Cross-validation repetitions	Error of fitted model
Round 1	0.1026
Round 2	0.4386
Round 3	0.0883
Round 4	0.1082
Round 5	0.0706
Round 6	0.1502
Round 7	0.1445
Round 8	0.1096
Round 9	0.3790
Round 10	0.1431
Mean error rate	**0.1735**

7.2.5 Cross-validation of SVM for the EV battery charge prediction model

For SVR model, the average error in ten iterations is equal to 0.117. It shows that the model can perform well (Table 32).

Table 32: Cross-validation repetitions and mean error rate for SVR model

Cross-validation repetitions	Error of fitted model
Round 1	0.0292
Round 2	0.2298
Round 3	0.0901
Round 4	0.0729
Round 5	0.1621
Round 6	0.0201
Round 7	0.0813
Round 8	0.0401
Round 9	0.1588
Round 10	0.2861
Mean error rate	**0.1170**

7.2.6 Cross-validation of NNETAR for the spare part planning model

Table 33 shows the mean and standard deviation of errors of tenfold CV for neural network based on NNETAR algorithm.

Table 33: Mean error of cross-validation for NNTAR

	Mean	Standard deviation
RMSE	0.7019	7.053e-02
AE	0.7019	8.476e-02

Details of cross-validation can be found in Appendix G. The values for RMSE shows that this model fails to perform satisfactorily. Therefore, the other approach used for scenario 3 in chapter 6 (time series forecasting with Tableau software, Figure 37) is more suitable to be adopted.

7.3 Discussion

Validity of identified MOL stakeholders and data and information needs. The results of identifying the stakeholders of MOL, data and information need and the validations of section 7.1 show that the findings from the industrial case studies of airplane and wind turbines (section 4.4) can be generalized to other products, such as motorboats and EVs. However, for a more comprehensive understanding, more research should be performed.

Validity of data analytics as a supportive tool to support MOL stakeholders. In this respect, section 7.1 provides arguments from state of the art. Based on these results, it can be said that data analytics can support MOL stakeholders. To improve the findings of this dissertation, the implementation scenarios should be put into practice and the feedback from the related MOL stakeholders should be gained.

Security of data exchange and privacy. Finding the right data for each decision and the right algorithm for the analytics are very important. However, this is just the initial step of a process that uses the output information for the decision-making. Another aspect, which influences the information needs of stakeholders, is the security of data exchange and privacy. It is essential to discuss the requirements of CL-PLM stakeholders from a security and privacy point of view. Moreover, finding a mechanism to share data by considering the security of data sharing is necessary. In case of the security for data sharing and data exchange among the stakeholders, misuse of data generated by the products might be possible. For instance, some sensor data from the product may have been tampered with, or a stakeholder may illegally (or accidentally) access information that he was not supposed to have. To deal with these types of challenges ongoing research is currently focussing on new decentralized methods for secure transactions and data sharing. Among these, Blockchain technology has recently gained significant attention (Wood, 2014). With this technology is it possible to track all the changes made to the data and therefore provide a level of trust for the stakeholders to work together (Nabati, et al., 2017). The interested reader may refer to the research on smart electronic contracts for more information.

Next chapter summarizes the findings and answers the research questions and provides possibilities of future work.

8 Conclusion

The conclusion chapter summarizes the work carried out in this dissertation and suggests the topics for future research. Section 8.1 starts with a summary of the dissertation and continues with gained results, innovation and contribution of this research. Later this section explains the answers to the research questions. Next, limitations of the research are presented and finally section outlook offers suggestions for future work.

8.1 Summary of findings and contribution of research

A concept for supporting stakeholders of product lifecycle was developed in this dissertation. Within this concept, recommendations, methods, and implementation scenarios were defined to promote a better understanding of product lifecycle stakeholders in MOL and the potential benefits that stakeholders can gain when utilizing newly emerged big data sources from product use and operation. This understanding can help organizations to manage challenges they face (as mentioned in chapter 1) in respect to digitalization, data sharing in product lifecycle as well as staying competitive in the market of trading products and services. Since stakeholders of product MOL had not been completely identified, there was a lack of research on a holistic concept that studies their information needs well. Particularly, information needs from new sources of product use (PUI). This dissertation addressed the gap and modeled the MOL stakeholders' data and information needs.

In order to identify the information needs, a research strategy consisting of real-world case studies, interviews, and qualitative analysis was defined and conducted in this dissertation to develop the concept. Following the identification of stakeholders and information needs, appropriate data analytical techniques were matched to the data and information needs. To reach this goal, a literature survey and experiments were conducted.

In addition, data analytics techniques were grouped in order to reduce complexity and variety of techniques and facilitate easy implementation of the concept. The criteria for grouping of data analytics tools derived from the characteristic of respective MOL data sources, type of information need and nature of the decision problem of stakeholder. After that, information needs were sharpened, so that they can be modeled with data analytics techniques.

Chapter 4 and 5 showed some of these shaped data and information needs for every stakeholder plus instances of adaption of information needs with data analytics techniques for every stakeholder. Chapter 6 presented three scenarios, where they show how data analytics works for realizing information needs. These scenarios, bridges and combines two main fields of data science and data and information needs from product MOL. These prototypical test cases give a more concrete picture of the applicability of the concept.

Following implementation, validation of models was conducted. Operational validation for implementation scenarios showed that data analytics could provide almost accurate prediction models for stakeholders in different fields and different stakeholders. Results from literature surveys also revealed that data analytics can be beneficial for stakeholders MOL and supports their data and information needs. Moreover, validation of quantitative models for each data analytics model was done by cross-validation technique. The finding from cross-validation showed which data analytics techniques perform better in each scenario.

Based on results of implementation and validation, it can be concluded that data analytics can consider as a supporting IT-based tool in future for all stakeholders of CL-PLM, from strategic and management decision-maker operational workers and planners.

In brief, a summary of contributions in this dissertation are as follows:

- Providing a better understanding of MOL and its mechanisms in terms of processes, data sources, beneficiaries and their needs
- Facilitating better and easier information management between the stakeholders of the product lifecycle by awareness of information needs
- Contributing to automatizing product lifecycle information flows
- Adopting right technologies and tools for serving stakeholders, processes and information sharing
- Demonstrating applications of data analytics in CL-PLM

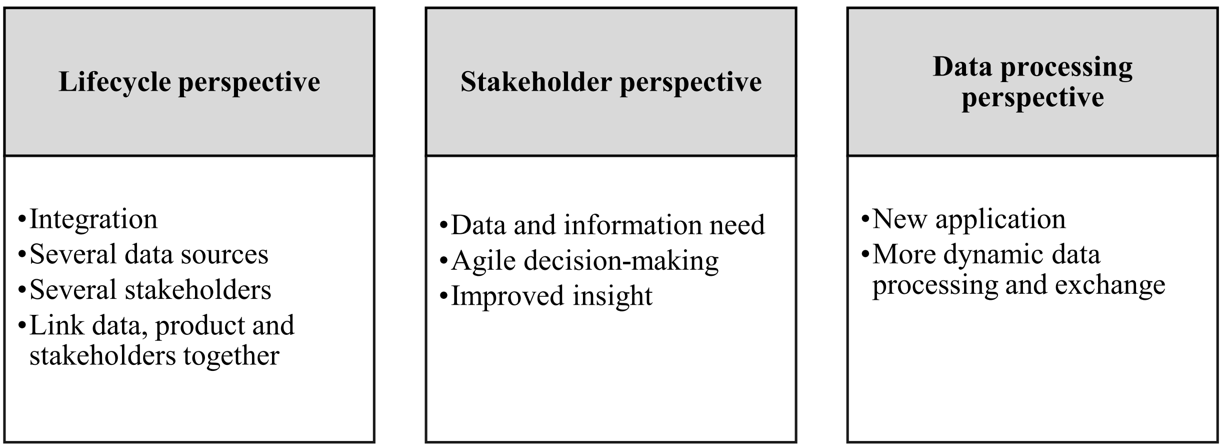

Figure 39: Contribution of the dissertation from three perspectives

Figure 39 shows the innovative aspects of this research work. From the perspective of data processing, applying data analytics to CL-PLM is a new area of application. To this end, considering a more dynamic data processing from MOL, as well as considering the exchange of data are two of novel aspects. Usually data analytics applications do not take these aspects into account.

From a lifecycle perspective, the integration of several data sources and several stakeholders is a contribution of this dissertation. Moreover, considering the impact of digitalization in MOL and its effect on lifecycle elements such as product, product services, information flows and stakeholders were studied. These findings enhance our understanding of product lifecycle in MOL.

RQ1: What are the sources in which the data are generated and where this information is needed?

The places where the information is needed and used include several lifecycle processes and information users involved in the product lifecycle. The information users (stakeholders) apply information as the input to lifecycle processes (for operation) or they use them for making decisions regarding the product or lifecycle processes.

This dissertation focuses on information, which comes from the operation of the product (PUI), and processes related to the product during its MOL. The reason and necessity to focus on this part of the information, is that still not enough research is done we focus on integrating these sources of information (as stated in chapter 2).

RQ2: What is the aim of using information from MOL? Who is interested in gaining knowledge from product MOL?

The flow of information can be seamless when PUI is delivered to the right place, person, at the right time with proper quality.

The aims of using information from MOL can be summarized as follows. First, people or stakeholder groups benefit from MOL information. To specify the respective groups, this dissertation studied MOL stakeholders and their information needs. Different stakeholders along the lifecycle own the information; they use the information for decision-making.

The second aim of using information was formulated from a technical point of view. There is a need for new concepts, which enable organizations to manage the lifecycle of the product. To this end, this dissertation considered the concept of CL-PLM and completed this concept from the aspect of stakeholders of the lifecycle and their decision-making. The goal of bringing more transparency to the product lifecycle has been addressed in this dissertation. Moreover, reaching the goal was tested by specifying possible data and information needs of product lifecycle beneficiaries. A challenge to determine information needs was that data sources from product operation are not still structured and that they do not have an integrated and uniform data structure in the product lifecycle information systems.

The results of this research can contribute to establishing the flows of information between different elements of the lifecycle and between different phases. Therefore, it is crucial to use MOL information.

