MASTER

Data-driven planned maintenance for MRI-Scanners

van Breugel, R.

Award date:
2018

Graduation thesis

Data-driven planned maintenance for MRI-Scanners

In partial fulfilment of the requirements for the degree of

Master of Science

in Operations Management and Logistics

Author:	ing. R. (Rutger) van Breugel
Student ID:	0983634
University:	Eindhoven University of Technology
Program:	Operations Management & Logistics
Capacity group:	Information Systems
Supervisor:	Dr. Ir. H. (Rik) Eshuis
Second supervisor:	Dr. Ir. A.E. (Alp) Akçay
Company:	Philips Healthcare
Company supervisor:	Dr. Ir. G. (Godfried) Webers
Date:	November 23, 2018

The Technical University of Eindhoven, Faculty of Industrial Engineering & Innovation Sciences

Series Master Theses Operations Management and Logistics

Subject headings: Industrial engineering, condition-based maintenance, data-driven maintenance, planned maintenance, preventive maintenance, healthcare, artifact-centric process modeling, case model and management notation, decision intensive processes

Abstract

This master thesis describes a graduation project conducted at Philips Healthcare in Best. The research looks into optimizing the design process of planned maintenance activities on effectivity and frequency for MRI-Scanners, by improving the process and applying historical maintenance data. To do so, a new process of defining planned maintenance activities is designed and proposed, integrating planned maintenance as an integral part of the component design. Additionally, a methodology and manual are developed that describe how to review existing planned maintenance activities on effectivity and frequency using historical maintenance data. The combination can create a new process for designing planned maintenance activities, improving planned maintenance activities on effectivity and frequency.

Management summary

This research focusses on optimizing existing planned maintenance (PM) activities of Philips Healthcare, with the use of big data and artifact-centric process modeling for an actual case in professional healthcare. Currently, Philips has maintenance visitations on fixed time intervals, the so-called PM visitations. Since a few years, Philips has easier access to historical test, condition-, and usage data of their machines. Therefore, using this data Philips wants to research scheduling PM activities based on data, optimizing on effectivity and frequency, resulting in the following problem statement.

The process of designing and reviewing planned maintenance activities is not optimized for effectivity and frequency and does not use historic planned maintenance activity data

Currently, Philips Healthcare uses PM manuals that describe all the required PM activities for each product model. Subject matter experts (SMEs) in Research and Development (R&D) and Service Innovation (SI) have defined PM activities in the past, designing them based on engineering judgment. There is no documentation explaining why PM activities are in the current PM manual, the reasoning for having these activities is implicit knowledge of SMEs.

To bridge this gap between the current and desired situation, two deliverables are developed. Firstly, there is a proposed methodology on how to define the required PM activities. This methodology contains a framework with all the possible motivations and types of PM activities, creating the opportunity to categorize the required PM activities. Secondly, the proposed process contains an aspect on reviewing PM activities; this second deliverable describes how to implement the aspect of reviewing PM in a data-driven way, verified by applying this process to a subset of nine PM activities.

Structural research methodologies help to analyze the problem and design towards a solution. The problem consists of two aspects, process-related and data related, and therefore, two different research methods are used, the regulative cycle developed by van Strien (Aken & Andriessen, 2011) applied until the design phase, and the more data focused Cross-industry standard process for data mining (CRISP-DM) method, illustrated in Figure 1.

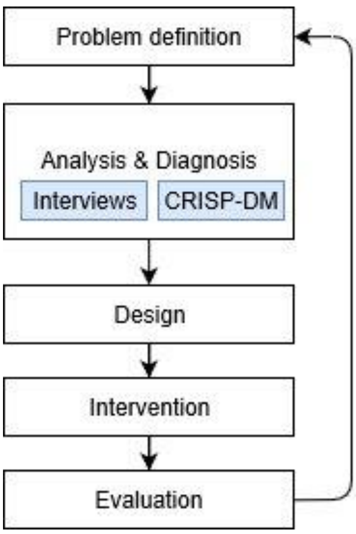

Figure 1 - Regulative cycle with CRISP-DM

During the analysis & diagnosis phase, there are two inputs for the design phase, the interviews and data analysis. The goal of the interviews is to understand the motivations and types of PM, while also gaining insight into the current process of defining PM activities, deriving the results from the interview transcriptions by using thematic analysis. The goal of the data analysis is to review the potential of the PM activities, by analyzing the opportunities and limitations using their respective historical data. In addition, the analysis looks into what valuable data is available in different data warehouses. Together, they are the input to create optimal PM.

The approach was to define the motivations and characteristics by analyzing the existing PM manual with an SME, with the interviews, followed by a verification with other SMEs to see if the lists are comprehensive. For motivations the result consisted of two main categories, (1) maintain physical integrity, and (2) risk management, where the main distinction between these aspects is that maintain physical integrity are solely choices of Philips Healthcare, and thus analysis on effectivity and frequency is of great value. This is not necessarily the case for risk management, since these activities can be mandatory for meeting legislation or standards in the healthcare industry. The process of defining characteristics was similar, and Table 1 contains the verified characteristics of PM, the percentage indicates the share in the total spent PM time.

Table 1 - Characteristics of planned maintenance

Functionality test (24%)	Calibration (6%)	Replacement (0%)	Lubrication (1%)
Inspection (25%)	Cleaning (25%)	Software revision (3%)	*Workflow (16%)*

In the current process, one of the most important learnings is that there are two departments relevant for PM activities, research & development (R&D) for the component design and thus the required PM activities, and service innovation (SI) who are responsible for prescribing the PM activities. These departments have different goals in terms of PM activities, but in the desired situation, they have a common business goal for Philips Healthcare. The proposed design of deciding on PM activities should enable a holistic evaluation considering PM during the component design, resulting in the best decision for Philips Healthcare as an organization

The relevant data consisted of three aspects; firstly, there is the Vertica database, the tables used were the maintenance history, usage data related to the number of scans, logging data, and machine configuration data. Secondly, an SAP extract was important to define what visitations were actual PM and those that were not, this data was only available since 2016. Thirdly, there were the service work orders (SWO), relevant for the selection of PM activities to analyze rather than analysis on effectivity and frequency.

Analyzing these activities was an iterative process using CRISP-DM, for each activity analysis there were learnings on how to improve or speed up the process. These learnings are then applicable for the next analysis, and as a result, the researcher gained the knowledge to create a focused methodology in an artifact-centric way to describe the process on how to do this analysis for PM activities. Although this process proved to give a conclusion with business value for the nine analyzed activities, specifically that four of these might be redundant, there was no unbiased validation of the process, which is a potential improvement in the research.

With the knowledge gained from the interviews and the data analysis methodology complete, it was possible to design the new methodology for defining the required PM activities. Based on the learning that the component design determines which PM activities are required, optimization for effectivity and frequency requires integrating PM in the design choices of the component design. The final component design is dependent on multiple component attributes, where one of these is PM. In the design proposal, the choice of the component design is a holistic consideration between all component attributes, and their respective lifetime costs compared to the component costs of the design, making the best choice for Philips Healthcare.

With this research, Philips Healthcare can start analyzing PM activities with the designed methodology. The analyzed sample of PM activities shows that a group of activities may be redundant, where the extrapolated saved time by not having to perform PM activities is large enough to justify the required effort. Therefore, the recommendation is to analyze the activities for effectivity and frequency using the designed data analysis methodology.

Finally, in terms of extending this research on academic grounds, there is potential for condition-based maintenance using daily available data, which is an opportunity for future work. The current research successfully used artifact-centric process modeling for doing data analysis, but it would be interesting to explore the use of artifact-centric process modeling on condition-based maintenance, especially in combination with decision-intensive processes. This is feasible for corrective maintenance activities at Philips Healthcare using the described opportunities for condition-based maintenance that have daily available data.

Acknowledgements

Firstly, I would like to thank my company supervisor Godfried Webers of Philips Healthcare. I have to thank him for making time for me whenever required, discussing the topics in a very structured manner. An example is his initiative of scheduling a weekly meeting for the entire period of the internship. Nevertheless, I could always approach him outside of these meetings, where his input during the Internship was indispensable, bringing my thesis to a higher level both on practical application for Philips Healthcare as academic standards for the university.

Outside of the excellent supervision from the company, there was strong support from the University of Eindhoven as well. Under the direct supervision of Rik Eshuis, I always got a quick response to any questions or concept versions. Additionally, I was always welcome for a meeting, and he always made sure he had time for me when needed.

In addition, there was the supervision of Alp Akçay. With his knowledge of maintenance, he was a strong supplemental supervisor. I would like to thank him for his notes on the research proposal and his valuable domain knowledge.

Outside of the people who have a direct influence on this thesis, I would like to acknowledge my parents. They have always supported me during my studies, from acquiring my high school diploma, graduating from a University of applied sciences, up to now, the final stage of a Masters diploma.

Contents

Table of Figures

List of Tables

Table of Abbreviations

Abbreviation	Meaning
BA	Business artifacts
BALSA	Business Artifacts, Lifecycles, Services and Associations
BPM	Business Process Modeling
BPMN	Business Process Modeling Notation
CBM	Condition-Based Maintenance
CM	Corrective Maintenance
CMMN	Case Management Model and Notation
Crisp-DM	Cross-Industry Standard Process for Data Mining
DIP	Decision Intensive Processes
DMN	Decision Model and Notation
e.g.	Exempli Gratia - "For example"
ERP	Enterprise Resource Planning
GSM	Guard-Stage-Milestone
IS	Information systems
MR	Magnetic Resonance
MRI	Magnetic Resonance Imaging
OEM	Original Equipment Manufacturer
OMG	Object Management Group
PM	Planned Maintenance
R&D	Research & Development
RQ	Research Question
SI	Service Innovation
SME	Subject Matter Expert
SQ	Sub-Question
SQL	Structured Query Language
SRN	System Reference Number
SWO	Service Work Order
TU/e	Technical University Eindhoven
UML	Unified Modeling Language

1 Introduction

In terms of accounting, maintenance is a cost to preserve an asset's operational status. Therefore, doing maintenance only when it is critical enhances the financial margins. Companies that have a combination of good product reliability and that can accurately predict maintenance costs will have an important competitive advantage (Meeker & Hong, 2014).

This research focusses on improving existing planned maintenance (PM) of Philips Healthcare, with the use of big data and artifact-centric process modeling for an actual case in professional healthcare. Where big data are datasets too large to process with traditional methods, artifact-centric process modeling focuses on the flow of business information artifacts rather than of material artifacts (Nigam & Caswell, 2003). There is research on both data-driven maintenance as artifact-centric modeling, although the combination of these research areas is untouched.

Traditional maintenance is an activity-centric process; there is a strict sequence of activities when maintaining a machine. Traditional business process models do not consider the informational perspective and informal context entirely, while modeling processes in an artifact-centric way consider the changes and evolution of a process in an informal context.

Condition-based maintenance (CBM) is one form of data-driven maintenance; it attempts to avoid unnecessary maintenance tasks by performing operations only when there is quantitative evidence, (Liao, Wang, & Pan, 2012) with the recent expansion of data sources, it is easier to implement data-driven CBM (Bumblauskas, Gemmill, Igou, & Anzengruber, 2017), giving potential to optimize maintenance strategies, reducing costs and unplanned downtime.

The product is MRI-Scanners in professional healthcare, produced by Philips Healthcare, as described in the company background in Appendix A. Philips operates as the original equipment manufacturer in this supply chain, installing machines worldwide. An important performance criterion is downtime, which can result in treatment rescheduling and other costs.

Currently, Philips Healthcare has PM visitations on fixed time intervals, where they complete maintenance activities, the so-called PM activities. Since a few years, Philips Healthcare has easier access to historical test, condition-, and usage data of their machines. Therefore, using this data Philips Healthcare wants to research the feasibility of scheduling PM activities based on available data improving their PM activities in terms of costs while not decreasing the machine performance and enhancing the customer satisfaction.

1.1 Problem Context

Philips Healthcare currently uses planned maintenance (PM) manuals that describe all the required PM activities for each product model. Subject matter experts (SMEs) in Research and Development (R&D) and Service Innovation (SI) have defined PM activities in the past, designing them based on engineering judgment. There is no documentation explaining why particular activities are in the current PM manual, knowledge of why and how these activities are structured is implicit knowledge of employees.

Because this knowledge is implicit, it requires substantial effort and therefore time, to re-engineer the reason for having certain PM activities. There are no explicit definitions for why there are PM activities, nor types of PM activities. As a result, reviewing or discontinuing existing PM activities rarely happens. Although history proves that the current set of PM activities achieves its goal in terms of safety, Philips Healthcare believes there is potential to enhance customer satisfaction and reduce PM duration and costs.

Typically, two scenarios are seen that drive the current PM activities. First, there is historical repair data, such as part replacement records. Second, there are the guidelines R&D engineers provide. These guidelines indicate when the engineers would expect the equipment to require service based on the design. To make their estimates, they re-use the guidelines from older models without reevaluating them. Therefore, these processes are too dependent on expert knowledge, instead of on facts, that can cause activities to be redundant or to be performed too frequently.