Another aim of using MOL data was to improve the quality of decisions that stakeholders make. Data were processed, as mentioned in chapter 3, to provide the stakeholders with the information they need. Suggestions on relevant tools and scenarios were provided in chapter 5 & 6.

Another aim of using information from MOL was to improve collaboration among product lifecycle stakeholders. To this end, chapter 5 identified collaborative information sharing in MOL via a holistic model. Moreover, discussions about aspects of collaboration, such as security of information sharing, and effective contract design have been briefly presented in this dissertation.

RQ3: Can the information support decision-making in lifecycle if the information from lifecycle phases are provided to beneficiaries? If not, what is needed to be accomplished?

This research showed that it is not enough to make information from MOL available. Information can support decision-making only if the information from MOL are first selected based on the needs and roles of stakeholders and subsequently summarized and extracted as well as delivered in a proper visual format to beneficiaries. The information from MOL, which has the mentioned characteristics, can be called right information. In this dissertation, the right information is reflected in data and information needs.

Moreover, the information should be processed (summarized, extracted, visualized) with the relevant tool before it is delivered to the beneficiary. The results of a review of the literature (state of the art) in chapter 2 and 3 showed that there was no mechanism and concept for supporting MOL stakeholders with PUI. Therefore, this dissertation addressed this research gap in chapters 4 to 6.

RQ4: Can enhancing the richness of data improve the quality of decisions? If so, what is the appropriate mechanism?

The results of the validation of the model (section 8.3) show that improving MOL data with data analytics tools has a positive impact on operational decisions. Comparing the finding with other research (published in state of the art) also confirms this finding (Elbert & Scharf, 2018). However, decision-making for strategic and mid-term decisions still do not use and apply the findings from data analytics. Moreover, levels of awareness of organizations about big data technology, their level of trust to IT-based systems and skills on using the outcome of data analytics tools have an impact in real world on decisions of stakeholders.

8.2 Limitations

Finally, a number of limitations need to be considered from this dissertation. First, the current study has not evaluated “stakeholder relations with each other and with the product” (chapter 4) for all the stakeholders. That is, the finding was tested based on scenarios. However, all the MOL stakeholders and their data needs could not be tested and evaluated comprehensively. Having workshops with several stakeholders and getting their feedback on the information needs can improve the performance and applicability of the findings.

Another limitation is that the sample size for building the holistic concept (for identifying stakeholders and their requirements) was relatively small even though this model was built based on several interviews and surveys. Comparing to its scale, more interviews and implementation scenarios are needed in order to make sure that the concept can accommodate all types of information needs.

Application of this concept should be modified for every industry and value change. The third limitation is that although stakeholders can have several potential data needs, they do not know themselves what they can expect from smart engineered products in the future. Moreover, most stakeholders do not have information about data analytics tools or are reluctant to learn about them. These limitations affect finding the right data and information need of stakeholders. In addition, it makes building a good tool based on their information needs harder.

8.3 Outlook

It is recommended that further research is undertaken in the following areas:

First, a more extensive study with respect to the collaboration of stakeholders and their collaborative information needs. Chapter 4 briefly presented the changes due to the concept of PSS, contracting and their challenges regarding stakeholder relations. However, more research is needed to study the information needs of CL-PLM stakeholders in mutual information sharing.

Second, more topics for future research can be:

- Linking the proposed concept to CL-PLM systems from the technical point of view. For example, further modeling the identified information requirements, as well as functional requirements by using LML, XML language.
- Testing the model of stakeholder relations with each other and with the product in other industries rather than aerospace, automobile, and wind energy.
- Study issues of trust, security, integrity, and privacy regarding stakeholders’ relationships in CL-PLM and data sharing in MOL.

Third, when having large volumes and complex sources of MOL data, future work should address the following questions:

- How to create metadata and ontologies from the gained insights of data analytics? The challenge is that, with heterogeneous data sources, it is hard to implement these tasks.
- How to define a standardized procedure for acquiring information needs and how to make the results of data analytics reproducible?
- How to preserve data quality across different and interrelated processes of product lifecycle?

Processing the PUI with advanced analytics, techniques contain several transformations. When data are combined with data analytical techniques, to go a step back and regenerate the data is very difficult. Therefore, tracking a piece of data in combined mode is very hard, also in CL-PLM systems.

Apart from mentioned future research opportunities, the issue of preserving data and information for the concept presented in this dissertation (either input PUI or the output information), as well as preserving insights from stakeholders, is another aspect that should be investigated.

9 References

Abramovici, M., Aidi, Y. & Dang, H. B., 2013. *Knowledge-based lifecycle management approach for product service systems (PSS).* Nantes, France, IFIP International Conference on Product Lifecycle Management, pp. 239-248.

Abramovici, M., Fathi, M., Holland, A. & Neubach, M., 2008. *Integration of product use information into PLM.* LCE 2008, 15th CIRP International Conference on Life Cycle Engineering, pp. 438-443.

Abramovici, M., Neubach, M., Schulze, M. & Spura, C., 2009. *Metadata Reference Model for IPS 2 Lifecycle Management,* Cranfield University, UK: Proceedings of the 19th CIRP Design Conference–Competitive Design.

Acatech, 2013. *Recommendations for implementing the strategic initiative INDUSTRIE 4.0.,* Berlin: acatech – National Academy of Science and Engineering.

Active Sensors Company, 2016. *Active Sensors; Aerospace.* [Online]
Available at: http://www.activesensors.com/markets/aerospace

Ahlemeyer-Stubbe, A. & Coleman, S., 2014. *A Practical Guide to Data Mining for Business and Industry.* UK: John Wiley & Sons.

Air transport action group, 2011. *Global airtransport stakeholders.* [Online]
Available at: www.atag.org

Airlines, D., 2016. [Online]
Available at: http://www.deltatechops.com/services/view/category/contract-types

Allianz, 2016. *Allianz Energy Insurance.* [Online]
Available at: http://www.agcs.allianz.com/services/energy/

Ameri, F. & Dutta, D., 2005. Product lifecycle management: closing the knowledge loops. *Computer-Aided Design and Applications,* 2(5), pp. 577-590.

Anitha, P. & Prabhu, B., 2012. *Integrating requirements engineering and user experience design in product life cycle management.* Zurich, Switzerland, IEEE Press, pp. 12-17.

Anninni, A., 2014. *Big data in the wind power industry,* Grenoble: Grenoble Graduate School of Business.

Ashford, N., Stanton, M. & Moore, C., 1996. *Airport Operations.* 2nd ed. USA: McGraw-Hills.

Baines, T. S. L. et al., 2007. State-of-the-art in product-service systems. *Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture,* 222(10), pp. 1543-1552.

Bakker, S., Maat, K. & van Wee, B., 2014. Stakeholders interests, expectations, and strategies regarding the development and implementation of electric vehicles: The case of the Netherlands. *Transportation Research Part A: Policy and Practice,* Volume 66, pp. 52-64.

Belkadi, F. et al., 2008. Innovative PLM-based approach for collaborative design between OEM and suppliers: Case study of aeronautic industry. *IFIP International Federation for Information Processing,* Volume 277, pp. 157-168.

Bischof, M., 2012. *Der Einfluss internationaler Leitlinien auf das CSR- und Stakeholder-Management.* Wien: Magisterarbeit, universität Wien.

Bishop, C. M., 2006. *Pattern recognition and machine learning.* 1st ed. New York: Springer.

Borsato, M., 2014. Bridging the gap between product lifecycle management and sustainability in manufacturing through ontology building. *Computers in Industry,* p. 258–269.

Breiman, L., Friedman, J. H., Olshen, R. A. & Stone, C. J., 1984. *Classification and Regression Trees.* Belmont, CA.: Wadsworth International Group.

Brockwell, P. & Davis, R. A., 2002. *Introduction to Time Series and Forecasting.* 2nd ed. New york: Springer.

Bryson, J. M., 2003. *What To Do When Stakeholders Matter: A Guide to Stakeholder Identification and Analysis Techniques.* Wasington, USA, National Public Management Research Conference.

Butler, S., 2012. *Prognostic Algorithms for Condition Monitoring and Remaining Useful Life Estimation.* Ireland: National University of Ireland.

Cadigan, S., 2017. *How To Future-Proof Your Career In A Digital Economy.* [Online]
Available at: https://www.forbes.com/sites/stevecadigan/2017/03/21/how-to-thrive-in-an-evolving-digital-landscape/#4302fefe74fc

Cassina, J. et al., 2006. *Development of the semantic object model for a PDKM system.* Milan, Italy., IEEE, pp. 383-390.

Chandrasegaran, S. K. et al., 2013. The evolution, challenges, and future of knowledge representation in product design systems. *Computer-Aided Design,* 45(2), pp. 204-228.

Chen, H., Chiang, R. H. & Storey, V. C., 2012. Business intelligence and analytics: From big data to big impact. *MIS quarterly,* 36(4), pp. 1165-1188.

CIMdata, 2017. *Making the Connection: The Path to Cloud PLM,* USA: Oracle.

Colas, M. et al., 2014. *Cracking the Data Conundrum : How Successful Companies Make Big Data Operational,* 1-44: Capgemini Consulting.

Collinge, B., 2011. *Understanding stakeholder requirements on an NHS hospital project: application of semiotics-rooted theories.* Bristol, UK, 27th ARCOM annual conference, pp. 963-972.

Dassult Systems, 2015. [Online]
Available at: http://www.3ds.com/industries/energy-process-utilities/sustainable-wind-turbines/sustainable-blade-design/

De'ath, G. & Fabricius, K. E., 2000. Classification and regression trees: a powerful yet simple technique for ecological data analysis. *Ecology,* 81(11), pp. 3178-3192.

Delen, D. & Zolbanin, H. M., 2018. The analytics paradigm in business research. *Journal of Business Research,* Volume 90, p. 186–195.

DELWP, A., 2018. *Department of Environment, Land, Water and Planning.* [Online]
Available at: http://www.dse.vic.gov.au/effective-engagement/toolkit/tool-stakehoder-analysis-stakeholder- matrix

Deng, Q., Franke, M., Hribernik, K. & Thoben, K. D., 2017. *Exploring the integration of social media feedback for user-oriented product development..* Vancouver, Canada, DS 87-4 Proceedings of the 21st International Conference on Engineering Design (ICED 17).