For the last five years, Philips Healthcare has used a big data resource called Vertica. This resource contains, among other things, historical data on all the PM activities, while making accessing the raw data simpler. As noted, it is unknown to the researchers if historical data analysis ever happened, and if PM activities are quantitatively motivated. Therefore, there is an opportunity to analyze and review the existing PM activities on effectivity by analyzing the historical data.

1.2 Research goals

The goal of the research is to define solutions to optimize planned maintenance (PM) activities for effectivity and frequency. Firstly, in the desired future, formulating PM activities happen in a structural way. When formulating PM activities, evaluation of components happens in a consistent manner, creating insight in the reasoning behind PM activities. This is in contrast to the current method, where the process is implicit and consists out of reusing old lists with slight modifications, possibly resulting in missing or redundant activities.

Activities may become redundant after a redesign. In the desired situation, Philips Healthcare will be evaluating the PM activities using the historical PM activity data, quantitatively verifying PM activities on effectiveness and frequency. This analysis enables Philips to take action, like omitting redundant PM activities or reconsidering the PM frequency.

By quantitatively motivating what activities to execute during the PM visitation from historical data, Philips Healthcare can (1) reduce the number of PM activities, resulting in a cost reduction, and (2) execute only the necessary activities, at the right time, reducing breakdowns.

Breakdowns have the highest impact on customer satisfaction and maintenance costs of Philips Healthcare. Eventually, the goal is to move from a time-heuristic PM strategy to a data-driven PM strategy, to minimize breakdowns. This research is the first step, by evaluating the design process of PM activities and analyzing existing PM activities based on their historical data.

Therefore, to go from the current situation as described in 1.1 Problem Context to the desired situation, it is required to start (1) designing PM activities via an explicit process, (2) with clear motivations (3) and goals while (4) Reviewing the PM activities whenever required, optimizing the activities for effectivity and frequency, as illustrated in Table 2 - GAP analysis.

Table 2 - GAP analysis

As-Is	To-Be
Effectivity per PM activity not quantitatively proven	Effectivity per PM activity quantitatively proven
Required frequency per PM activity not quantitatively motivated	Required frequency per PM activity quantitatively motivated
Motivations for doing PM activities implicit	Motivations for doing PM activities explicit
Types of PM activities implicit	Types of PM activities explicit
The design process of PM activities unstructured	The design process of PM activities structured

1.3 Deliverables

To bridge the gap between the current and desired situation, two deliverables are required and therefore developed. Firstly, there is a methodology for defining the required PM activities. This methodology contains a framework with all the possible motivations and types of PM activities, creating the opportunity to categorize the required PM activities. Since these are dependent on the product design, such a framework categorizes all the activities. Additionally, it gives some insight into what potential categories R&D has not thought of.

Because of the limited timeframe of this research, it is not feasible to analyze each PM activity. Therefore, the second deliverable of the research is a data analysis methodology on how to review existing PM activities based on PM activity data, proving the effectiveness and defining the required frequency in a quantitative manner. Reviewing with this data does not happen in the current situation; therefore, this is an addition to the existing process.

Combining these deliverables redesigns the existing process that defines the PM activities. Figure 2 – Desired process of defining PM activities illustrates the existing process, marking the deliverables that will result from this project.

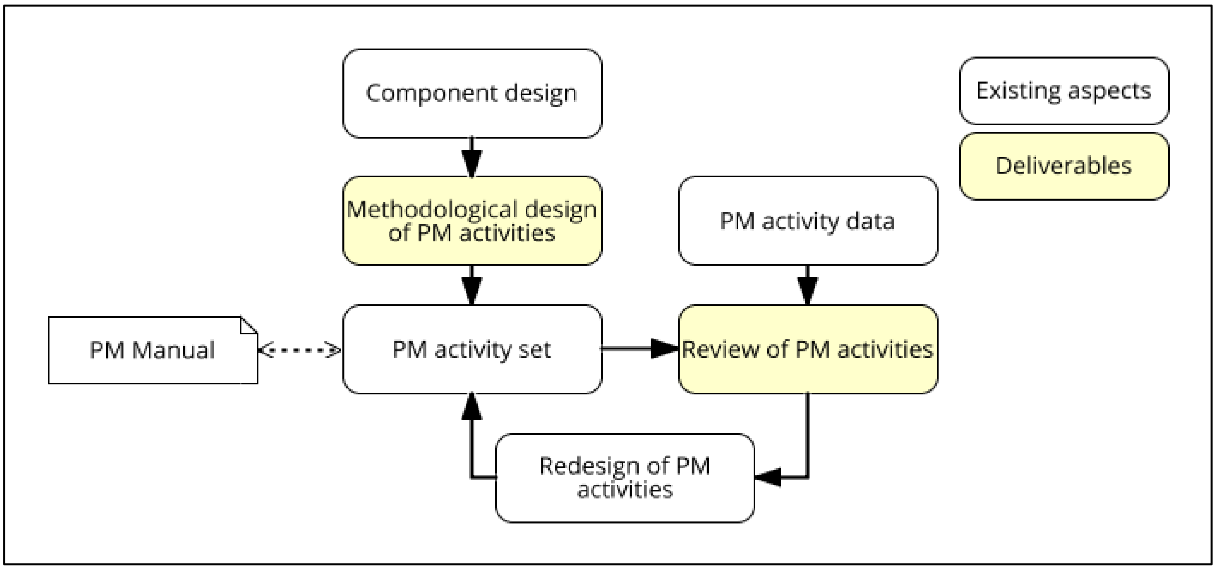

Figure 2 – Desired process of defining PM activities

1.4 Problem statement

The problem statement summarizes the goal of the research and its relevance.

The process of designing and reviewing planned maintenance activities is not optimized for effectivity and frequency and does not use historic planned maintenance activity data.

1.5 Research questions

The research goal is to develop a solution to solve the problem statement. To do so, formulating research questions (RQ) with respective sub-questions (SQ) for the current and desired situation, this gives the tools and knowledge to create a design to solve the problem statement.

Currently, there is a lack of knowledge about the PM activities, and to gain insight the question is where the current PM activities come from. The goal of this question is to understand the why and how of the formulation in the current list of PM activities, understanding the implicit motivations, types and the entire unstructured design process of PM activities, providing the input to bridge the gaps caused by a lack of knowledge.

RQ1 How are planned maintenance activities currently designed?

- SQ1.1 What are the motivations for doing planned maintenance activities?
- SQ1.2 What are the types of planned maintenance activities?
- SQ1.3 Who determines the current planned maintenance activities?
- SQ1.4 How are the current planned maintenance activities determined?

To understand the potential of PM activity data for quantitatively proving PM activities, the input to bridge the first two gaps, it is required to understand what data is available and relevant.

RQ2 What PM activity data is available to analyze current planned maintenance activities?

- SQ2.1 What are the opportunities and limitations of planned maintenance activity data?
- SQ2.2 What other relevant data is currently available at Philips Healthcare?

Knowing the potential of the data, designing a methodology to apply this data to create quantitative proof bridges the gap of proving PM activities on frequency and effectivity.

RQ3 How should the selected data be used to review decision making for PM activities?

- SQ3.1 How to evaluate the effectiveness of a PM activity?
- SQ3.2 How to evaluate the defined frequency of a PM activity?

The next step is to apply the knowledge from RQ1 and RQ2 to design the desired situation, a structured design process using explicit definitions of PM activities motivations and types.

RQ4 How should the process of designing planned maintenance activities look like?

- SQ4.1 Who should determine the planned maintenance activities?
- SQ4.2 How should they determine the planned maintenance activities?

1.6 Scope

With nearly 30 types of product models, there are many different MRI-Scanners. In addition, the data volume is vast, with over 100TB/year, and a velocity of about 40-70 GB/day in compressed text form. To be able to handle this big data set, the research is limited to the following aspects:

- The analysis focuses solely on planned maintenance activities and PM activity data
- Since defining PM activities happens per Product model, the distinction per product model makes sense. The focus is on the MRI-machines 'Ingenia', since these are recent but still have a few years of historical data, while having a representative installed base.

1.7 Scientific relevance

Data-driven predictive modeling for maintenance is not unique (Hsu & Chen, 2016). There is even some research on predictive modeling for maintenance in healthcare (M. A. Patil, Patil, Krishnamoorthy, & John, 2016). The scientific relevance of this research exists in combining data-driven predictive modeling with business artifacts and decision intensive processes.

Nigam and Caswell introduced the concept of business artifacts (BA) and the notion of modeling business processes in terms of artifact lifecycles (Nigam & Caswell, 2003). Pioneer of the business artifacts paradigm is Richard (Rick) Hull. He led a group performing research on data-centric workflow and BPM. A primary focus was on the use of a novel marriage of data and process. The goal was to build a unifying foundation for the management of data-centric workflows and business operations (Hull, 2018).

There is one extensive research found on data-intensive processes in relation to business artifacts. It stated that decision-intensive processes are similar to case management; a user has much more control over the possible flow choices. Therefore, modeling DIP processes in an artifact-centric way has several advantages. The hierarchical structuring corresponds to how the domain experts think about their activities, improves process understanding, and support structured visibility, tracing and visualization of decision tasks (Vaculín et al., 2011).

Combining the scientific knowledge of business artifacts, decision intensive process and data-driven maintenance to optimize a (planned) maintenance strategy is unexplored to the best of knowledge of the researcher. This research tries to apply this new way of working, to the case of planned maintenance activities at Philips Healthcare for MRI-Scanners.

1.8 Report structure

In the introduction, there is an exact description of the problem context and its distillation to a researchable subject. With determining the relevant deliverables, problem statement, research questions, scope and scientific relevance, this document continues by describing the core of the research. This paragraph attempts to explain the structure of this report, to understand how to read the report and why all the aspects are relevant.

Firstly, there is a theoretical background to gain insight in the literature related to the subject. Based on a systematic literature review, this chapter goes in-depth about maintenance, artifact-centric modelling and decision intensive processes, to gain sufficient theoretical knowledge, and to be able to achieve the research goals, an example is to learn how to create an artifact-centric process design.

With the problem environment description of chapter 1 and the theoretical background of chapter 2, there is the description of a methodology on how to solve the problem in chapter 3. In this chapter, there is a precise description of how the research is performed, by describing the used methodologies and the activities per phase. Additionally, this chapter includes descriptions on the gathering of the input, including what input, the process of how the input is gathered, and the application of the input in the final design.

After a detailed description of how the research is performed in chapter 3, then chapter 4 describes the actual execution of these steps in terms of gathering, analyzing and diagnosing the input. Therefore, this chapter gathers all the relevant data, describes how it will be relevant to the design, and what the key aspects are that influence the design, effectively answering the RQ1 and RQ2, that relate to the current situation.

With all the preparation in gathering the right input, chapter 5 delivers the result to solve the problem context, thus designs of the deliverables that answer RQ3 and RQ4 of the desired situation. From answering these RQs, it is feasible to answer the problem statement and thus solve the business problem of Philips Healthcare.

Finally, the research concludes with the key aspects in the research, including the limitations of the research. These limitations are then the input for formulating recommendations on improving the current work, future academic research and implications for Philips Healthcare.

2 Theoretical background

The theoretical background goes in-depth about (data-driven) maintenance, decision-intensive processes and business artifacts. The aim of this chapter is presenting relevant work while identifying the gap of knowledge in existing literature. This chapter merely illustrates the key findings, for used databased and other factors required for reproducibility; the entire systematic literature review is attached in Appendix G.

The value of the theoretical background is illustrated using the conceptual project design model (Aken & Andriessen, 2011) illustrated in Figure 3. The box in the top-left contains the theoretical aspects, illustrated in the top-right; they form the basis of data-driven planned maintenance (PM). Theoretical knowledge is a critical success factor in introducing and implementing a new concept in existing business.

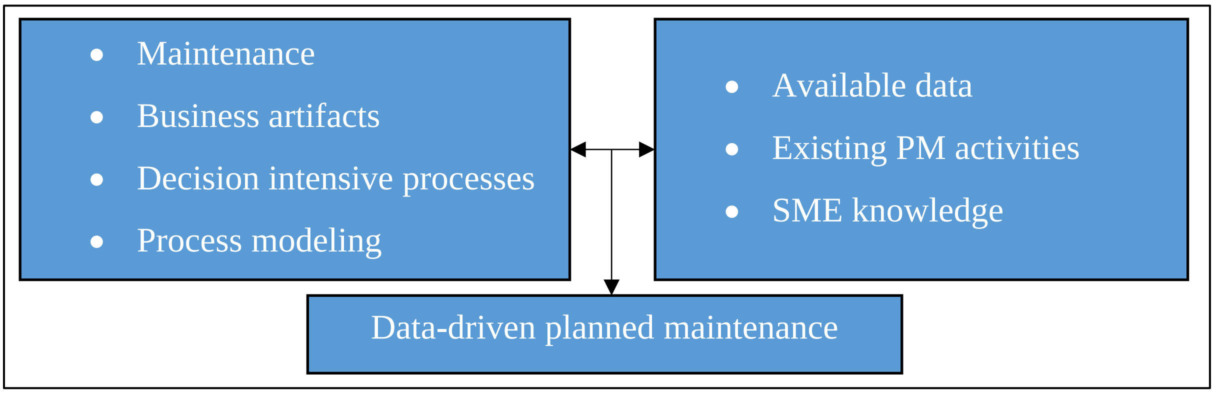

Figure 3 - Conceptual project design

2.1 Maintenance

There is an increasing demand of maintenance management policies as well as associated information systems in order to reduce unexpected failures, eliminate unscheduled downtimes, and minimize maintenance-related costs (Alexandros Bousdekis, Papageorgiou, Magoutas, Apostolou, & Mentzas, 2017; Mosaddar & Shojaie, 2013). Carrying out PM simply on some fixed schedule is labour-intensive and ineffective (Sipos, Fradkin, Moerchen, & Wang, 2014).