Dienst, S., Ansari-Ch, F., Holland, A. & Fathi, M., 2010. Necessity of using Dynamic Bayesian Networks for feedback analysis into product development. *IEEE International Conference on Systems Man and Cybernetics,* pp. 939-946.

Dienst, S., Fathi, M., Abramovici, M. & Lindner, A., 2011. *A Conceptual Data Management Model of a Feedback Assistance System to support Product Improvement.* USA, IEEE, pp. 446-451.

Dienst, S. et al., 2012. *Concept for improving industrial goods via contextual knowledge provision.* Graz, ACM.

Dudovskiy, J., 2018. *Inductive Approach (Inductive Reasoning).* [Online]
Available at: https://research-methodology.net/research-methodology/research-approach/inductive-approach-2/

El Kadiri, S. & Kiritsis, D., 2015. Ontologies in the context of product lifecycle management: state of the art literature review. *International Journal of Production Research,* 53(18), pp. 5657-5668.

Elbert, R. & Scharf, K., 2018. *Analysis of the Choice Behavior for Container Transport Services in the Maritime Hinterland.* Bremen, Germany, Springer, pp. 199-203.

Estevan, H., Schaefer, B. & Adell, A., 2018. *Life Cycle Costing State of the art report,* SPP Regions, EU: ICLEI – Local Governments for Sustainability, European Secretariat.

European Commission, 2017. *European countries join forces to digitise industry.* [Online]
Available at: https://ec.europa.eu/digital-single-market/en/news/european-countries-join-forces-digitise-industry

Eyer, W. W., 1979. Sale, Leasing and Financing of Aircraft. *Journal of air law and commerce,* 217(45).

Fathi, M. & Holland, A., 2009. *Knowledge-based feedback integration to facilitate sustainable product innovation.* Mallorca, Spain, IEEE.

Fathi, M., Holland, A., Abramovici, M. & Neubach, M., 2007. *Advanced Condition Monitoring Services in Product Lifecycle Management.* USA, IEEE, pp. 245-250.

Fernández, A. et al., 2014. Pattern Recognition in Latin America in the "Big Data" Era. *Pattern Recognition,* 48(4), p. 1185–1196.

First Market Controls, 2016. [Online]
Available at: http://www.firstmarkcontrols.com/newsfdr.htm

Ford, G., McMahon, C. & Rowley, C., 2015. *An Examination of Significant Issues in Naval Maintenance.* UK, The Fourth International Conference on Through-life Engineering Services, pp. 197-203.

Fosso-Wamba, S. et al., 2015. How ‘Big Data’ Can Make Big Impact: Findings from a Systematic Review and a Longitudinal Case Study. *International Journal of Production Economics,* Volume 165, pp. 234-246.

Freeman, E., 2010. *Strategic Management: A Stakeholder Approach.* 1st ed. Cambridge, UK: Cambridge University Press.

Freitag, M., Kück, M., Ait Alla, A. & Lütjen, M., 2015. Potentiazale von Data Science in Produktion und Logistik. *Industrie 4.0,* pp. 22-26.

Gandomi, A. & Haider, M., 2015. Beyond the hype: Big data concepts , methods , and analytics. *International Journal of Information Management,* Volume 35, pp. 137-144.

García Márquez, F. P., Tobias, A. M., Pinar Pérez, J. & Papaelias, M., 2012. Condition monitoring of wind turbines: Techniques and methods. *Renewable Energy,* Volume 46, pp. 169-178.

Garg, J., 2013. *Maintenance: Spare Parts Optimization,* Paris: Capgemini Consulting.

Gartner, 2013. [Online]
Available at: https://readwrite.com/2013/09/18/gartner-on-big-data-everyones-doing-it-no-one-knows-why/

General Electrics Corporation, 2013. *The Case for an Industrial Big Data Platform,* USA: gesoftware.com.

Ghasemaghaei, M., Ebrahimi, S. & Hassanein, K., 2018. Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems,* 27(1), pp. 101-113.

Gilliam, E., 2017. *30 Best Customer Feedback Tools: an overview.* [Online]
Available at: https://mopinion.com/30-best-customer-feedback-tools-an-overview/

Glover, F., Klingma, D. & Phillip, N., 1992. Decision Support System for Network Models. In: *Network Models in Optimization and their Application in Practice.* USA: John Wiley and Sons.

Gopsill, J., McAlpine, H. & Hicks, B., 2011. *Trends in technology and their possible implications on PLM: looking towards 2020..* Bath, UK, s.n., pp. 1-12.

Greiner, S. et al., 2015. *German Offshore Wind Operation Guide,* Bremen: Hochschule Bremen.

Gulledge, T., Hiroshige, S. & Iyer, R., 2010. Condition-based Maintenance and the product improvement process. *Computers in Industry,* 61(9), pp. 813-832.

Günther, F. & Fritsch, S., 2010. neuralnet: Training of Neural Networks. *The R Journal,* 2(1), pp. 30-38.

Hadaya, P. & Marchildon, P., 2012. Understanding product lifecycle management and supporting systems. *Industrial Management & Data Systems,* 112(4), pp. 559-583.

Hahmann, M., Schröder, G. & Grosse, P., 2011. *Universität Dresden.* [Online]
Available at: https://wwwdb.inf.tu-dresden.de/misc/WS1112/FK/01_data_analytics.pdf
[Accessed 09 2015].

Hameed, Z. et al., 2009. Condition monitoring and fault detection of wind turbines and related algorithms: A review. *Renewable and Sustainable Energy Reviews,* 13(1), pp. 1-39.

Hashem, I. A. T. et al., 2014. The rise of "Big Data" on cloud computing: Review and open research issues. *Information Systems,* Volume 47, pp. 98-115.

Hastie, T., Tibshirani, R. & Friedman, J., 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* 2nd ed. Stanford: Springer.

Horton, G., 2016. *Impuls für Innovation.* [Online]
Available at: http://www.zephram.de/blog/geschaeftsmodellinnovation/beispiel-servitization/

Hribernik, K., Franke, M., Coscia, E. & Thoben, K.-D., 2017. *Towards a Platform for Integrating Product Usage Information into Innovative Product-Service Design.* Madeira, Portugal, Proceedings of the 23rd ICE Conference 2017.

Hribernik, K., Stietencron, M., Hans, C. & Thoben, K., 2011. Intelligent Products to Support Closed-Loop Reverse Logistics. *Glocalized Solutions for Sustainability in Manufacturing,* pp. 486-491.

Hyndman, R. et al., 2018. *Package 'forecast',* Australia: R-CRAN Repository.

IATA, 2017. *Ground Operations.* [Online]
Available at: http://www.iata.org/whatwedo/ops-infra/ground-operations/Pages/index.aspx

Iatrou, K. & Oretti, M., 2016. *Airline choices for the future: from alliances to mergers.* USA: Routledge.

Jansen-Vullers, M., Van Dorp, C. & Beulens, A., 2003. Managing Traceability Information in Manufacture. *International Journal of Information Management,* Volume 23, pp. 395-413.

Jennings, C., Wu, D. & Terpenny, J., 2016. *Forecasting Obsolescence Risk and Product Life Cycle With Machine Learning.* USA, IEEE Transactions on Components, Packaging and Manufacturing Technology, pp. 1428-1439.

Jun, H. B., Kiritsis, D. & Xirouchakis, P., 2007. Product life-cycle metadata modeling and its application with RDF. *IEEE Transactions on Knowledge and Data Engineering,* 19(12), pp. 1680-1693.

Jun, H.-B., Kiritsis, D. & Xirouchakis, P., 2007. Research issues on closed-loop PLM. *Computers in Industry,* Volume 8-9, pp. 855-868.

Jun, H.-B., Shin, J.-H., Kiritsis, D. & Xirouchakis, P., 2007. System architecture for closed-loop PLM. *International Journal of Computer Integrated Manufacturing,* 20(7), pp. 684-698.

Kaisler, S. H., Espinosa, J. A., Armour, F. & Money, W. H., 2014. *Advanced Analytics-Issues and Challenges in a Global Environment.* USA, IEEE, pp. 729-738.

Kammoun, M. A. & Rezg, N., 2018. Toward the optimal selective maintenance for multi-component systems using observed failure: applied to the FMS study case. *The International Journal of Advanced Manufacturing Technology.*

Kaufmann, M., 2007. *Data Mining: Practical Machine Learning Tools and Techniques.* 2nd ed. USA: Elsevier.

Kauschke, S., Fürnkranz, J. & Janssen, F., 2016. *Predicting Cargo Train Failures: A Machine Learning Approach for a Lightweight Prototype.* Italy, Springer, Cham, pp. 151-166.

Khademhosseinieh, B. & Khan, M. T., 2009. *Tools and Organisational Measures to Improve Information Flow,* Sweden: Master thesis.

Kiritsis, D., 2011. Closed-loop PLM for intelligent products in the era of the Internet of things. *Computer-Aided Design,* 43(5), pp. 479-501.

Klinke, J. & Klarmann, M., 2014. *Referenzprozessmodell für den Lebenszyklus von Offshore-Windparks,* Germany: BTC Wind Farm Center.

Kolodziejski, C. & Szöllösi-Brenig, V., 2015. *Big Data in a Transdisciplinary Perspective,* Hannover: VolkswagenStiftung.

Korpi, E. & Ala-Risku, T., 2008. Life cycle costing: a review of published case studies. *Managerial Auditing Journal,* 23(3), pp. 240-261.

Kowalczyk, A., 2014. *Support vector regression with R.* [Online]
Available at: https://www.svm-tutorial.com/2014/10/support-vector-regression-r/

KPMG, 2016. *Mit Daten Werte Schaffen,* Germany: KPMG AG.

KPMG, 2017. *KPMG operations excellence.* [Online].

KPMG, 2017. *Mit Daten Werte schaffen,* Germany: KPMG AG.