Information systems (IS) that help predict maintenance and determine costs has been improved in recent years because of faster processing, better storage capacity, and improved analytics tools (Bumblauskas et al., 2017; Meeker & Hong, 2014; Trunzer et al., 2017). The availability of real-time data allows organizations to react proactively to avert problems or make improvements to processes (Buchmann, 2014).

Maintenance is traditionally preventive maintenance, including PM and condition-based maintenance (CBM), or corrective maintenance (CM), as illustrated in Figure 4. (Cipollini, Oneto, Coraddu, Murphy, & Anguita, 2018; Huiguo, Rui, & Pecht, 2009; Niu & Pecht, 2009). CBM has been widely used and successfully applied in other industries (Ayo-Imoru & Cilliers, 2018). It incorporates reliability models, knowledge of equipment degradation as well as historical breakdown trends (Guan et al., 2011). The goal of preventive maintenance is to extend the period of time during which the system will function as desired and reduce the number of unnecessary delays and failures (Kisi, Durovic, Kovacevic, & Petrovic, 2015).

The goal of CBM is maintenance at the right time, this is not always easy to achieve (Niu & Pecht, 2009). An overall framework to introduce Condition-based maintenance (CBM) has a lack of grasp on reality (L. Wang, Qian, Li, & Liu, 2017). MIMOSA developed the most evolved standard for an overall framework named OSA-CBM (Schmidt, Wang, & Galar, 2017). Additionally, other authors created frameworks for CBM in three stages; (1) data acquisition to collect information, (2) data processing to handle information and (3) decision-making following maintenance policies (A Bousdekis, Magoutas, Apostolou, & Mentzas, 2015; Catelani, Ciani, & Venzi, 2017; Last, Sinaiski, & Subramania, 2011).

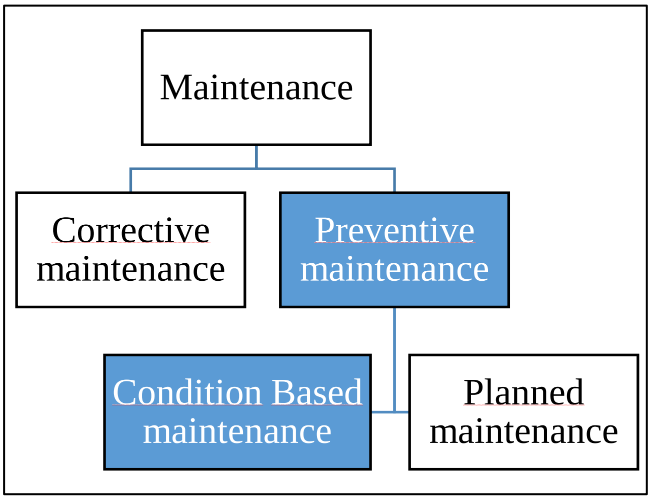

Figure 4 - Types of maintenance

Failure prognostics methods can be classified into three main approaches: physical-based, experience-based, and data-driven prognostics (Ayo-Imoru & Cilliers, 2018; Kinghorst et al., 2017; Kisi et al., 2015; Tobon-Mejia, Medjaher, Zerhouni, & Tripot, 2012). Each approach has advantages and disadvantages; consequently, combining them in hybrid applications is common (Galar, Thaduri, Catelani, & Ciani, 2015). Their location in the CBM paradigm is illustrated in Figure 5 based on (Asmai, Hussin, & Mohd Yusof, 2010).

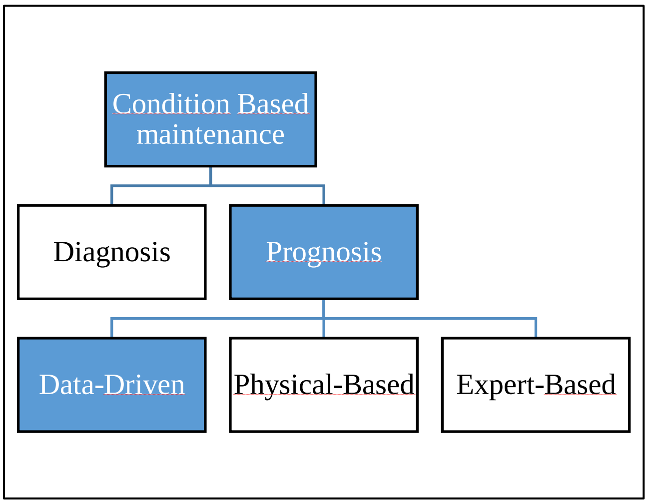

Figure 5 - Types of Conditional Based Maintenance

2.2 Data-driven aspects

This research focuses on a data-driven method based on log files for analyzing required PM activities; therefore, it is valuable to review the literature using log files for analysis, to gain insight into the different methods to implement this. There have been many success stories, even for medical systems at Philips (R. B. Patil, Patil, Ravi, & Naik, 2017). Like many databases, maintenance or log databases may contain noisy, redundant or missing values (Mosaddar & Shojaie, 2013). It is essential to review the quality of the data since this directly affects the quality of the analysis results.

Different forms of prediction techniques have complementary strengths and limitations. By combining diverse techniques, it is possible to improve the performance of prognostics systems (Baptista et al., 2018). There is a large number of techniques, divided into machine learning and statistical techniques. Both statistical methods as machine learning have their advantages and disadvantages. The essential difference is that with statistics they fit a model to the available data, while machine learning adapts its algorithm to the available data (Schabenberger, 2016).

The most common methodology to help with data mining is the CRISP-DM methodology. It is a cross-industry proven methodology, developed by IBM among others. It is an iterative process with six phases, (1) business understanding, (2) data understanding, (3) data preparation, (4) modeling, (5) evaluation and (6) deployment. It explains the relationships between these phases while providing an overview of the data mining lifecycle (IBM, 1996).

2.2.1 Log files

Since the main source of data at Philips is log data, a separate research using these as a data source is valuable. The logs contain all the software and sensor data, including results of all the maintenance activities ever done on a machine. It is possible to predict failures by checking daily logs for patterns consisting of multiple event codes. However, there are challenges as they; (1) rarely contain explicit information for failure prediction; (2) contain various kinds of heterogeneous data; and (3) contain massive amounts of data, composing computational challenges. (Sipos et al., 2014). To handle this, it is common to transform the problem into a classification problem (Last et al., 2011; Mosaddar & Shojaie, 2013; R. B. Patil et al., 2017; J. Wang, Li, Han, Sarkar, & Zhou, 2017). In addition, logs contain a large proportion of non-informative text that needs cleaning by removing unwanted spaces, numbers, punctuation and non-discriminating words. A text mining approach is possible to extract predictive indicators as well (Arif-Uz-Zaman, Cholette, Ma, & Karim, 2017).

2.2.2 Data analysis methods

Data analysis consists of comparing data with each other, like a certain sensor value over time, trying to find a linear trend. This would be a comparison of two continuous variables. It is also possible to do a comparison of categorical data with continuous data, e.g. the number of errors over time. These combinations have different potential and analysis methods. This paragraph focuses on inferential statistical methods instead of descriptive ones.

Analysis methods are ample, but only a handful fundamentally different algorithms exist, the following distinctions and descriptions are taken from (Provost, 2013):

1. *Classification and class probability estimation;* attempts to predict, for each individual in a population, which of a (small) set of classes this individual belongs too.
2. *Regression* ("value estimation"); attempts to estimate or predict, for each individual, the numerical value of some variable for that individual.
3. *Similarity matching;* attempts to identify similar individuals based on known data.
4. *Clustering;* attempts to group individuals in a population together by their similarity, but not driven by any specific purpose
5. *Co-occurrence grouping;* attempts to find associations between entities based on transactions involving them.
6. *Profiling;* attempts to characterize the behaviour of an individual, group, or population.
7. *Link prediction;* attempts to predict connections between data items, usually by suggesting that a link should exist, and possibly estimating the strength of the link.
8. *Data reduction*; attempts to take a large set of data and replace it with a smaller set of data that contains much of the important information in the larger set.
9. *Causal modeling;* attempts to help us understand what events or actions actually influence others.

What the best data analysis method is, depends on the available data from a fundamental and practical perspective. Therefore, the knowledge worker decides which method to use in respect to the environment of the data analysis and his own capabilities. This aspect is a prime example of why traditional activity-centric process modeling is not suitable for data analysis; it is a decision dependent on many aspects and should be more flexible in nature.

2.3 Decision intensive processes

Proper maintenance is dependent on domain knowledge, while the purpose of applying this knowledge is to reach the best possible decisions. Defining the specific procedures that individual knowledge workers or experts follow is difficult, maybe even impossible, since this procedure is flexible. A way of capturing processes in a flexible, conceptual way is using business artifacts. The complex decision-making, powered by versatile data is complex to model and analyze. Linking knowledge and conditional maintenance is attempted via Decision Intensive Processes (DIPs), DIPs are repeatable business processes whose conduct are dependent upon knowledge workers (Bromberg, 2007a).

Successful execution of a DIP depends on the right information reaching the right person, at the right time, in the right context. Just as for optimal preventive maintenance, where the optimal execution of doing an activity is at the right time, in the right context. DIPs must be structured to be agile, change, expand or scale to meet external pressures (Bromberg, 2007b).

DIP processes assist users in performing decision-intensive tasks and provide users with a guidance relevant to process execution context. DIP processes are by nature collaborative, data-driven, need to support various kinds of flexibility at the design and run time, and need to integrate with external services and information sources. DIPs are genuine knowledge, information and data-centric, require substantial flexibility at design, and runtime. In decision-intensive processes, a lot of knowledge and expertise is implicit, resides in experts knowledge or it has a form of (informal) best practices or organizational guidelines (Vaculín et al., 2011).

Therefore, to create a successful decision intensive process it is important to remember that:

1. DIPs are by nature collaborative
2. DIPs needs to be agile and support flexibility over various characteristics
3. DIPs are information and data-centric in an implicit way
4. DIPs are dependent on information reaching the right person
5. DIPs are powered and dependent on knowledge workers

2.3.1 Decision intensive processes and maintenance decisions

In the decision making process on which PM activities are required, there are three types of categories motivating the decision to do the PM activity (EN 13306:2010 standard, 2010).

1. *Predetermined maintenance*: Predetermined maintenance is the traditional planned maintenance technique. Measuring the total usage of an asset, conducting maintenance when reaching a certain threshold, often component age.
2. *Condition-based maintenance*: This maintenance strategy includes three major steps (Jardine, Lin, & Banjevic, 2006), data acquisition, data processing, and maintenance decision-making.
3. *Opportunistic maintenance*: an additional variant, when making fixed costs already, additional maintenance activities become worthwhile. For example, when there is a corrective maintenance call, the engineer might replace a different component since he is already there, while you would not replace it when you would have to send him there just for that activity.

The decision if a PM activity is done for condition-based maintenance normally happens when a value crossed a certain threshold. At one point in time, the expected costs for doing a PM activity preventively is lower than not doing the activity and risk corrective maintenance costs. Having and applying this information is the key difference with traditional maintenance.

Although, opportunistic maintenance implies that there are two thresholds for doing the PM activity since it does not require the fixed costs. Figure 6 – Decision-making for condition-based maintenance illustrates these thresholds, which are the deciding variables in the decision-making, partly derived from (Feng, Li, & Sun, 2012).

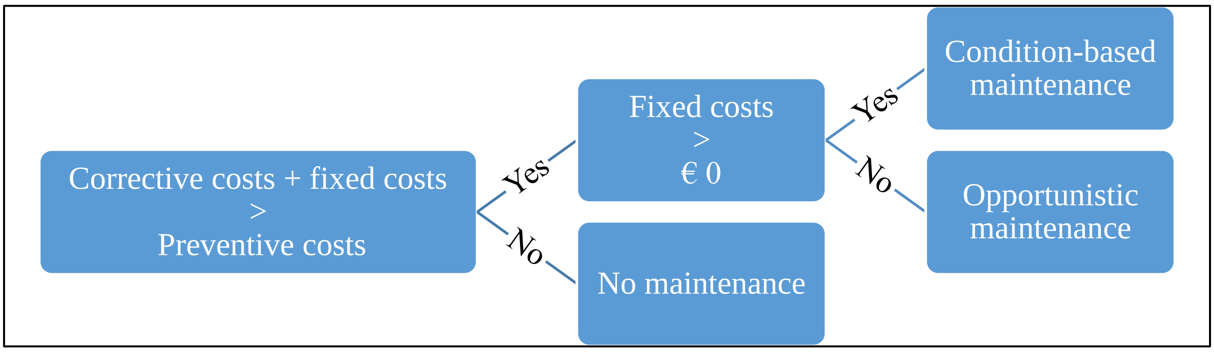

Figure 6 – Decision-making for condition-based maintenance

2.4 Business artifacts

The key challenge in business process modeling is to find mechanisms whereby business executives, analysts, and subject matter experts can specify, in an intuitive yet concise way, the framework and specifics of how the operations of a business are conducted (Hull, 2008).

The area of business artifacts (BA) is an approach to business process modeling centred on a holistic combination of data and process (Cohn & Hull, 2009; Hull, 2011; Limonad, Boaz, Hull, Vaculin, & Heath, 2012; Sun, Hull, & Vaculín, 2012). Perception of the role of data in business processes is subjective and hence varies considerably (Marrella et al., 2015). Artifact-centric business process modeling has emerged as it supports a more flexible process compared with traditional activity-centric models (Cohn & Hull, 2009; Köpke, 2016; Solomakhin, Montali, Tessaris, & De Masellis, 2013; Yongchareon, Liu, & Zhao, 2012). Therefore, making it more suitable for a DIP which is dependent, and their outcomes change, on the process information.

The two major paradigms of process modeling are activity-centric and artifact-centric. They focus on different first-class modeling constructs and therefore, they are eligible for different scenarios (Meyer & Weske, 2013). Where the most famous process model, Business Process Model and Notation (BPMN), is activity-centric.

Artifact-centric process modeling provides a higher level of robustness and flexibility for describing process specifications (Yongchareon, Liu, & Zhao, 2011). The artifact-centric approach has more potential than the activity-centric one for human-centric processes in which different courses of action are possible and the choices dependent on human decisions (Bruno, 2013) since they enhance rich and natural communication (Cohn & Hull, 2009).

The BA approach can be laid in a four-dimensional framework called BALSA, where the four dimensions are business artifacts, lifecycles, services and associations (Benatallah et al., 2015; Bruno, 2013; Hull, 2008). Based on BALSA, several artifact-centric meta-models have emerged in recent years. Although all the proposed models claim to support the artifact-centric approach, their support in the BALSA elements is not clearly described in the existing literature (Benatallah et al., 2015). One of the most known notations is the guard-stage milestone notation (GSM), which supports the specification of BALSA elements. The object management group (OMG) standard for Case Management Model and Notation (CMMN) draws key foundational elements from GSM, resulting in an effective tool as well (Koutsos & Vianu, 2017; Marrella et al., 2015; Russo, Mecella, Patrizi, & Montali, 2013; Sun et al., 2012).

2.5 Unifying paradigms with Decision intensive processes

Effective maintenance is dependent on knowledge, while the purpose of knowledge is to reach the best possible decisions. Defining the specific procedures that individual knowledge workers or experts follow is difficult since this process is flexible. A way of capturing processes in a flexible, conceptual way is using business artifacts. Modeling this complex data-driven decision-making is attempted via Decision Intensive Processes (DIPs), DIPs are repeatable business processes whose conduct are dependent upon knowledge workers (Bromberg, 2007a).

Successful execution of a DIP depends on the right information reaching the right person, at the right time, in the right context. Just as for optimal preventive maintenance, where the optimal execution of doing an activity is at the right time, in the right context. DIPs must be structured to be agile, change, expand or scale to meet external pressures (Bromberg, 2007b).

Decision-intensive processes are similar to case management; a user has much more control over the possible flow choices. Derived from the GSM model is the structure of the case management model CMMN. Case management specifically focuses on supporting the work of knowledge workers that is usually ad-hoc, centred around knowledge and relies on data and their flows (Vaculín et al., 2011).

Another business process modeling method that works well with DIP and CMMN is the decision model and notation (DMN). A standard published by the OMG group as well, which is an approach to model repeatable decisions. This group considered the combination of BPMN, CMMN and DMN the triple crown of process modeling, which together can model the range of working methods used across most organizations. OMG Group acknowledges that independent usage of BPMN, CMMN and DMN is possible, but while the designing them the focus was on creating complementary methods (OMG Group, 2016).

3 Methodology

Structural research methodologies help to analyze the problem and design towards a solution. This chapter defines the methodology used, illustrating how the researcher is planning to solve the problem mess. The problem consists out of two aspects, process-related and data related, which is a challenge in terms of a structural research methodology.

Therefore, two different research methods are used, the regulative cycle developed by van Strien (Aken & Andriessen, 2011) and the more data focused CRISP-DM method. The goal of the regulative cycle is to gain qualitative knowledge and insight, and support in the full project process.

Additionally, the researcher uses the CRISP-DM methodology to support in the analysis & diagnosis phase of the regulative cycle for the data analysis of existing PM activities. Combining these methods creates a hybrid data-driven approach by combining expert knowledge and data, which is deemed an effective method (Galar et al., 2015).

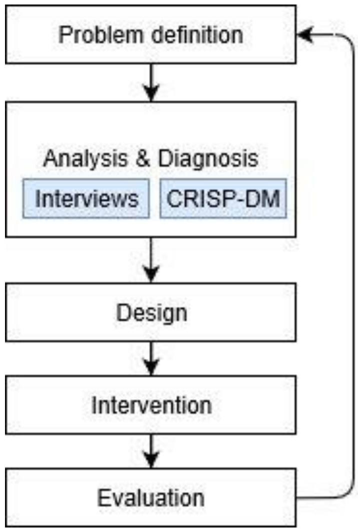

Figure 7 - Regulative cycle with CRISP-DM

3.1 Problem definition

The first phase of the regulative cycle is the problem definition, which is effectively chapter 1 of this report, defining the problem with a problem statement and multiple research questions (RQs) to progress towards a solution. This chapter describes the methods used to answer these RQs, and the method to resolve the problem.

3.2 Analysis & Diagnosis

The actual work starts with an analysis and diagnosis of the environment, the goal is to gain insight into existing PM activities and the current design process of new PM activities. This paragraph describes methods used for answering the RQs and therefore completing the analysis.

Firstly, gathering qualitative information happens with interviewing, described in detail in 3.2.1. The goal is to understand the motivations and types of PM, while also attempting to gain insight into the current process of defining PM activities.

Secondly, the CRISP-DM methodology is applied for the data analysis of existing PM activities, described in detail in 3.2.2. The goal of this analysis is to determine if answering hypothesis about effectivity and frequency of PM activities is feasible with the available data.

3.2.1 Interviews

The goal of the interviews is to understand the motivations and types of PM, while also attempting to gain insight into the current process of defining PM activities. Firstly, this chapter discusses the interviewees, then the structure and finally the type of analysis used. These interviews have the goal to answer the RQ1 *“how are PM activities currently designed?”*

The interviewed SMEs do not have a complete overview of the PM activities; however, they do know what is required for their product domain. Due to the highly diffused expert knowledge on PM activities, it is required to combine information from multiple interviewees. As a resource, product subject matter experts (SMEs) from service innovation (SI) and research & development (R&D) are used. These departments are relevant since R&D is responsible for the formulation of the PM activities while SI prescribes the PM activities to the markets.

The interview consists of two sections, firstly, a structured section that has a goal to describe all the existing motivations and types of PM activities. The options are pre-determined by analyzing the list of existing PM activities and then verified using the structured interviewing method because the expectation is that the question would be too open otherwise. The second section of the interview relates more towards the process of who and how they define PM activities, including the problems in the current process. This happens in a semi-structured way since it is unknown how these conversations will evolve during the interviewing.

After the interviews, they are transcribed and coded. The defined themes in the coding are for the first section the resulting characteristics and motivation. For the second section, it will require familiarization with the data before identifying relevant themes, since this section is semi-structured. In Figure 8 - Phases of Thematic Analysis illustrates this process, derived from (Gallardo-Echenique, 2015), there is an illustration of the applied phases of thematic analysis.

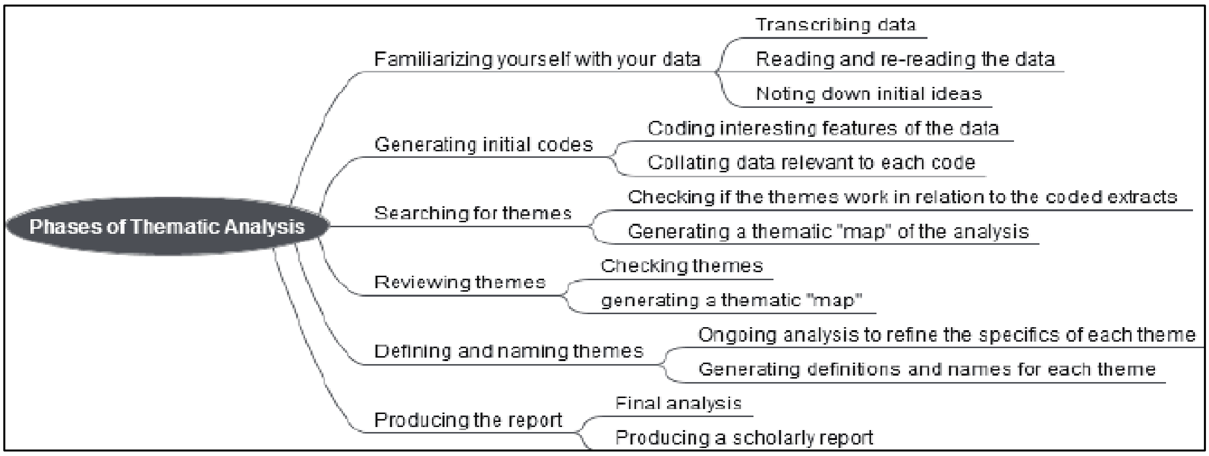

Figure 8 - Phases of Thematic Analysis

3.2.2 Data analysis

Proper data processing and analysis are key to the success of the project. A methodology often used in data-heavy projects is Crisp-DM, making it the methodology of choice for data-analysis of the PM activities. This chapter explains CRISP-DM and the goal of this phase.

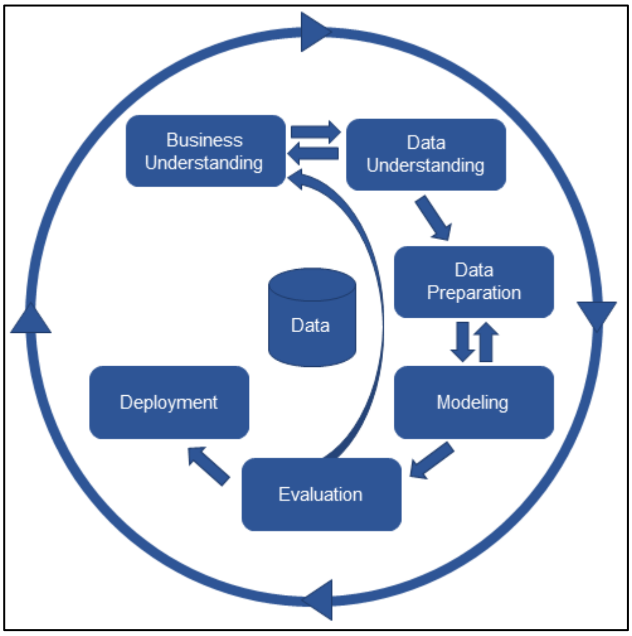

Figure 9 - Crisp-DM methodology in a process diagram

Crisp-DM exists of six phases, where moving back and forth between phases is required. In addition, the methodology has a cyclic nature since lessons learned during the process often give the potential for further improvement of the main problem, as illustrated in Figure 9. The phases consist of:

- *Business understanding*; understanding the project objectives from a business perspective and converting it to data problem, reviewing existing PM activities
- *Data understanding*; Collecting, familiarizing, quality analysis and other tasks to get a basic insight into the data to form the first hypotheses
- *Data preparation*; Changing raw data to a data set which is filtered and ready for use
- *Modeling*; Application and optimization of various data analyzation tools on the data
- *Evaluation*; Evaluate model quality thoroughly and check if all expectations formulated in the business understanding phase are met and includes all aspects
- *Deployment*; the scope of this research is limited to organizing and presenting the model results, creating the opportunity for SMEs to make deployment decisions

This phase has the goal to understand the potential of the available PM activity data. This is achieved by analyzing the opportunities and limitations of the historical data of PM activities. In addition, the analysis also looks into what other valuable data is available in both the Vertica data warehouse and other systems at Philips, answering RQ2 "*What PM activity data is available to analyze current planned maintenance activities?*"

3.3 Design

At this stage, there is knowledge about why Philips Healthcare has PM activities, and the potential of data analysis to review those activities. Now, it is required to combine these aspects to propose a new design for defining PM activities. In this design, two aspects are deliverables of this research, (1) the review of PM activities and (2) the methodological design of PM activities.

Knowing what data is available, does not complete the analysis since it is required to know how to apply the data to evaluate PM on effectiveness and frequency. The next step is to apply the data gained from RQ2 in the design, answering RQ3 *"How should the selected data be used to review decision making for PM activities"* creating the process of reviewing PM activities.

The review process of PM activities originates from doing multiple repetitions of analyzing different PM activities using the CRISP-DM methodology. Since the analysis happens in the specific environment of PM activities at Philips Healthcare, defining the recurring steps is feasible, resulting in a process specifically for data analysis of PM activities at Philips.