Kuhn, O., Liese, H. & Stjepandic, J., 2008. Engineering Optimisation by Means of Knowledge Sharing and Reuse. *International federation of information processing,* pp. 95-106.

Lachmayer, R., Mozgova, I., Sauthoff, B. & Gottwald, P., 2014. Evolutionary approach for an optimized analysis of product life cycle data. *Procedia Technology,* Volume 15, pp. 359-368.

Lau, B., Ma, E. & Pecht, M., 2012. *Review of offshore wind turbine failures and fault prognostic methods.* China, IEEE, pp. 1-5.

Law, D., G. C. & Eberhardt, J., 2014. *Do you know big data.* [Online].

Lawley, B., 2016. *Full Product Management Infographic.* [Online]
Available at: https://280group.com/product-management-blog/full-product-management-infographic/#lightbox/1/

Lee, J., Davari Ardakani, H., Yang, S. & Bagheri, B., 2015. *Industrial big data analytics and cyber-physical systems for future maintenance & service innovation.* Cranfield, UK, The Fourth International Conference on Through-life Engineering Services.

Lee, Y. C., Sheu, L. C. & Tsou, Y. G., 2008. Quality function deployment implementation based on Fuzzy Kano model: An application in PLM system. *Computers & Industrial Engineering,* 55(1), pp. 48-63.

Lessard, M., 2009. *Airline Business & Law.* [Online]
Available at: https://www.mcgill.ca/iasl/files/iasl/ASPL614_Aircraft-Buy-Lease_Lessard.pdf

Lindström, J., 2016. *Improving Functional Product availability: software-related measures planned and taken.* Cranfield, Procedia CIRP.

Lufthansa Technik, 2016. [Online]
Available at: http://www.lufthansa-technik.com/de/

Luh, Y. P., Pan, C. C. & Chu, C. H., 2011. A hierarchical deployment of distributed product lifecycle management system in collaborative product development. *International Journal of Computer Integrated Manufacturing,* Volume 24, pp. 471-483.

Madenas, N., 2014. *Integrating product lifecycle management systems with maintenance information across the supply chain for root cause analysis.* Cranfield, UK, United Kingdom: Cranfield University.

Maglaras, L. et al., 2017. Mobile Networks and Applications. pp. 1-3.

Mahdjoub, M., Monticolo, D., Gomes, S. & Sagot, J., 2010. A collaborative Design for Usability approach supported by Virtual Reality and a Multi-Agent System embedded in a PLM environment. *Computer-Aided Design,* Volume 42, p. 402–413.

Majava, J. & Haapasalo, H., 2015. *The Roles of Stakeholders in an NPD Project: A Case Study.* Bari, Italy, Proceedings of the MakeLearn and TIIM Joint International Conference 2015, pp. 199-205.

Makan, A. et al., 2015. Stakeholder analysis of the Programme for Improving Mental health carE (PRIME): baseline findings. *International Journal of Mental Health Systems,* 27(9), pp. 1-12.

Manyika, J. et al., 2016. *Digital globalization: The new era of global flows,* New York: McKinsey Global Institute.

Marr, B., 2015. *That's Data Science: Airbus Puts 10,000 Sensors in Every Single Wing.* [Online]
Available at: http://www.datasciencecentral.com/profiles/blogs/that-s-data-science-airbus-puts-10-000-sensors-in-every-single

Marr, B., 2017. What Is Digital Twin Technology - And Why Is It So Important?. *Forbes Magazin*.

McFarlane, D. & Cuthbert, R., 2012. Modelling information requirements in complex engineering services. *Computers in Industry,* 63(4), pp. 349-360.

McKinsey & Company, 2014. Product Excellence. In: *Driving success through true product excellence.* USA: McKinsey.

Meer, D. & Dasgupta, A., 2012. *How to enable business decisions with "big data" and analytics,* New York: PWC.

Meyer, G., Främling, K. & Holmström, J., 2009. Intelligent Products: A survey. *Computers in Industry,* 60(3), pp. 137-148.

Montana, P. J. & Charnov, B. H., 2008. *Management.* 4th ed. NY: Barron's Business Review Series.

Montgomery, D. C., Peck, E. A. & Vining, G. G., 2012. *Introduction to linear regression analysis.* 4th ed. USA: John Wiley & Sons.

Moustafa, N., Creech, G. & Slay, J., 2017. Big Data Analytics for Intrusion Detection System: Statistical Decision-Making Using Finite Dirichlet Mixture Models. In: S. International, ed. *Data Analytics and Decision Support for Cybersecurity.* UK: Springer, pp. 127-156.

Muller, A., Marquez, A. & Iung, B., 2008. On the concept of e-maintenance: Review and current research. *Reliability Engineering & System Safety,* 93(8), pp. 1165-1187.

Müller, P., 2016. *Industrielles Internet & PLM: Produktentwicklung, Markt,strategische Effekte.* Kaiserslautern: Contact Software.

Nabati, E. G. & Thoben, K. D., 2016. *On Applicability of Big Data Analytics in the Closed-Loop Product Lifecycle: Integration of CRISP-DM Standard.* South Carolina, USA, Springer, pp. 457-467.

Nabati, E. G., Thoben, K. D. & Daudi, M., 2017. Stakeholders in the middle of life of complex products: understanding the role and information needs. *International Journal of Product Lifecycle Management,* 10(3), pp. 231-257.

Neuman, W., 2003. *Social Research Methods: Qualitative and Quantitative Approaches.* 7th ed. USA: Pearson education.

Nilsson, P. & Fagerström, B., 2006. Managing stakeholder requirements in a product modelling system. *Computers in Industry,* Volume 2, pp. 167-177.

Oelker, S., Lewandowski, M. & Freitag, M., 2015. Konzept einer Preagierenden Instandhaltungsstrategie. *Industrie 4.0 Management,* 31(5), pp. 40-43.

Pátkai, B. & McFarlane, D., 2006. *RFID-based sensor integration in aerospace,* University of Cambridge, UK: Auto-ID Lab.

Pelham, J. G., Fan, I. S., Jennions, I. & McFeat, J., 2015. *Application of an AIS to the problem of through life health management of remotely piloted aircraft.* UK, AIAA Infotech, pp. 1797-1810.

PennWell Corporation, 2016. [Online]
Available at: http://www.laserfocusworld.com/articles/2008/04/piezoelectric-actuators-and-sensors-embedded-in-aircraft-wings.html

Popovič, A., Hackney, R., Coelho, P. S. & Jaklič, J., 2012. Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. *Decision Support Systems,* Volume 54, p. 729–739.

PROMISE PEID Grouping, 2009. *The information exchange for closed-loop product lifecycle management.* [Online]
Available at: http://cl2m.com/wiki/75

PROMISE, 2010. *Product Lifecycle Management and Information Tracking Using Smart Embedded Systems.* [Online]
Available at: http://promise-innovation.com/promise-plm
[Accessed 2014].

Pulvino, T., 1998. Do Asset Fire Sales Exist? An Empirical Investigation of Commercial Aircraft Transactions. *The Journal of Finance,* 53(3), pp. 939-978.

Rachuri, S. et al., 2008. Information sharing and exchange in the context of product lifecycle management: Role of standards. *Computer-Aided Design,* 40(7), pp. 789-800.

Rahej, D., Lllinas, J., Nagi, R. & Romanowski, C., 2006. Data fusion/data mining-based architecture for condition-based maintenance. *International Journal of Production Research,* 44(14), p. 2869–2887.

Rasmussen, P., 2017. *Just What Are Engineered Products Anyway?.* [Online]
Available at: https://news.ewmfg.com/blog/just-what-are-engineered-products-anyway

Rolls Royce Corp., 2016. *Rolls-Royce Partners Finance.* [Online]
Available at: http://www.rolls-royce.com/media/insights/rolls-royce-partners-finance.aspx

Rolls-Royce Corp., 2016. *Rolls-Royce Partners Finance.* [Online]
Available at: http://www.rolls-royce.com/media/insights/rolls-royce-partners-finance.aspx

Rolls-Royce, 2016. [Online]
Available at: http://www.rolls-royce.com/about/our-technology/enabling-technologies/engine-health-management.aspx#sense

Rosenblatt, F., 1961. *Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms.* Washington DC: Spartan Books.

Rouse, M., 2016. *Data ingestion.* [Online]
Available at: https://whatis.techtarget.com/definition/data-ingestion

Schaar, D. & Sherry, L., 2010. *Analysis of airport stakeholders.* Integrated Communications Navigation and Surveillance Conference (ICNS), IEEE, pp. J4-1-J4-17.

Schallmo, D. R., 2016. *Digitale Transformation von Geschäftsmodellen.* Wiesbaden: Springer Gabler.

Schilli, B., 2006. Collaborative life cycle management between suppliers and OEM. *Computers in Industry,* Volume 57, pp. 725-731.

Schmeer, K., 1999. *Guidelines for Conducting a Stakeholder Analysis,* USA: Bethesda, MD: Partnerships for Health Reform, Abt Associates Inc.

Schoenthaler, F. A. D. &. K. T., 2015. *Design and governance of collaborative business processes in industry 4.0..* Lisbon, In Proceedings of the Workshop on Cross-organizational and Cross-company BPM (XOC-BPM) co-located with the 17th IEEE Conference on Business Informatics (CBI 2015), pp. 1-8.

Schuh, G., Rozenfeld, H., Assmus, D. & Zancul, E., 2008. Process oriented framework to support PLM implementation. *Computers in industry,* 59(2-3), pp. 210-218.

Schulte, S., 2008. Customer centric PLM: integrating customers' feedback into product data and lifecycle processes. *International Journal of Product Lifecycle Management,* 3(4), pp. 295-307.

Sharp, H., Finkelstein, A. & Galal, G., 1999. *Stakeholder Identification in the Requirements Engineering Process.* Proceedings. Tenth International Workshop on IEEE ed. Florence, Italy: Database and Expert Systems Applications.

Shearer, C., 2000. The CRISP-DM model: the new blueprint for data mining. *Journal of data warehousing,* 5(4), pp. 13-22.

Sheng, S., 2015. *Improving component reliability through performance and condition monitoring data analysis.* Houston, Texas: National revenewable energy labratory (NREL).