In terms of business artifacts, this means that the *key* artifacts become apparent, the artifacts which return for every analysis. The characterization of an artifact is a data object that is manipulated and therefore influences the process; with these iterations, the goal is to understand which manipulations have this influence. Understanding what variables are the key manipulators is required to understand the decision-making for each scenario. Because the essence of decision intensive processes is that although their conduct and execution are heavily dependent on knowledge workers, they are repeatable (Bromberg, 2007b).

Defining the tests to analyze requires a small explorative study up front. This explorative study looks into a few aspects of maintenance costs, derived from the service work orders, and performance of PM activities. This results in top 25 lists for (1) Defective parts costs, (2) Maintenance labour costs, (3) Quantity of replacements and (4) probability of error for all maintenance activities. In consultation with SMEs, it is possible to relate costs per part to groups of PM activities. Therefore, it became feasible to determine a subset of PM activities to analyze by using the combination of the probability of error of PM activities, and their cost impact. The goal of this subset is to do an analysis on effectiveness and frequency using CRISP-DM, applying the learnings and experiences from all the previous analysis into the following analysis, to enhance the methodology over time.

The data analysis is an implicit process, dependent on the employee doing the process. Therefore, the choice is to model the process in case management model notation (CMMN) since it is more flexible and data-driven. This is an iterative process; therefore, processing lessons learned from the previous analysis, considering them when starting a new data analysis.

With the deliverable of reviewing PM activities finalized, the second aspect is the methodological design of PM activities, answering the last remaining RQ4 with these deliverables, namely *"How should the process of designing planned maintenance activities look like?"* With the combination of a new methodology for designing PM activities, while optimizing the current and future PM with data analysis, improving the activities in terms of effectivity and frequency.

Defining the methodological design of PM activities happens with the knowledge gained from the interviews. By discussing the current situations with its drawbacks, and the wishes of the interviewees, the desired situation becomes apparent. The second aspect is more process related, namely the methodological design of PM activities. Learning from the interviews about how they do it now, makes it feasible to identify the root causes of the problems in this process. Then altering this process to mitigate root causes, by changing responsibilities and decision-making, creating a desired situation.

4 Analysis & Diagnosis

The goal of this phase is to obtain more insight into the status quo of Philips Healthcare, especially in the Planned Maintenance (PM) policy, to gain input for developing a solution for the main problem. The first paragraph explains the documentation of PM activities, the chapter continues with describing the process of the conducted interviews, including their results relating to the motivations and characteristics of PM activities, and the process of defining PM activities. Followed by a description of the process of reviewing existing PM activities using data analysis, completing answering RQ1 and RQ2 respectively. Additionally, there are descriptions of condition-based maintenance opportunities, identified during this phase.

4.1 Ingenia MRI-Scanners PM manual

For each product model, there is a PM manual, describing the required PM activities, including descriptions of actual execution and decision-making. The manual contains a timetable summarizing all the activities with their execution frequencies as well.

Table 3 - Sample Ingenia planned maintenance timetable shows the format of this timetable, including four example rows. The columns named “1”, “2”, “3”, and “4” are the PM visitation numbers in a range of two years, if a column has a value for a row, the activity is required. Therefore, if all columns have a value for one row, the activity happens four times in two years. The value in the column is the expected minutes the field service engineer needs for the activity.

Table 3 - Sample Ingenia planned maintenance timetable

Subsystem/ module	Task	Execution	Input	Spec	1	2	3	4
-	Log in to MR service	Workflow step	-	-	3	3	3	3
Magnet system	Check vent pipe	Mandatory (safety)	Pass/fail	Pass	-	15	-	15
Software	Clean up disks	Mandatory	Pass/fail	Pass	1	1	1	1
Measure & Adjustments	Check VSWR	Mandatory	Pass/fail	Pass	-	4	-	4

While the magnetic resonance service department in Best communicates the required PM visitations, the market groups of Philips Healthcare are responsible for the execution of the visitations. This is the reason why Philips Healthcare treats the condition-based maintenance activities they have separate from the PM activities since it is up to the markets to schedule when to execute the maintenance activity. If the markets want to combine condition-based with planned maintenance. Therefore, if it is feasible to create a condition-based maintenance activity, it will disappear from the PM manual.

4.2 Interviews for acquiring PM knowledge

Acquiring knowledge about planned maintenance activities in terms of their characteristics, motivations, and their design process happens via subject matter expert (SME) interviews. The assumption is that product SMEs have the most expertise in PM activities relevant to their component domain, compared to colleagues or other employees of Philips Healthcare.

It is worthwhile understanding the motivation per PM activity to gain a better understanding of the activities, explaining *why* Philips Healthcare is doing certain activities. These motivations give the possibility to do a more focused data-analysis or skip the analysis altogether when the activity is mandatory. For example, if an activity happens every half year because of legislation, it is of no use to do a data-analysis reevaluating the effectivity or frequency.

Characteristics are the returning aspects seen in PM activities. It answers the question of *what* Philips Healthcare is doing during PM. These characteristics give insight into what the goal of doing the activity is, making it possible to understand and interpret the data correctly. Therefore, it is a step in the business understanding in the CRISP-DM methodology. It makes it feasible to formulate the right hypothesis, and thus have a correct conclusion.

The interviews consist of two sections. Firstly, a structured section with a table that cross compares pre-defined characteristics and motivations. This challenges the SME to think, for each combination, if it is relevant for their component, validating the pre-defined characteristics and motivations while following up with questions if the SME was missing anything. This interview section happens in a structured manner since the expectation is that asking ‘why’ and ‘how’ PM is done, is a too open way of asking questions for practical results.

Secondly, there is a semi-structured interview section with a set of more open questions divided into content and process questions. The content questions relate back to the earlier presented table, to ensure the table is comprehensive, while the process questions relate to how defining PM activities happens. The goal of this section is to explore the current implicit and explicit decision processes.

4.2.1 Interview process

This paragraph describes the process of conducting the interviews, which consists of two aspects. Firstly, there is a structured cross-comparison of pre-defined PM activity characteristics and motivations for doing them. Secondly, there is a semi-structured section with open questions focusing on the process of defining PM activities. Appendix F contains the interview questions and encoded interview transcriptions, including the used codebook.

The creation of this cross-comparison table was in collaboration with a Service architect, by analyzing the current list of PM activities for the Ingenia systems. By comparing the finalized table to the PM activity list of a different MR system, the Achieva, validated the comprehensives of the table and did not result in any modifications.

Six interviews, with different SMEs, validate the categories and characteristics and investigate the current process of defining PM activities. Additionally, there was expert knowledge from a Service architect who accompanied every interview as well, to support with his domain knowledge asking relevant or critical questions from his background. Table 4 - Descriptive statistics interviews illustrates the basic features of the interviews.

Table 4 - Descriptive statistics interviews

#	Role	Department	Product	Duration	Words
1	Hardware designer	R&D	Gradient amplifier	45:51	5.399
2	System architect	R&D	General / Safety	40:55	6.437
3	Hardware engineer	R&D	Patient table	54:51	8.978
4	Product expert	R&D	Gradient cooling	56:24	5.923
5	Mechanical engineer	R&D	Mechanics	33:23	4.371
6	Product specialist	Service	Rf amplifier	56:40	9.737
				04:48:04	**39.845**

Since the first part of the interview was about verifying the characteristics and motivations with a pre-defined table, the interview process was therefore in a closed-interview nature. Therefore, the group of nodes in this phase directly relate to the pre-defined characteristics and motivations, making them the nodes in the thematic analysis. The second section of the interview was about the process of designing PM activities; therefore, the encapsulating node is "Process" that consists of the nodes: component design, effectivity PM, feedback, frequency, incentive and review activities.

Sometimes, during these interviews, discussions arose about potential opportunities for introducing condition-based maintenance. This is valuable information for any potential follow-up research, and therefore tracking was valuable as well, this happened using a separate node "Condition-based opportunity". Figure 10 - Thematic analysis illustrates the thematic analysis for the PM activity design process, including nodes, with their corresponding hierarchy.

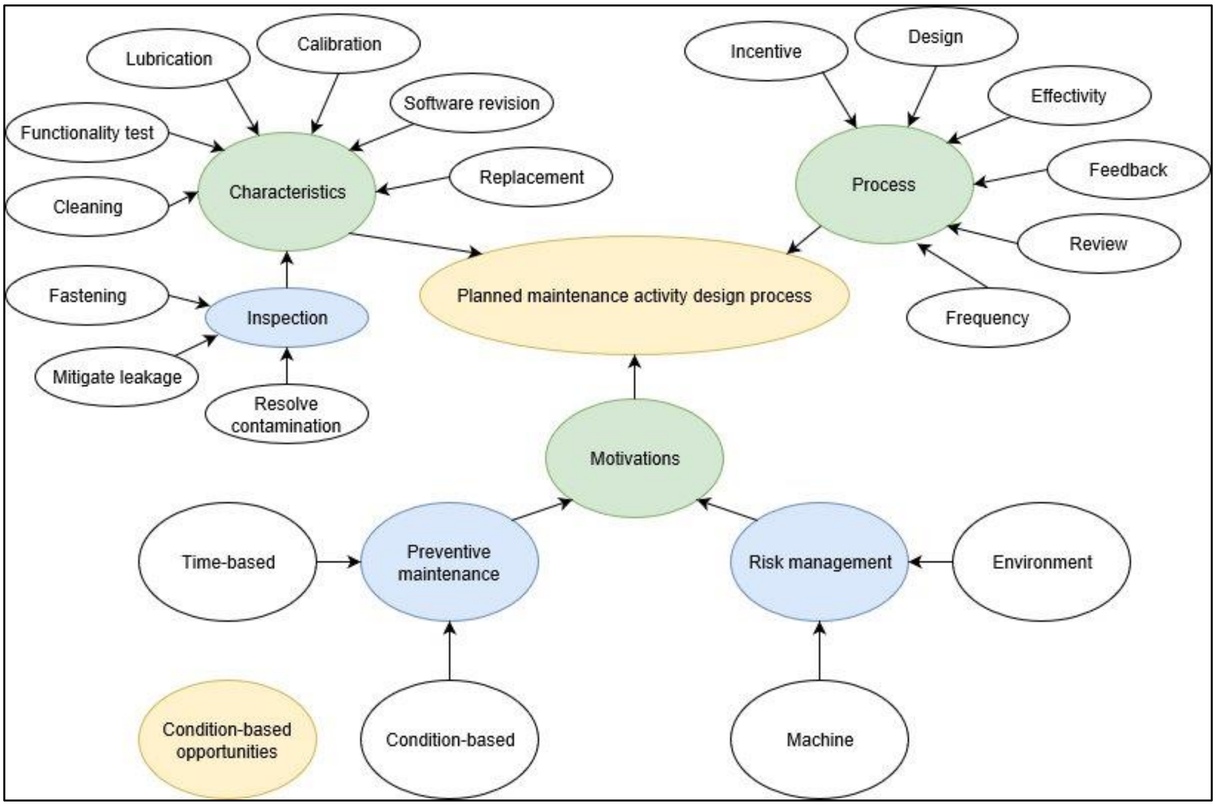

Figure 10 - Thematic analysis

4.3 Motivation categories for PM

There are many motivators for planned maintenance (PM) activities; this paragraph defines categories for these motivators in such a way that it is possible assigning all PM activities to a category. Literature defines the motivation for PM as a form of preventive maintenance solely in place to prevent corrective maintenance (Wood, 2003), there was no material found in literature about other motivators for PM. Therefore, the approach had to be more practical, defining the categories by analyzing the existing manual, in collaboration with a service architect, resulting in Figure 11 - Categories for PM motivators.

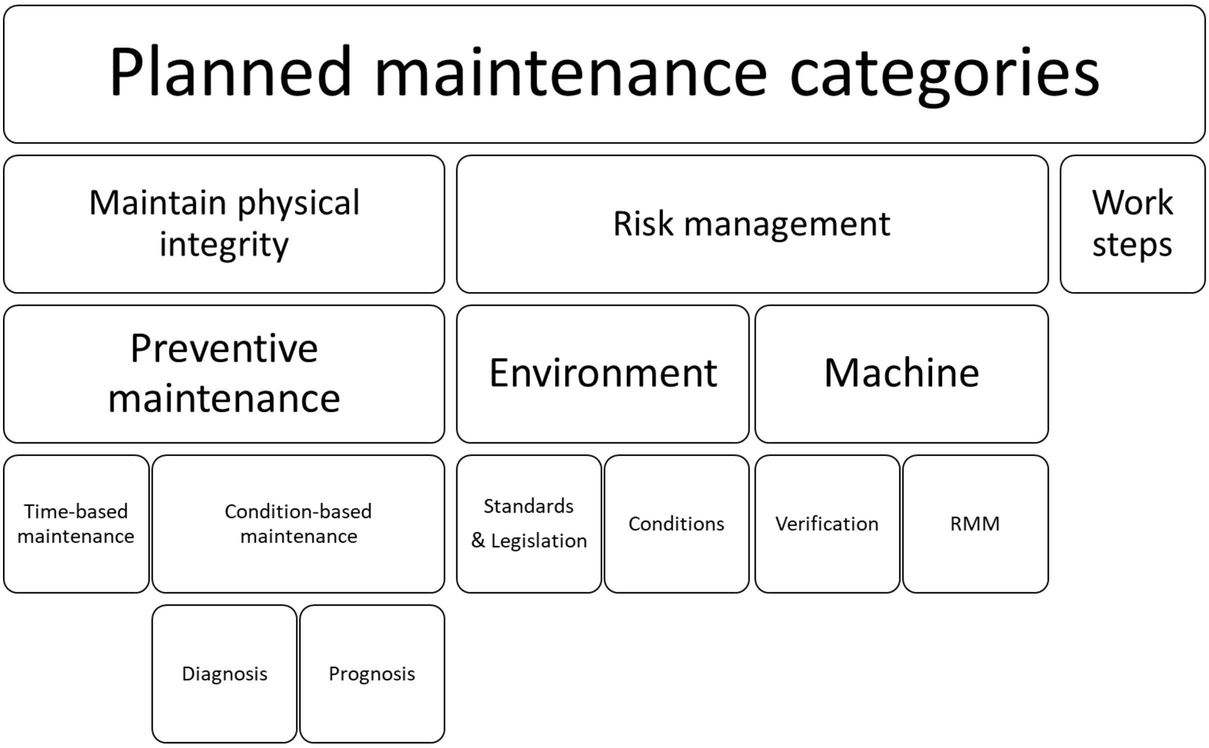

Figure 11 - Categories for PM motivators

The first layer of PM categories exists of three groups. Firstly, work steps, these are basic steps required to do the PM activities but have no direct benefit or value, for example removing a cover. The further distinction is between physical integrity and risk management.