Shilivitsky, O., 2014. *Beyond PLM.* [Online]
Available at: http://beyondplm.com/2014/08/05/the-end-of-single-plm-database-architecture-is-coming/

Siemens PLM Software, I., 2011. *Open product lifecycle data sharing using XML,* USA: Siemens PLM Software.

Sivarajah, U., Kamal, M. M., Irani, Z. & Weerakkody, V., 2017. Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research,* Volume 70, pp. 263-286.

Sprague, R. A., Singh, K. J. & Wood, R. T., 1991. Concurrent engineering in product development. *IEEE Design & Test of Computers,* 8(1), pp. 6-13.

ST Aero, 2015. *Singapore Technologies Aerospace Ltd.* [Online]
Available at: https://www.staero.aero/

Stark, J., 2006. *Product lifecycle management, 21st Century paradigm for product realisation.* London: Springer.

Sudarsan, R., Fenves, S., Sriram, R. & Wang, F., 2005. A product information modeling framework for product lifecycle management. *Computer-Aided Design,* 37(13), pp. 1399-1411.

SunGard solutions, 2013. *White paper- Big Data Challenges and Opportunities for the energy industry,* FIS Corporate: https://www.sungard.com/.

Tableau, 2018. *How Forecasting Works in Tableau.* [Online]
Available at: http://onlinehelp.tableau.com/current/pro/desktop/en-us/forecast_how_it_works.html

Taefi, T. T., Fink, A. & Stütz, S., 2016. *Increasing the Mileage of Battery Electric Medium-Duty Vehicles: A Recipe for Competitiveness?,* Hamburg: HSU Institute of Computer Science Research Paper Series.

Taigel, F., Tueno, A. K. & Pibernik, R., 2018. Privacy-preserving condition-based forecasting using machine learning. *Journal of Business Economics,* pp. 1-30.

Taisch, M., Cammarino, B. & Cassina, J., 2011. Life cycle data management: first step towards a new product lifecycle management standard. *International Journal of Computer Integrated Manufacturing,* 24(12), pp. 1117-1135.

Takoutsing, P. et al., 2014. Wind Turbine Condition Monitoring: State-of-the-Art Review, New Trends, and Future Challenges. *Energies,* 7(4), pp. 2595-2630.

Tempest, R., 2015. *The Future of PLM resides in Brussels,* Brussels: http://www.plm-irf.org/.

Terzi, S., Panetto, H., Morel, G. & Garetti, M., 2007. A holonic metamodel for product traceability in Product Lifecycle Management. *International Journal of Product Lifecycle Management,* 2(3), pp. 253-289.

The renewable energy hub, 2016. *The renewable energy hub UK: Wind turbines.* [Online]
Available at: https://www.renewableenergyhub.co.uk/certification-bodies-for-wind-turbines.html#sthash.F6Da60Ye.dpuf

Thoben, K. D. et al., 1999. *From Product Data to Product Data and Knowledge Management - Requirements and Research Perspective.* Den Hauge, Netherlands, mcbuchner.com, pp. 147-155.

Todd, J. & Thorstensen, L., 2013. *Creating the green energy economy- Analysis of electric vehicle industry,* Washington DC, USA: International Economic Development Council.

Tracht, K., Westerholt, J. & Schuh, P., 2013. Spare parts planning for offshore wind turbines subject to restrictive maintenance conditions. *Procedia CIRP,* Volume 7, pp. 563-568.

Tukker, A. & Tischner, U., 2006. *New business for old Europe: product-service development, competitiveness and sustainability.* Sheffield: Greenleaf Publications.

TÜV Reinland, 2018. *Electric Vehicle Charging System Testing.* [Online]
Available at: https://www.tuv.com/world/en/electric-vehicle-charging-system-testing.html

U.S. Department of Energy, 2018. *Electric Vehicles: Stakeholder Solution Center.* [Online]
Available at: https://www.energy.gov/eere/electricvehicles/electric-vehicles-stakeholder-solution-center

Umeda, Y. et al., 2012. Toward integrated product and process life cycle planning - An environmental perspective. *CIRP Annals - Manufacturing Technology,* Volume 61, pp. 681-702.

United Nations, 1998. *Kyoto Protocol.* [Online]
Available at: http://unfccc.int/kyoto_protocol/items/2830.php

United Nations, 2015. *Paris Agreement.* [Online]
Available at: http://unfccc.int/paris_agreement/items/9485.php

Vera-baquero, A., Colomo-palacios, R. & Molloy, O., 2014. Towards a process to guide Big Data based Decision Support Systems for Business Processes. *Procedia Technology,* Volume 16, pp. 11-21.

Wellsandt, S. et al., 2016. A survey of product lifecycle models: towards complex products and service offers. *International Journal of Product Lifecycle Management,* 9(4), pp. 353-390.

Wellsandt, S., Thoben, K. D. & Hribernik, K., 2015. *Sources and chracteristics of information about product use.* Israel, Procedia CIRP, p. 242–247.

Werner, D. C., Weidlich, R., Guenther, B. & Blaurock, G., 2004. Engineers' CAx education- it's not only CAD. *Computer-Aided Design,* 36(14), pp. 1439-1450.

Wiesner, S. et al., 2014. *Requirements engineering for cyber-physical systems.* Germany, Springer Berlin Heidelberg, pp. 281-288.

Wind Energy Hamburg, 2016. *Wind Enenrgy Hamburg trade fair: brach categories.* [Online]
Available at: http://www.windenergyhamburg.com/en/the-fair/exhibitors-products/exhibitor-directory-2016/#/search/t=10

Wuest, T., 2014. *Approach to identify product and process state drivers in manufacturing systems using supervised machine learning,* Bremen, Germany: PhD Thesis, University of Bremen.

Wuest, T., Hribernik, K. & Thoben, K. D., 2014. *Capturing, managing and sharing product information along the lifecycle for design improvement.* Gommern, 10th international workshop on integrated design engineering.

Xiao, S., Xudong, C., Li, Z. & Guanghong, G., 2010. Modeling framework for product lifecycle information.. *Simulation Modelling Practice and Theory,* 18(8), pp. 1080-1091.

Yu, H., Xie, T., Paszczynski, S. & Wilamowski, B. M., 2011. Advantages of radial basis function networks for dynamic system design. *IEEE Transactions on Industrial Electronics,* 58(12), pp. 5438-5450.

Zell, H., 2007. *Projektmanagement. - lernen, lehren und für die Praxis.* Norderstedt: Books on Demand GmbH.

Zhang, Y., Ren, S., Liu, Y. & Si, S., 2017. A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products. *Journal of Cleaner Production,* Volume 142, pp. 626-641.

Zuccolotto, M. et al., 2013. I2MS2C: Intelligent maintenance system architecture proposal. *Chemical Engineering Transactions,* Volume 33, pp. 241-246.

Zügn, D., 2017. *Quo Vadis "PLM" in the Age of Digitization?,* Germany: PTC.

10 Appendix

10.1 Appendix A: Techniques and tools in PLM/CL-PLM systems for enhancing the value of information

The literature shows a variety of approaches, tools and technologies that organizations use for enriching information in different processes of product lifecycle. In this section, from the perspective of methods and techniques, the following questions are investigated: (a) How stakeholders of lifecycle find their information needs? (b) Which systems or technologies they made for it?

The literature tends to address information enhancement in the lifecycle from major areas. Thus, this dissertation categorizes the body of literature into the following three major areas. Selection of the tool mainly depends on the task to be performed or on the area (process) in which the enhancement should happen.

Base on literature review, companies used the following approaches to get their information. The approaches can be categorized into three groups.

1. Enhancing information access
2. Enhancement of the use/reuse of information and knowledge
3. Enhancement of information to support decision-making

Papers which address the first category mainly show the IT and more technical aspects of data integration. For example, unifying the data format by using a common programming language. In the domain of PLM using XML and recently LML has been extensively used.

The second category demonstrates the techniques, which are used to reduce uncertainties, gain new knowledge or to promote the use of a specific information/knowledge for facilitating its use. In this context, aspects of improving information can be seen from the viewpoint of enhancing the quality of information or improving flow of information. For an instance, improving flow of information is categorized to the following (Khademhosseinieh & Khan, 2009): send, receive, store, find , distribute, aggregate, categorize, extract, filter, version, index, log, interpret, customize, process, migrate, backup, evaluate and activities of data route.

The third category of approaches to enrich information in the lifecycle are the techniques, which focus on decision-making. Therefore, they help the stakeholders in rationalizing the alternatives, reducing the uncertainty of decision and enhancing the collaboration among the stakeholders. Table 34 shows some of the major approaches and tools which is used in the product lifecycle for enhancing data, information, knowledge or collaboration.

Table 34: Tools and approaches for enriching data, knowledge, information from the product lifecycle

Data	Knowledge	Information	Collaboration
Information systems	Knowledge-based engineering (KBE)	Key performance indicators	Video meeting
Digital twins	Semantic networks	Decision support systems (recommender system)	Collaboration tools: email, phone, chat applications
Big data technology	Knowledge discovery systems	Decision analysis by using pay off table, possible decision criteria, expected monetary value, risks, decision tree, sensitivity analysis	Agent-based simulation
Data warehousing	Configuration management	Optimization	
XML	Simulation	Data analysis and visualization	
Simulation	Knowledge management	Reporting system	
	Ontology engineering	Workflow management & document management (Ameri & Dutta, 2005)	
	Optimization	Search engines	

In the following, these methods and approaches are described.

Decision support systems. This systems support organization in decision-making. The main modules include (a) decision analysis module (b) database and (c) optimization models module. Organizations can receive alternative solutions for decision-making or get insight into their decision context from the information that the system provides them. For more information, refer to (Glover, et al., 1992).

Simulation. One of the technologies to extract knowledge and reuse it in the PLM for the aim of improving the PLM is a multi-agent system. Mahdjoub et al. (Mahdjoub, et al., 2010) proposed using the multi-agent system technology in collaborative product development. By integrating the knowledge based on the design processes it can help designer to use the engineering know-how more effectively.