Secondly, there is maintaining physical integrity. Currently, the scope is already on PM categories so it is possible to consider just preventive maintenance, otherwise, other aspects would be things like corrective maintenance. Preventive maintenance can both be time-based or condition-based, where condition-based maintenance can be on a diagnosis or prognosis basis (Niu & Pecht, 2009).

Risk management consists of two aspects and is in place to mitigate hazards. Firstly, there is the environment, which has to fulfill certain conditions, these are in order to meet the functionality guarantees of the machine, like the humidity and the temperature of the technical room. Additionally, in the environment, there are the standards and legislation, these are externally motivated and usually present for safety reasons, like checking outlet pipes.

In principle, doing preventive maintenance is always the choice of Philips Healthcare while for risk management this is not always the case, there may be external reasons like legislation or conforming to standards forcing Philips Healthcare to manage risks. When the latter is the case, it is not worthwhile to analyze the PM activity, since the frequency might be mandatory.

Machine specific risks are documented in the risk management matrix (RMM), which mitigates risks by ensuring functionality of certain machine aspects by an intervention, additionally, there is verification; this is a check if the machine is still functional. Validation of the pre-defined PM motivations resulted in only one small change, which was renaming work steps to workflow. For each sub-category, there were at least three but often more SMEs confirming the category.

With the categories defined, using expert knowledge from the Service architect, all the PM activities are appointed to their respective category. Resulting in an overview of spent time per category on PM, illustrated in Figure 12 - Spent time in 2 years of visitations per category.

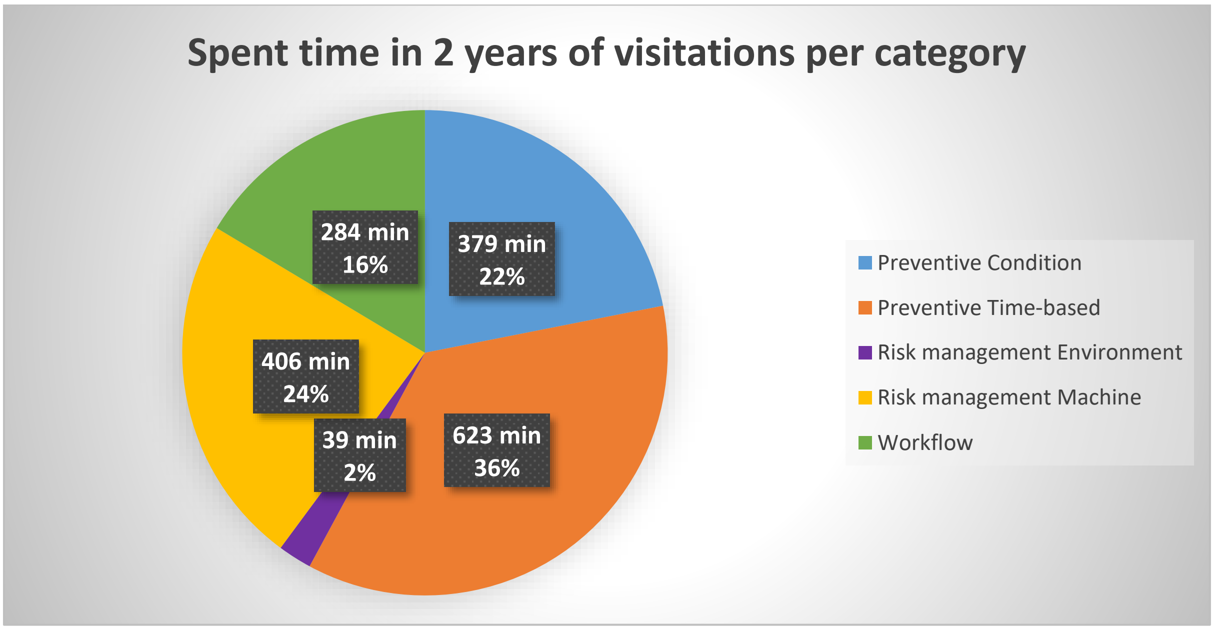

Figure 12 - Spent time in 2 years of visitations per category

4.4 Characteristics of PM activities

After analyzing different planned maintenance (PM) activities, it became apparent that they had certain characteristics to achieve their goal category. By dividing the activities per characteristic, it becomes apparent what the types of PM activities are. The primary results are illustrated in Table 5 - Characteristics of planned maintenance activities, which are created by analyzing the existing list of activities in collaboration with a Service architect.

Table 5 - Characteristics of planned maintenance activities

Characteristic	**Description**	**Example**
Functionality test	Test if the component is still working	Check door switch
Inspection	Test to see what current status is	Check site condition
Cleaning	Non-aesthetic cleaning	Computer cleaning
Filter replacement	Replacement of filters	Dust filter
Calibration	Measurement compared to standards	Power control loop calibration
Software revision	Activities for software	Specification files
Fastening	Mechanical joint checks	Bolts
Lubrication	Apply substance to minimize friction	Apply grease
Resolve contamination	Resolving impure liquids	Water Drain
Mitigate leakage	Solving leakage issues	Re-pressurize cooling loops
Workflow	Prerequisite to do PM activities	Remove covers

With the interviews, there was verification of the characteristics with SMEs to see if the list is comprehensive, the first learning from the interviews was that the characteristic "inspection" is more comprehensive, in the new definition the PM activities that 'check' for something happening that is not designed for, like leakage, are part of the inspection characteristic. As a result, the characteristics "fastening", "mitigate leakage", and "resolve contamination" changed to "inspection". Additionally, the characteristic "filter replacement" changed to a more general definition by changing it to "replacement". Table 6 - New characteristics of planned maintenance activities illustrate this review, the conclusion is that multiple SMEs confirmed that these characteristics should be comprehensive for all PM activities.

Table 6 - New characteristics of planned maintenance activities

Functionality test	*Calibration*	*Replacement*	*Lubrication*
Inspection	*Cleaning*	*Software revision*	

With the characteristics defined and expert knowledge from the Service architect, it is possible to appoint all the existing PM activities to their respective characteristic, resulting in an overview of spent time per category on PM, illustrated in Figure 13 - Spent time in 2 years of visitations per characteristic.

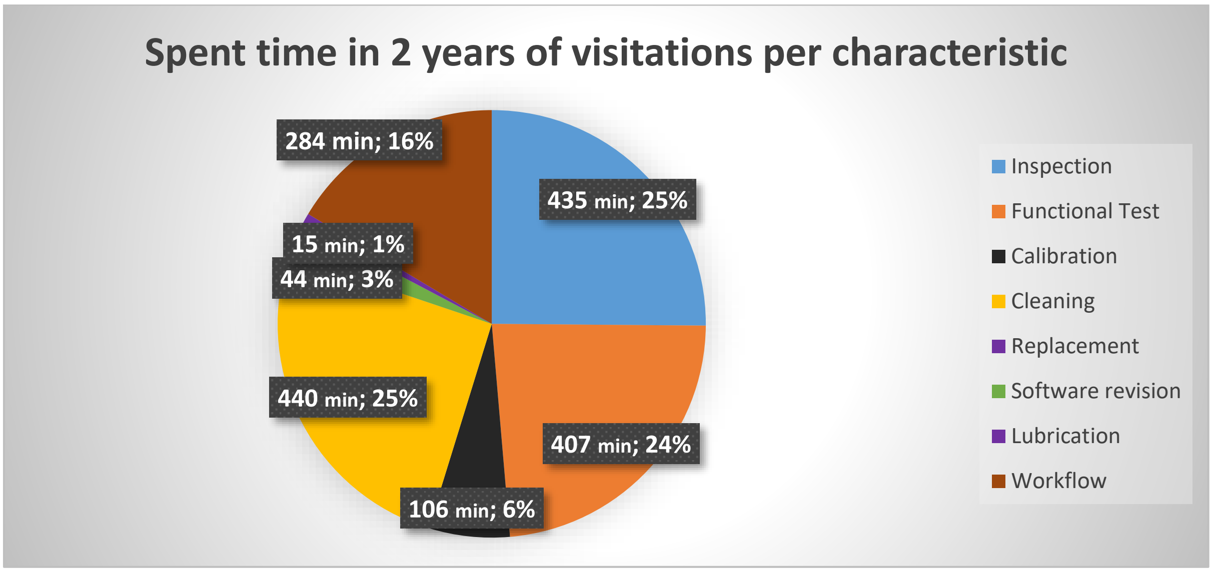

Figure 13 - Spent time in 2 years of visitations per characteristic

Table 7 illustrates the amount of time with motivations compared to the characteristics, providing insight into the relative importance of PM characteristics. The conclusion is that the most important characteristics are functional tests, inspections, calibrations and cleaning while replacement, software and lubrication are of less importance, making strategies to decrease the time spent on the first four characteristics of higher importance. Deriving these figures happens via the timetable defined in appendix D, defining activities to one motivation and characteristic, and summing the times for all four periods in the manual.

Table 7 - Spent time per motivation and characteristic per two years

	Preventive maintenance		**Risk management**	
	Time-based	*Condition-based*	*Machine*	*Environment*
Functional test	225 min	0 min	182 min	0 min
Inspection	25 min	191 min	180 min	39 min
Cleaning	252 min	188 min	0 min	0 min
Calibration	106 min	0 min	0 min	0 min
Replacement	0 min	0 min	0 min	0 min
Software revision	0 min	0 min	44 min	0 min
Lubrication	15 min	0 min	0 min	0 min
Workflow	284 min			

4.5 The process of defining planned maintenance activities

During the interviews, there was a recognition of six themes, namely (1) Component design, (2) Incentive, (3) Effectivity of PM, (4) Feedback between Research and Development (R&D) and Service Innovation (SI), (5) Reviewing of PM and (6) Frequency of PM. This paragraph discusses the learnings for each theme and their implication for the research.

R&D owns the process of defining planned maintenance (PM) activities since they create the component design, which is the origin of the required PM activities. In contrast, service innovation is responsible for creating a manual and prescribing of these PM activities defined by R&D. One of the issues is that R&D has limited insight into how and when PM actually happen, and it is unclear if R&D actually systematically prescribes the required PM activities.

Design changes have the most potential to decrease PM, and while it is a design rule to prevent PM, the fact is that PM still exists. This is because the foremost goal is component functionality, and preventing PM is more of a desire. Although decision-making about PM and other aspects happen in a design team, the choices made are from an R&D perspective instead of in a holistic way, while it depends per engineer how much effort there is in preventing PM.

The incentive for designing and reviewing PM is different between R&D and SI, the responsibility for the prescription of PM is at SI, they have to pay the costs related to PM, making their goal minimizing the costs of PM. In contrary to the R&D department, whose goal is to meet the reliability targets set internally independent of PM costs, they rather have more PM than less. Although the ambition is to reduce PM activities, according to a hardware engineer, the vision of individual system owners of R&D sustain this, and there is no explicit trade-off analysis between maintenance costs, component costs and other design characteristics. Therefore, these ambitions remain implicit and dependent on the vision of individuals.

SMEs question the effectivity of some PM activities, e.g. there are doubts about the effectivity of functional tests for components with random failures, and another example is inspections and calibrations that already happen remotely, undermining their effectiveness during PM.

The feedback loop between R&D and service innovation (SI) is weak and implicit, while they are dependent on input from SI, e.g. in terms of field performance and the process of PM. Therefore, R&D has no insight in the root causes of maintenance costs making it difficult for them to have effective and efficient decision-making in a holistic way in terms of PM activities.

There is a lack of knowledge about the existing PM activities, caused by the process of re-using existing PM lists, resulting in the fact that the domain of PM is no longer an integral part of personnel of SI. Currently, reviewing PM happens in a reactive way, caused by corrective maintenance (CM) issues in the field. The formulation of some existing PM activities happened a long time ago, while the most recent structural review has been years ago, thus structurally reviewing PM activities rarely happens. Additionally, it is undesirable to do a review for a single PM activity, since scheduling visitations happens a year beforehand, thus any changes in the visitations have to be significant in terms of costs to compensate the effort of rescheduling, requiring a disruptive change.