In another work, Xiao et al. (Xiao, et al., 2010) proposed adding the simulation data of virtual products to the PLM systems. They presented an architecture for this model, which visualize the geometry, tasks, virtual prototypes and information related to simulations.

Ontology engineering. Ontologies are used to define concepts in a specific domain and the relationships between these concepts within a particular domain (Kuhn, et al., 2008). El Kadiri & Kiritsis (El Kadiri & Kiritsis, 2015) provided a comprehensive review of ontologies in the context of product lifecycle managing. Some examples of application for this field in the PLM are the models, which try to represent the aspects of product design, in other words “functions of concepts, to mathematical models for exploring and optimizing the design space, to multidisciplinary models, which are used to represent the behavior of designs” mathematically or virtually (Chandrasegaran, et al., 2013). Identifying and using the ontologies in the different PLM phases is still a research challenge and for example, defining ontologies for a product and its taxonomies, defining the concepts and rules of performance monitoring and quality control during the manufacturing phase of the product are under study.

Semantic networks. In a semantic network, all relationships between members of a group are clarified by the context, in other words, by the meaning of the data. Unified Modeling Language (UML), Bayesian networks, Petri nets belong to semantics networks (Jun, et al., 2007).

Knowledge management. The knowledge management aims to collect the knowledge, which exits across the organization and store it in a database or centrally. This approach enhances the effective use and reuse of knowledge. The types of captured knowledge in the PLM systems are the product, process and organization’s knowledge.

Knowledge-based engineering (KBE). This approach is based on knowledge of experts in design process and helping design related problem-solving task. It uses object-oriented approach. Already used in aeronautics and automobile industry. KBE is designed for a specific area of knowledge, due to its limited capacity; it is sensitive to size of the dataset.

10.2 Appendix B: Stakeholder analysis Freeman results

More details of applying Freeman's four-step method is provided in this appendix. For identification of stakeholder groups first, several groups of stakeholders are identified. For example, for a commercial airplane, this research initially found 51 stakeholder groups who are in direct contact with these engineered products. These groups are further analyzed. The analysis is done based on scoring their impact or influence of the stakeholder on the product and its PUI. This analysis aims to identify the stakeholder groups, which are the most important and influential on the product. Stakeholder analysis matrix from (Bryson, 2003; Freeman, 2010) is used for assessing the importance of stakeholders and mapping them. For applying this method first each of identified stakeholder groups (e.g., each of 51 groups) are ranked twice. The ranking criteria were based on 1) power/ influence 2) concern/ interest. Powers shows the extent of stakeholders' power or influence on the product/service. Interest or concern means the extent the product affect the stakeholder, namely, the extent the stakeholder is interested in the product or is concerned about it or its services.

For every criterion a Likert scale is used. The scales of measurements are: unknown or not at all (0)/ little (1)/ some (2) /significant (3). Figure 56 shows the schema of this ranking matrix.

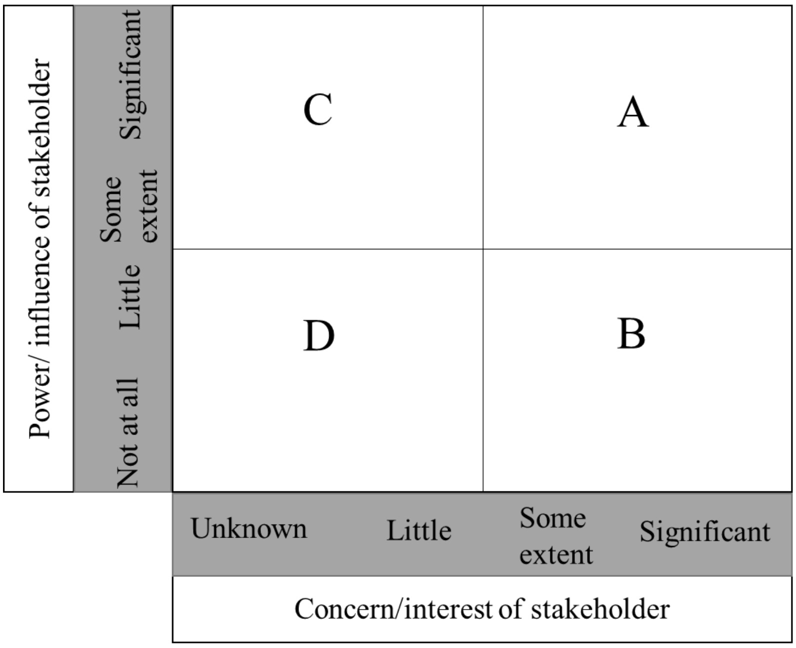

Figure 40: Stakeholder-mapping matrix (DELWP, 2018)

A similar method is used for stakeholders of MOL for airplane and wind turbine to assess their power and influence on the product and its lifecycle. The results of this analysis for the major stakeholder groups are presented in Figure 41 and Figure 42.

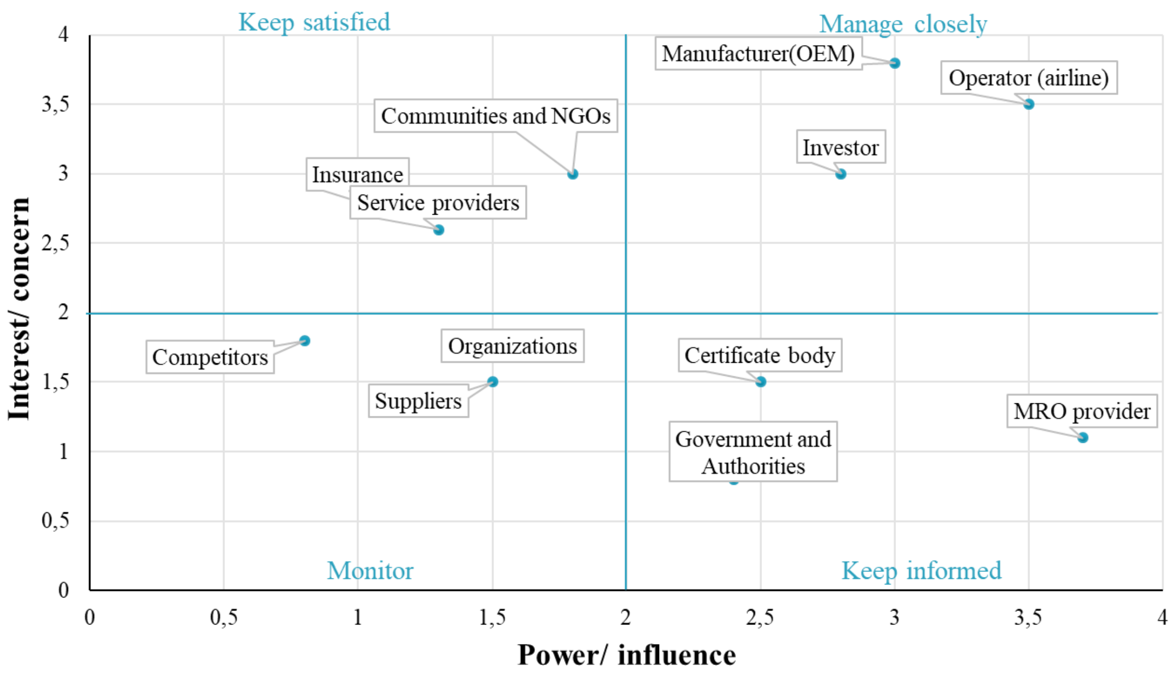

Figure 41: Stakeholder mapping by power/ interest assessment for wind turbine

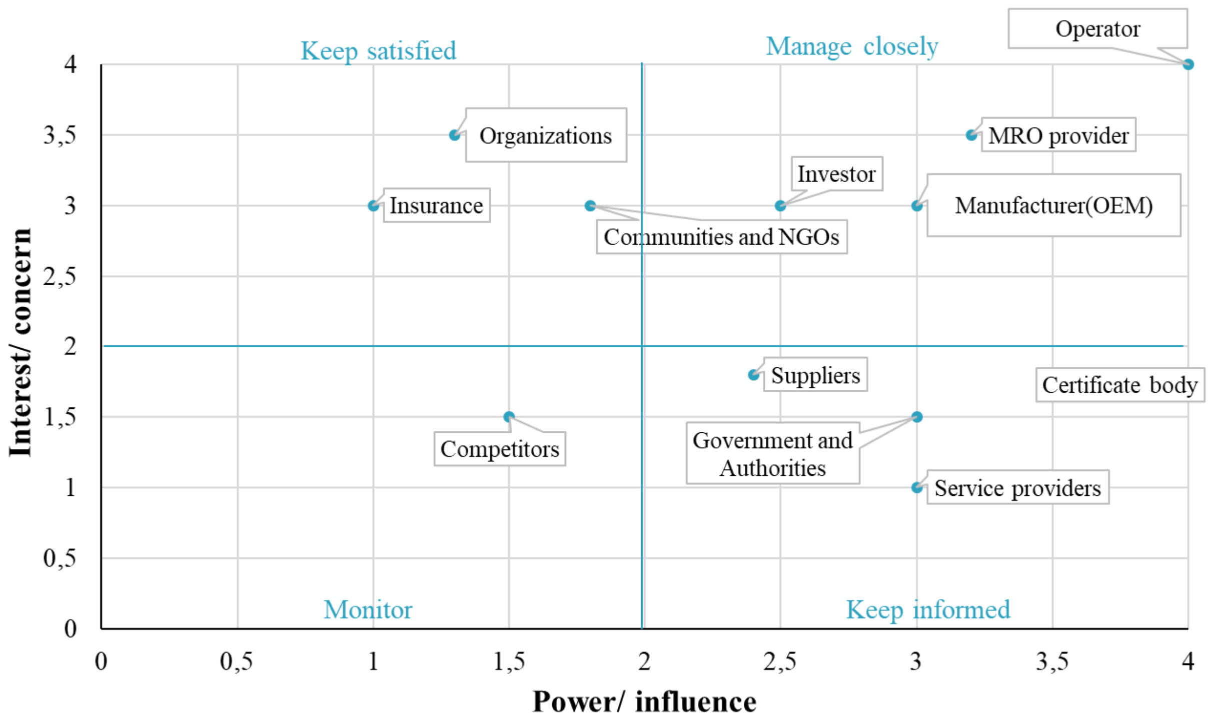

Figure 42: Stakeholder mapping by power/ interest assessment for commercial airplane

From Figure 41 and Figure 42, we can find out that operator, MRO, investor and OEM has the highest scores among other stakeholder groups. Therefore, they play a more important role in respect to affecting the product or being affected by data and information in the product lifecycle. More information is provided in chapter 4.