Currently, an incentive to review the current PM activities is high CM costs, but when there are no issues in the field in terms of CM costs, it is still possible that PM activities are optimizable. There is no structural process of reviewing for ineffective PM activities; interviewees stated that this is complicated since departments do not have sufficient insight into what the other departments reasoning is for having certain PM activities.

One learning in terms of frequency of PM is that the SMEs could motivate the time-frequency of doing PM for some activities in terms of safety. Outside of the legislation-defined frequencies, defining the motivation were terms as:

- Engineer judgement (System Architect, Product specialist)
- Based on the past / Rule of thumb (Product specialist)
- Gut feeling (Product specialist, Product expert)
- Knowledge of people (Service Architect, Product expert)
- When you are there any way (Hardware engineer)

To summarize, R&D designs components with the design rule to have no PM, but their main goal is to increase reliability and achieve functionality while minimizing bill of material (BOM) costs, so purely component parts. To do so, they would require input from SI in terms of field performance and maintenance costs, which is currently unstructured and insufficient. Because of the different goals to reduce PM and a lack of a holistic approach, R&D has a lower incentive to review the formulated PM activities, while SI does not have the component design knowledge to do so.

4.6 Reviewing existing planned maintenance activities

This chapter focuses on the reviewing of PM activities using data by discussing the relevant data available, and then, making a choice for a subset of PM activities, motivated by their direct and indirect costs. The next step is analyzing these PM activities in depth; this chapter will not focus on the detailed steps of these analyses, but rather on the results and the process.

4.6.1 Relevant data

Development of the internet of things (IoT) is also happening at Philips Healthcare, connecting their MRI-scanners with their servers exchanging data, which can be live data and historical data, like machine logs. The logging of this data has been going on for years, but it only became recent that this data became easily accessible with the introduction of Vertica, creating opportunities for monitoring, controlling, analysis and efficiency improvements from distance.

Vertica Systems is a column-oriented database management system used at Philips Healthcare, and it is the system containing most of the 'big data', where data can be extracted using SQL queries in the JDBC format using SQL clients like Squirrel, or, with additional libraries, R. Since Vertica is a shared resource Philips Healthcare wide, queries with low computational requirements are preferred. Therefore, for simple aggregating queries researchers use SQL clients, while for joining tables and other computational heavy analysis, R is preferred, since this uses the local computer resources instead of Vertica's resources.

Data is ample in Vertica, divided between over hundreds of different tables. This research makes use of the logging of every maintenance activity ever done for each system, including resulting parameters and results. Furthermore, Vertica also contains data like the number of scans per machine, machine configurations and their product models. Although, Vertica does not contain everything that is relevant; therefore, other data sources are still beneficial. An additional limitation is that some data tables do not contain the full history of a system, e.g., the number of scans per machine is only available since 2016.

Philips Healthcare also has the enterprise resource planning system SAP, this data is not directly accessible by the researcher but he can request extractions, the value of some of the data in SAP can be worth this extra effort. For example, it is possible to derive all PM visitation dates per system reference number (SRN) from SAP, making it possible to merge these datasets to Vertica tables and create a Boolean if the visitation was a scheduled PM visitation; although, registering this data only happens since 2016.

It differed per analysis what data was relevant to prove the hypothesis, although the PM activity data (STTdata), PM visitation data and setup registry (Product model) were always required, while always using the SRN as an identifier to merge the tables. The list below describes this and other data used to analyze the subset of PM activities, while Figure 14 - Entity-relation model illustrates their underlying relations.

- SRN; System reference number, unique number per machine
- TestId & ParId; Identification IDs in the software for each maintenance activity
- TestName & ParName; Human-readable names for each maintenance activity
- TestResult & ParResult; a discrete variable if the activity passed or something else
- ParDt & Date; the timestamp of the moment a certain test was done
- ValFloat & Specs; Float value that had to be Lesser than (Lt) or Greater than (Gt)
- DaysSinceFirstTest; Days between minimum ParDt and now for that SRN and ParId
- Normalized_Scans; Scans done by that SRN at ParDt normalized over log completeness
- PMvisitation; Boolean if the date is known to be a PM visitation for that SRN
- ProductModel; the product model of the SRN, filtered to always be an Ingenia type
- MagnetType; the magnet type of the SRN, filtered to always be Magnet_1 or Magnet_2
- RfType; the Rf type of the SRN, filtered to be Rf_1, Rf_2, Rf_3 or Rf_4

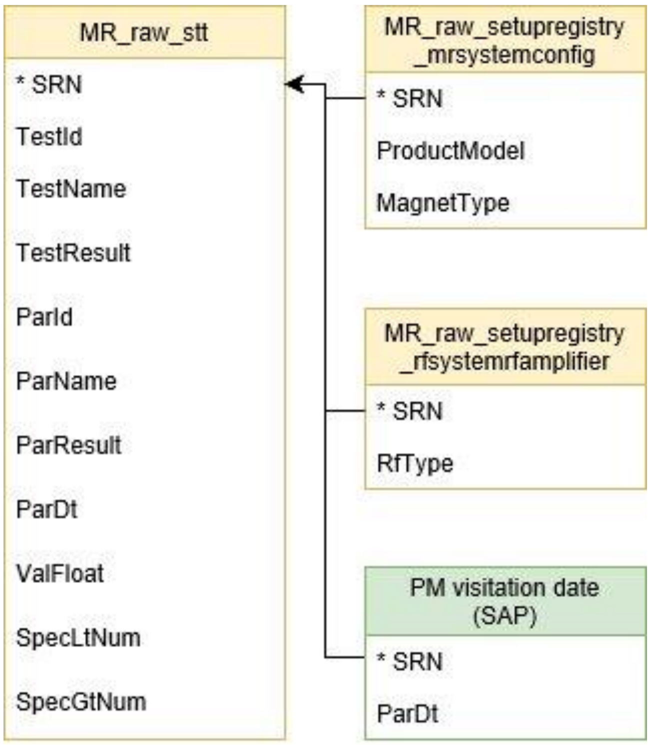

Figure 14 - Entity-relation model

4.6.2 A subset of planned maintenance activities

To define the subset of planned maintenance (PM) activities to analyze, there is an explorative research on the characteristics and motivations, plus the implicit costs per PM activity.

Service work orders (SWO) contain the data about corrective maintenance, including part costs and the spent time of the field service engineer (FSE). This data requires a manual process to retrieve, so is not ideal for continuous monitoring but suitable for periodic analysis of CM costs. This data is the input for deciding which components, hence PM activities, are interesting to analyze by comparing the CM costs, and required time for the FSE per PM activity.

In terms of motivation, the main categories defined are preventive maintenance and risk management. In principle, doing preventive maintenance is always the choice of Philips Healthcare while for risk management this is not always the case, there may be external reasons like legislation or conforming to standards forcing Philips Healthcare to manage certain risks. When the latter is the case, it is not worthwhile to analyze the PM activity, since the frequency might be mandatory.

The second categorization happens over the different kind of characteristics a PM activity can have, in total there is the definition of seven characteristics. The functional tests, inspections and calibrations have data available in the test data table, while the four characteristics that lack data in the Vertica table are replacement, software revision, lubrication and cleaning.

Costs for PM activities primarily consist of the labour cost for field service engineers (FSEs), making the number of activities and their frequency the highest influencers of the PM costs. To endorse having a PM activity it should be cheaper than the alternative of having CM costs, thus the effectivity and frequency of a PM activity are of great importance. Therefore, it is logical to start analyzing the PM activities that have the highest financial impact, selecting these PM activities by

1. Relating them to components with the highest corrective maintenance costs
2. Taking the ones with the highest error rates
3. Map which of them require the most time of field service engineers

The analysis on these aspects happens in detail in Appendix B. In terms of corrective maintenance (CM) costs, there are pivot charts illustrating corrective costs in terms of quantity of replacements, component costs and component plus employee costs for Ingenia systems. The conclusion is that the Rf amplifier, including its subcomponents, is the most expensive module in terms of CM. Other expensive components are the power module, compressor and gradient coil, where especially gradient coils are expensive to replace in terms of component costs, although the quantity of replacements is relatively low.

In terms of error rates, there is a table illustrating the error rates per PM activity, while these results are harder to interpret because they do not consist solely of PM activities, one of the important findings was the high error rate for some of the Rf amplifier related tests.

Finally, there is a table illustrating the amount of maintenance time spent per activity, and the frequency of the activity, resulting in an average time spent per year. Interesting observations were the high amount of time spent on verification activities and cleaning activities.

Considering the corrective maintenance costs, error rates and maintenance times. Table 8 - Analyzed planned maintenance activities illustrates the chosen tests to analyze the category preventive maintenance on a time basis. This category is a prerequisite because for risk management it is unsure if the activity is mandatory from a legislation or adhering a standard. There is a limitation on functional tests and calibrations since cleaning does not have data available while not considering the characteristic inspections due to time constraints of the research.

Table 8 - Analyzed planned maintenance activities

PM Test	**Motivation**
Calibration 1	High corrective maintenance costs
Calibration 2	High corrective maintenance costs
Calibration 3	High error rate
Calibration 4	High error rate
Calibration 5	High corrective maintenance costs
Functional test 1	High error rate
Functional test 2	High error rate
Functional test 3	High error rate
Functional test 4	Indirect corrective maintenance costs

4.6.3 Learnings CRISP-DM

The first time an analysis was started on the "*Calibration time-based 1*" activity, the CRISP-DM methodology gave the first structure in the way of working. When doing the analysis for following PM activities, the learnings from the previous analysis were implemented, creating a consistency in the way of working. Therefore, this chapter describes all the learnings per CRISP-DM step. Creating input for designing a new methodology and thus essential input for answering RQ3, specific on how to do data analysis for PM activities at Philips Healthcare, modelled in Case Management Model and Notation. CRISP-DM consists of the phases, (1) Business understanding, (2) Data understanding, (3) Data Preparation, (4) Data modeling, (5) Evaluation and (6) Deployment, which will be the following subparagraphs.

4.6.3.1 Business understanding

The data analysis of PM activities at Philips Healthcare has two major aspects. Firstly, the entry criterions, define when there is a business desire to do the actual analysis. Currently, three distinct entry criterions are recognized, components which cause high corrective maintenance (CM) costs, tests with high error rates and components which require a lot of PM time. In the case of high CM costs, it is required to define the relevant existing PM activities including its test and parameter IDs, done by consulting the respective subject matter expert (SME). While in case of the tests with high error rates or PM time, the support in defining the relevant identification codes is not required.

Secondly, it is important and a challenge to define the goal of the analysis comprehensively, to ensure that there is no time spillage by doing a non-beneficial analysis. Therefore, it is important to involve key stakeholders to ensure that everyone is on the same page, so that they all agree on the business objectives, understand the current situation and agree with the goals of the data analysis, resulting in business goals and a hypothesis encapsulating all these aspects.

One of the most important learnings was that the goal of the data analysis is always to give a recommendation about the effectivity and frequency of the test during PM. The analysis should motivate the actual existence of the PM activity, and if the activity is required, in what frequency then, illustrated in Figure 15 - Hypothesis format data analysis PM activities.

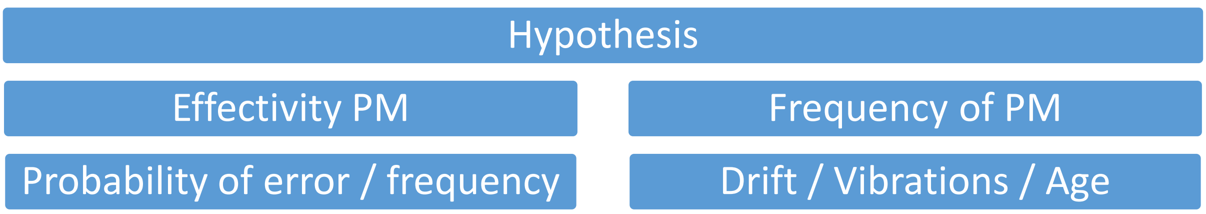

Figure 15 - Hypothesis format data analysis PM activities

4.6.3.2 Data understanding

With the business understanding and goals agreed on, the data analyst becomes responsible for achieving these goals, firstly by selecting the relevant data to analyze like the test results, machine configurations or other parameters, deciding with SMEs what and how the variables should be included in the analysis. This data is not limited to Vertica, firstly knowledge from other SMEs regarding the test can be valuable to include or exclude certain parameters or categories beforehand. Secondly, it is possible to use other databases, like SAP, for PM visitation data or Service work orders (SWO) for component replacement.

In terms of the test data, the test result is always a discrete variable, directly dependent on the parameter results, which are a passed or failed Boolean. Optionally, the parameter results are dependent on a float value, which has to be in a defined specification range. Figure 16 – Test result dependency illustrates the interdependency of these variables.