10.3 Appendix C: Support Vector Regression (SVR)

SVR uses a subset of training data, known as support vectors and map them to produce regression boundaries (Hastie, et al., 2009). Equation 1 shows the general formula of SVR, which shows the mechanism behind building boundaries. In equation 1, $f(x)$ is the function, which produces boundaries, x is the vector containing input variables for prediction, x^t is transpose matrix of x. β is a vector of weights given to each variable x for balancing them and tuning the model. β_0 is the intercept. The value of β_0 is equal to expected mean value of $f(x)$ when x=0.

$$f(x) = x^t\beta + \beta_0 \tag{1}$$

In the simplest form, an SVR resembles a linear regression, which divides the feature space by a line into two parts. Please see Figure 43. The line in the figure shows the boundary produced by SVR.

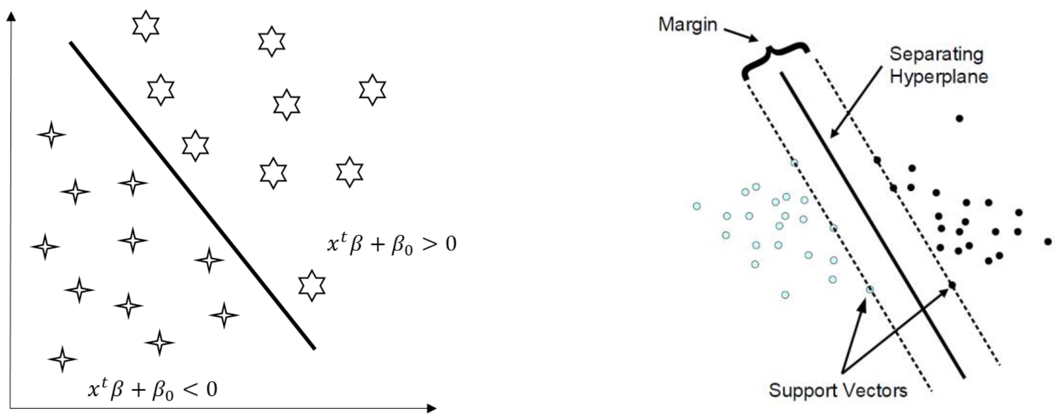

Figure 43: Simple demonstration of support vector regression (adopted from (Wuest, 2014))

When x has higher dimensions, the weights of each variable in vector x should be calculated. For calculating these weights in general, the following equation should be solved:

$$H(\beta, \beta_0) = \sum_{i=1}^{N} V(y_i - f(x_i)) + \frac{\lambda}{2}\|\beta\|^2 \tag{2}$$

And

$$V_{\in} = \begin{cases} 0 & if \ \ |r| < \in, \\ |r| - \in, & otherweise \end{cases} \tag{3}$$

Equation 3 is called loss function. Loss function shows a measure of the error in the model. It calculates the residual of $y_i - f(x_i)$. Moreover, it defines a criteria for handeling errors (equation 3). Eplison (∈) is a parameter for adjusting level of accepted error in the model. Base on this formula, errors of size less than ∈ are ignored and considered equal to 0. It means that points on the correct side of decision boundary and far away from it, are ignored in optimization.

If $\hat{\beta}$ and $\hat{\beta}_0$ are the estimed values, which minimize the function H, the solution function, can be shown to have the form:

$$\hat{\beta} = \sum_{i=1}^{N} (\hat{\alpha}_i^* - \hat{\alpha}_i) x_i, \tag{4}$$

$$\hat{f}(x) = \sum_{i=1}^{N} (\hat{\alpha}_i^* - \hat{\alpha}_i) \langle x, x_i \rangle + \beta_0 \tag{5}$$

In equation 4 and 5 , $\hat{\alpha}_i^*$ and $\hat{\alpha}_i$ solve the quadratic programming problem, as follows:

$$\min_{\alpha_i, \alpha_{i,}^*} \epsilon \sum_{i=1}^{N} (\alpha_i^* + \alpha_i) + \sum_{i=1}^{N} (\alpha_i^* - \alpha_i) y_i + \frac{1}{2} \sum_{i,\acute{i}=1}^{N} (\alpha_i^* - \alpha_i)\ (\alpha_i^* - \alpha_i) \langle\, x_i, x_i \rangle \tag{6}$$

In high dimensions (more than one variable), the boundary is no more linear. It resembles a curve. For finding this boundary, the following equation should be solved.

Subject to: (7)

$$0 \leq \alpha_i, \alpha_i^* \leq {}^{1}/_{\lambda},$$

$$\sum_{i=1}^{N} (\alpha_i^* + \alpha_i) = 0,$$

$$\alpha_i^* \alpha_i = 0$$

The parameters, ϵ and λ, have different roles in creation of model. ϵ is a parameter of the loss function V_ϵ. Note that V_ϵ depend on the scale of y and hence r.
The other parameter, λ is a regularization parameter. It value can be estimated for example by cross-validation (Hastie, et al., 2009).

Tuning SVR parameters. As mentioned before SVR model has several parameters. Here tuning of the following parameters, for our SVR model of EV battery charge prediction, are discussed: (a) Kernel type, (b) Cost function and (c) Epsilon.

(a) The type of kernel function used in this study is radial kernel. Kernel functions are varied. SVR can use kernels such as radial, polynomial and linear kernels. In this study, we used radial kernel. Radial kernel (Hastie, et al., 2009) is used because of its advantages such as simple design, good generalization, strong tolerance to input noise, and online learning ability. Therefore, it is selected for this study. The linear kernel has less tolerance and flexibility in a higher dimension (Yu, et al., 2011). Therefore, it is not the right choice for our dataset. For more information regarding different applications of kernels please, refer to (Hastie, et al., 2009).

(b) Cost parameter is another important parameter of SVR model. Cost parameter adjusts the model fitting and avoids overfitting of SVR model on data. Avoiding overfitting is essential to reduce bias of model and avoid modeling the noise instead of real values. For more information regarding overfitting please refer to (Hastie, et al., 2009). In this study, we try values 1 to 20 for the cost (with 0.2 step increase between each of the two tested values).

(c) Next parameter, which needs tuning is epsilon (ϵ). We try different values between 0 to 1 for epsilon to find appropriate amount (Note that ϵ can have a range between $0 \leq \epsilon \leq 1$).

The methodology for parameter tuning here is grid search. By doing a grid search (Kowalczyk, 2014), we build and train many models for the different values of ϵ and cost. Moreover, for selection of best combination of parameters "10-fold cross-validation" is used. This method is described in chapter 7. Range of values for epsilon and cost are as follows; epsilon = seq (0,1,0.1), cost = seq (1,20, 0.2).

The final SVR model is built after applying tuned values. It has the following values for epsilon and cost: Epsilon =0.1 & cost= 3.4. The new model is reported in subsection 6.2.3.

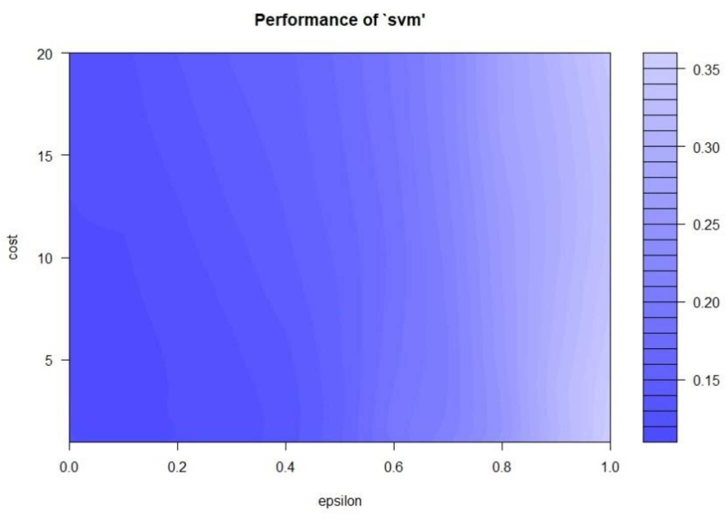

Figure 44: Visualization of results of parameter tuning (tune 2)

10.4 Appendix D: MOL Stakeholders of EV and discussion on validity

In this section, the stakeholders of an EV are shown. In order to implement the concept of stakeholder's information needs (chapter 4) on the EVs there is a need to assure that this concept is valid and applicable on the EVs. In this section, first stakeholder identification is carried out for EV and later the results are compared with the identified MOL stakeholders in chapter 4 (subsection 4.4.4). This identification has been done independently from the two scenarios of wind turbine and commercial airplane.

The information and comparisons provided in this section are used for showing the validity of EVs as a relevant scenario in the dissertation; as well as validating the concept of stakeholders' information needs.

Table 35: MOL Stakeholder identification for EV

Stakeholder group	Role	Typical internal stakeholders in this group
Car user	Owner of car, operator, car renter	OEM, end-user, car rental organization, logistic companies, organizations
Municipality	Effect of car on city, traffic zoning, parking place provider, charge place provider	Traffic directors, city and urban logistics planners, city development
Electricity provider	Produces electricity or buys energy (renewable or classic) from electricity produces	Electric utility company, power distributers, retailers in electricity market, charging network operator, charging station owner
General public	Are affected by sound, noise and environmental impact of electric car	People in the neighborhood, society
OEM	Long-term viability of business, compliance with emissions regulations; freedom to develop product portfolio (Bakker, et al., 2014)	Assemblers, machinists, production managers
Spare parts supplier	Provide components and spare part, supply new systems for electric cars, new gear boxes, electric power steering, water, pumps to cool the electric engine, battery packs and cell component (Todd & Thorstensen, 2013)	Automotive part suppliers, automotive components suppliers, battery makers, cell component makers
MRO	Provide maintenance and service	Automotive service technicians, mechanics

Stakeholder group	Role	Typical internal stakeholders in this group
Communities, Educational institutes	Training emergency personnel, developing incentives and educating the public based on the experiences (U.S. Department of Energy, 2018)	Research institutes
Government	Leading transportation and climate initiatives, create, administer and amend planning processes, rules and regulations, including in zoning, parking and permitting (U.S. Department of Energy, 2018). Jurisdiction issues of electric transportation Tax incentives for purchase of cars	National government, local government
Car owner	Possess the car as an asset	Individuals, car rental services, OEM
Fossil fuel providers	Substitute product service providers, affected by decrease in demand	Tank stations, petrochemical companies, power and gas companies
Insurance	Risks of using EV, accidents involving EV	Insurance company, firefighters, police
Logistic providers	Transport the cars to the point of sell on large scale	Railway and ship lines
Service providers	Define services based on using cars.	Logistic companies (such as post)
Organizations	Provide plug-in service for charging cars, use EV for their own service such as restaurants	Retailers (restaurants, shops), Electric charging facility installer
Electricity grid operators	Provide grid stability through balanced supply and demand, provide public recharging infrastructure	Renewable energy purchaser, power plants, smart grid developer
Inspection and certification companies	EV charging system testing, supply equipment testing and certification, data communication testing (TÜV Reinland, 2018)	Certification bodies, total quality assurance provider
Investors for the infrastructure	Providing finance for charging stations and supporting infrastructure	Banks, government, local authorities, municipalities

Table 35 shows the stakeholders of EV. Comparing to the identified beneficiaries (in engineered products) which were shown in subsection 4.4.4, it can be said that most of the 12 stakeholder groups also exist in electric vehicles. To illustrate more here are some examples of this similarity between stakeholder groups. OEM is a major beneficiary of an EV. His information requirement about the awareness of product performance during the product's operation is similar to what we have identified in Chapter 4. The same is the requirement of the OEM to know the environmental effects of the product.

Regarding the differences of stakeholders of EV and findings of subsection 4.4.4 followings can be stated. In the case of an electric car, an information need for OEM is to know whether the business is sustainable in the future and whether it is profitable. This information need was not observed as important in engineered products (in this research), because it is likely that OEM of airplanes and wind turbines design and produce a product based on an issued order from a customer. Therefore, they are less likely to seek market for an already manufactured product. At the same time, in most cases, airplane and wind turbine manufacturers use state subsidies. Therefore, focus on only profit making is not the primary goal of these stakeholders.

Another difference among stakeholders is that Certification bodies have a smaller role in MOL of EVs than wind turbines and airplanes. Though organizations like TÜV checks the performance and health of components in an EV, they record little information about the processes of MOL in the EV and how the EV is being operated.

A further distinction of EV is that the owner of an EV can be a person, while in airplanes and wind turbines are more likely to be one or several companies. This affects the gathering and sharing of information to meet the information needs of MOL other stakeholders. The issue of information privacy also raises in this case.

Apart from these differences, the data and information needs of some MOL stakeholders such as OEM, MRO, spare parts supplier, government and municipality are similar to the findings of this dissertation in chapter 4. As a result, it can be assumed that the recommendation of supporting information needs with data analytics techniques (chapter 5) can be tested on EVs. Therefore section 6.2 examines the implementation of data analytics for supporting information needs of OEM from usage data of EVs.

10.5 Appendix E: Linear regression for battery charge prediction

Linear multiple regression. Linear regression is a prediction model, which describes the relationship of a response variable with two or more (several) exploratory variables. The exploratory variables can be continuous or discrete. The model for multiple linear regression given n observations and p explanatory variables is:

$$y_i = \beta_0 + \beta_{1x_{i1}} + \beta_{2x_{i2}} + \cdots + \beta_{px_{ip}} + \varepsilon_i \qquad i = 1,2,\ldots,n \qquad (9)$$

Where y is the response variable, x_1 to x_p are exploratory variables and β are the weights, ε is the random error. For more information about characteristics of parameters in such a model, please refer to (Montgomery, et al., 2012). This regression builds a linear function of exploratory variables on y. In a simple case, when only two exploratory variables exist, multiple linear regression a multiple regression model resembles a hyperplane in 2-dimensional space of two explanatory variables. Please see Figure 45 for an example of multiple linear regression.

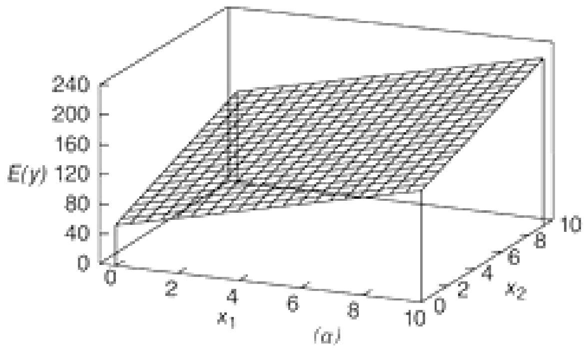

Figure 45: Regression plane when having two exploratory variables x1 & x2 (adopted from (Montgomery, et al., 2012))

In our case (subsection 6.2.1), y denotes the percent of battery charge and x_1 to x_p are the exploratory variables such as motor power, trip duration, etc., these exploratory variables are utilized for building the model of battery charge.

The importance of using multiple linear regression for our scenario in the prediction of battery charge in EVs can be established based on explanations of (Montgomery, et al., 2012). As stated by Montgomery et al. (Montgomery, et al., 2012), these models are often used as empirical models. Such empirical models are useful for approximating a function, which describes the relationship between y and x_1, x_2, …, x_p. When these variables and their effects on y are unknown, nevertheless, over a certain range of exploratory variables, the relation between them can be built. Furthermore, the linear model is an adequate approximation to the true unknown function (Montgomery, et al., 2012).

Table 36: Implementing multiple linear regression on battery charge data

Variables (coefficients)	**Estimate Std.**	**Error**	**t value**	**Pr (>|t|)**
(Intercept)	3.329e-16	0.032	0	1
Left.indicator	0.01.873	0.034	0.544	0.586
Rigth.indicator	0.0746	0.003	2.241	0.025 *
Position.hand.brake	-9.319e-03	0.035	-0.261	0.794
Ligths	.016	0.033	0.486	0.626
AC.switch	0.644	0.035	1.827	0.068 .
Brake.ligths	0.018	0.038	0.467	0.640
Drive.mode	-0.167	0.036	-4.600	4.8e-06 ***
Motor.power	5.479e-03	0.036	0.151	0.880
Trip.number	-0.215	0.037	-5.811	8.76e-09 ***
Signif. codes	0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			
Residual standard error	Residual standard error: 0.957 on 851 degrees of freedom			
Multiple R-squared	Multiple R-squared: 0.093, Adjusted R-squared: 0.084			
F-statistic	9.77 on 9 and 851 DF, p-value: 2.558e-14			
Residuals	Min 1Q Median 3Q Max			
	-2.664 -0.593 0.100 0.739 1.706			

Table 36 and Table 37 shows the results of running a linear regression on the dataset of EV. Interpretation of results of linear regression is that this model and selected variables do not adequately describe the relationship between the amount of battery charge and other measurements. The value of R-square is 0.09 (9%) and this is low for a model to be an accurate model. Therefore, for modeling battery charge algorithms, which can model nonlinear relationships, are used in this dissertation. They are explained in section6.2.

Table 37: Predicted results vs. real values for average battery charge (percent) with a linear regression model

Predicted value	**Real value**
95.80	98.00
81.54	97.33
70.83	97.25
92.86	97.00
84.62	87.25
88.76	84.86

10.6 Appendix F: Evaluation of CART models for scenario 1

Data for making the CART regression tree (Figure 25 and Table 18) is divided into test and train set. The test set contains 25% of data and train dataset contains 75% data. The following shows a part of the predicted values from our tree model versus real values of energy.

Table 38: Values of decision tree model (prediction of electricity based on affecting variables, for model error calculation)

No.	Prediction	Real energy value
1	21,049.47	61,459.48
2	267,188.73	103,463.98
3	153,270.82	77,933.64
4	214,800.73	285,652.54
5	115,643.53	156,091.64
6	153,270.82	201,576.11
7	69,828.48	68,001.51
8	69,828.48	45,446.07
9	214,800.73	272,734.33
10	21,049.47	27,683.02
11	21,049.47	29,416.89
12	214,800.73	289,424.72
13	69,828.48	54,402.96
14	153,270.82	130,084.20
15	214,800.73	122,446.48

Mean Absolute Error: 42947.1957 and RMSE: 10376.13 are the error rates for regression tree (unscaled values). Evaluation of classification tree (Table 19) for speed also was performed on the test dataset. Since this problem is a classification problem and not a regression problem, the calculation of error rate is different. The error rate for classification tree can be reported in terms of the probability that an observation belongs to a class. If the model is an entirely accurate classifier, observation belongs to one of the classes with the probability of 1 and the value of probability for other classes equals to zero.

The results of the prediction show the probability that every observation belongs to each speed range. Figure 39 shows a part of the results.

Table 39: Part of evaluation results for wind classification model

Wind speed	0-5 m/s	5-10 m/s	10-15 m/s	15-20 m/s
Probability of belonging the observation to each category	0.20	0.51	0.25	0.04
	0.29	0.64	0.07	0.00
	0.29	0.64	0.07	0.00
	0.20	0.51	0.25	0.04
	0.20	0.51	0.25	0.04
	0.29	0.64	0.07	0.00
	0.20	0.51	0.25	0.04

The average value of prediction results (on test dataset) are as follows:

0-5 m/s	5-10 m/s	10-15 m/s	15-20 m/s
0.318	0.580	0.091	0.009

The results show that the classification tree is not very accurate. For example, probably that an observation correctly belongs to class 5-10 m/s is 58%. Therefore, we use visualization and run RF (see subsection 6.1.3) on the data as an alternative and to be able to build more accurate prediction models.

11 List of Tables

12 List of Figures