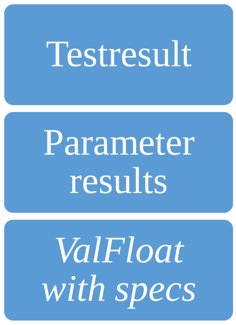

Figure 16 – Test result dependency

The data tables from Vertica have ample columns of information making it important to consider carefully which attributes are relevant for testing the hypothesis, after defining all the relevant attributes, it is important to reevaluate if proving or disproving the hypothesis using these attributes is possible. If not, it is needed to consider what additional data is required to test the hypothesis before going further. The minimum required data is the

- Full test data history
- Machine configurations, like the Magnet type and Product model
- Planned maintenance visitation data

When querying the Vertica database, it logically acquires all the information up to that day, but when acquiring data from a different source like SAP, it is important using only Vertica data up until this date when repeating the analysis to ensure merging data with the same timeframe. Additionally, when using different sources like this, it is essential to convert them to the same format before merging to prevent quality issues. For example, a timestamp in Vertica has a different format than in SAP, when comparing for equality this will return a false if not converted to the same format.

4.6.3.3 Data preparation

Processing of all the selected data entails the data preparation. With the selected data available, the different sources need to be connected. Every system has a unique identifier code used in most Vertica tables and SAP tables, namely the SRN. Therefore, the recommendation is to combine the data by this SRN code.

Filtering is the activity of removing data that meets certain conditions, which is possible at different moments in the process; the only requirement is that filtering happens before the modeling phase. Therefore, filtering in the SQL query already, just before feeding an analysis model or both are perfectly fine, although the recommendation is to keep track of the number of data points before and after filtering conditions, which is more complicated when filtering is in the query. It is important to estimate if the sample size is large enough to have conclusions and recommendations for the whole population, listed below are some examples of filters.

- *System age*; excluding values of new machines without wear and tear
- *TestId or ParId*; Only look at a specific test or parameter, identifiable by this column
- *Days since the first test*; excluding values lower than a threshold to exclude system installations or other startup moments
- *Log completeness*; when using daily log information, check for the number of available logs compared to system age, filter low log completeness (< 80%)
- *NA values*; filter rows where data is missing, like machines with an unknown configuration in terms of magnet type, product model or Rf amplifier type
- *Duplicates*; remove duplicate rows, possibly caused by joining tables
- *System configurations*; only consider systems which meet certain configurations, e.g. the product model has to be an Ingenia
- *Systems with configuration changes*; remove systems from the data who have changed in a configuration in the known history
- *Number of data points*; remove systems with too few data points, especially important for analysis on a system level, e.g. regression analysis
- *SRN range*; some SRN values are known to be test systems; therefore it is common practice to filter out SRN < 100 or SRN > 97000
- *Infeasible combinations*; the data might not be perfect; therefore some SRN configurations can be filtered by checking for data combinations which are not possible. For example, a magnet type with an incompatible RF amplifier type

Outliers and inconsistencies can be the result of faulty data, but this does not have to be the case. Therefore, it requires consultation with SMEs to explain outliers and inconsistencies, leaving them out of the analysis if they are explainable. Careful documentation is required for any removal or recognition of outliers or inconsistencies in the analysis, considering their effect for any conclusions or recommendations, indicating any assumptions made. The possible negative effects of outlier removal, like decreasing confidence in the conclusion, is smaller when the percentage of outliers compared to the dataset is smaller.

Granularity is the detail level of the analysis in terms of the timestamp, e.g. on a precision of seconds or days. When a system has multiple errors on one day, without intervention a model would start analyzing all these data points separately. In these situations, it can be interesting to increase the granularity if the goal is to discover trends over larger timeframes or when calculating error percentages. For example, analysis where machines have ten float values per day can improve by taking their mean value per day when considering large timeframes.

Before modeling, some data construction by creating additional analysis variables can be valuable for the model. Common constructor variables used are:

- *Timestamp truncated*; Multiple columns with a different granularity of the timestamp
- *Days Since First Test*; the number of days since the first test expressed as an integer
- *PM Boolean*; if the visitation is a PM activity TRUE or FALSE
- *Number of scans*; the number of scans made since the first scan and the test time

The final step before statistical modeling is converting the data to a consistent format, where the goal is to make the formats as easy as possible to interpret by the model and make them consistent so they are interchangeable between datasets. Examples are having timestamps in the same formats, scale extremely large or small values, define all numeric values as the same type (e.g. doubles) and categories as factors (in R).

4.6.3.4 Modeling

To start modeling there is a distinction between descriptive statistics and inferential statistics. The goal of descriptive statistics is to gain the first insights of the data, by organizing, summarizing and simplifying the data to relevant figures, by exploring the data with general figures and graphical illustrations. While inferential statistics are more descriptive in nature, their goal is to draw conclusions and predictions about the dataset based on a sample set.

The type of *descriptive statistics* are data dependent, Table 9 - Types of descriptive statistics describes some common mathematical figures per datatype, for the test data this usually results in analyzing ParResult and ValFloat. Often, to gain first insight in the grouping of the data, analyzing happens over different categories in the sample, e.g. product model. Further analysis on relationships among variables go into the inferential statistics section, descriptive statistics give a first insight into what inferential statistics might be interesting.

Table 9 - Types of descriptive statistics

Categorical *(ParResult)*	**Continuous** *(Valfloat)*
• Distinct values (e.g. # of SRN) • Frequency (e.g. # of Data points) • Percentage of total (e.g. Error rate)	• Mean, Median, Standard deviation • Range (min, max) • The interquartile range (LQ, UQ)

Just the general figures do not represent all the data characteristics, for further data exploration and interpretation, *graphical illustrations* are an excellent tool, where Table 10 - Types of descriptive graphical illustrations describes some common general figures per datatype. For example, they are effective to understand if the data has a parametric distribution, for the choice of inferential statistics to use; methods are the QQ-Plot, a fitted line over a histogram, and a Kolmogorov-Smirnov test to quantify the goodness-of-fit.

Table 10 - Types of descriptive graphical illustrations

Categorical *(ParResult)*	**Continuous** *(Valfloat)*
• (Clustered) Bar Chart (with error bars) • Pie chart • Dot plot (Comparing category to continuous) • Radar (Comparing category to continuous)	• Histogram (with fitted line) • Box & Whisker Plot • Scatter plot (with line) • Q-Q plot (parametric fit) • P-P plot (skewness)

By comparing the hypothesis with the gained knowledge from the descriptive statistics, it is possible to define the desired statistical tests, the so-called *inferential statistics*. These evaluate the remaining sample with the goal of accepting or rejecting the hypothesis. An important aspect is if the descriptive statistics show a parametric pattern in the data, if so, these statistical methods are preferred over non-parametric methods. Table 11 summarizes some of the most commonly used statistical methods for certain scenarios. For example, if you want to predict the continuous Valfloat value using the continuous Days since first test parameter, but the values do not show a parametric distribution, a kernel test is a suitable option for data analysis.

Table 11 - Common statistical methods (Common Statistical Tests, 2018; Zulfigar & Bhaskar, 2016)

Scenario	**Dependent variable**	**Independent variable**	**Parametric**	**Non-parametric**
Mean of two independent groups	Continuous /scale	Categorical/ nominal	Independent t-test	Mann–Whitney U test
Mean of 2 paired samples	Continuous /scale	Time variable (time before & after, t0 & t1)	Paired t-test	Wilcoxon signed-rank test
Mean of 3+ independent groups	Continuous /scale	Categorical/ nominal	One-way ANOVA	Kruskal–Wallis
3+ measurements on the same subject	Continuous /scale	Time variable	Repeated measures ANOVA	Friedman Test
Relationship between 2 continuous variables	Continuous /scale	Continuous /scale	Pearson's correlation coefficient	Spearman's correlation coefficient
Predicting the value of one variable from other	Continuous /scale	any	Linear regression	Kernel test
Survival distribution	Time to failure	Any, often time	Weibull life	Kaplan-Meier
Relationship between two categorical variables	Categorical/ nominal	Categorical/ nominal	-	Chi-squared test

Especially for the tests that compare continuous variables, analysis per category or even SRN are valuable. For example, a linear regression over the entire population of ValFloat and DaysSinceFirstTest may not return any results, while there are trends on a system level, thus doing the analysis per system reference number can give new insights.

4.6.3.5 Evaluation

It is common to gain new learnings during the data preparation process or modeling process in the (intermediate) results, requiring documentation and resolving the issue, while restarting the process before evaluating the conclusion. Only when the researcher is confident the results are comprehensive and statistically sound, it is of value to evaluate the results with the SME. One method is by creating a presentation with all the relevant documentation, the process becomes more efficient, where such a presentation should contain at least the following aspects:

- The Hypothesis which are to be tested
  - *Effectivity*: The probability that the PM activity passes during
  - *Frequency*: The probability that the PM activity passes consistently during a timeframe
- Relevant variables in the analysis (e.g. categories)
- Approach & Scope used to test the hypothesis
- Filters used on the dataset including;
  - Reasoning if the sample is representative
  - Descriptive statistics on the effects of the filters on the dataset
- Descriptive statistics including;
  - *General*; Parameter ID's, Data points, unique SRN, Error %, mean system ages, test frequency, test frequency according to PM,
  - *For Valfloat*; minimum, maximum, mean, standard deviation
  - *Data visualizations*; including a parametric test when working with Valfloat
- Inferential statistics including;
  - Tests to prove or disprove the statistical significance of categories and PM
  - Trend analysis when working with Valfloat values
  - Survival distribution to predict the probability of error over the timeframe
- Conclusion with a summary of the most important results and
- A final conclusion if the hypothesis is valid, rejected or if this is undecided

When the subject matter expert disapproves the conclusion, the comments have to be processed and the data analyst restarts the process. It is common that this happens multiple times. When both the data analyst and the subject matter expert agree on a conclusion, the last responsibility of the data analyst is to document the analysis comprehensively including his and the SMEs reasoning.

4.6.3.6 Deployment

The actual implementation of the results of data analysis is dependent on other factors as well; there might be implicit motivators not apparent from data analysis. From this phase, using the data analysis in any business decision making, the SME is responsible again. The initial goal is to reconsider doing the PM at all and the frequency. If the activity is considered effective for a given frequency, the SME will have to discuss the results with other SMEs, to see if the recommendations from the data-analysis are implementable.

The analysis returns probabilities in terms of PM activities of passing, optionally in a timeframe, while also knowing the required time to do these PM activities. The decision in keeping the PM activity depends on the height of the expected CM costs when not doing the PM activity, compared to the height of the known PM costs, when the PM costs are higher than the expected resulting CM maintenance costs; it proves that the PM activity is ineffective, thus redundant.

4.6.4 Key business artifacts

Business artifacts are data objects, whose manipulation define the process (Koutsos & Vianu, 2017). In this process, a *single* key artifact determines the course of the entire process, the maintenance activity. Its underlying aspects determine the choices made in the process, including how it ends. The key aspects of the business artifact consist of the following

- To what components the activity is related too, to define if the analysis is valuable
- For what motivations the activity is done, to define if the analysis is valuable
- What the characteristics the activity has, to define if and what analysis is valuable
- If the activity has enough data points, to make the dataset comprehensive enough
    - All the relevant data as specified in the entity-relation table
        - Where the deciding variables are the result and optionally float values
    - If the activity is Planned maintenance or not
    - If the data is of high quality or not (fit for use)
- If the data points relate to a hypothesis about effectivity and frequency

These are the identified critical aspects of the business artifact, if these are equal for two different maintenance activities; the conclusion in the process in terms of the hypothesis is equal. Effectively, resulting in a decision-intensive process because dependent on the data and this process, there is a repeatable decision.

4.7 Condition-based opportunities

During the interviews with the SMEs, multiple opportunities for CBM became apparent, summarized in Table 12 - Condition-based maintenance opportunities expert interviews.

Table 12 - Condition-based maintenance opportunities expert interviews

Role	Opportunity	Data required	Time
Hardware designer	Predict remaining useful life	Usage patterns in terms of load and number of cycles	09:03
System architect	Still vague	Adding sensors to ‘weak spots’	33:35
Hardware engineer	Replace fan before fail	The current speed of the fan	20:15
Product expert	Contamination measurement	Measurement of PH values	22:27
Product specialist	Power reference calibration	Indirect power drive variables	43:06

Additionally, during the data analysis other opportunities CBM became apparent as well, summarized in Table 13 - Condition-based maintenance opportunities data analysis.

Table 13 - Condition-based maintenance opportunities data analysis

PM activity	Opportunity	Data required
PSU self-test	Use self-test to predict degradation	Daily Fan data
F0 determination	Remote CBM already happening	Daily F0 data
Cleaning	Use temperature to estimate the amount of dust	Temperature of CPU

An important finding is that all of these opportunities require a constant flow of data, independent of a (planned) maintenance activity, where often a dataset of historical data is required to create a condition-based maintenance algorithm.

4.7.1 Condition-based planned maintenance

One of the methods recognized in the literature study to introduce data-driven planned maintenance (PM) was by using the concept of condition-based maintenance. Although, condition-based maintenance is not an effective solution with PM activity data instead of remote data for a multitude of reasons:

- *Data point usability:* For many tests, there are quite many data points outside of the PM activities, for example from corrective maintenance. This data is inherently different, unusable for the data analysis of PM activities. Using statistical methods like the Student t-test, ANOVA test, or histogram comparison, proves the assumption that these data points are indeed different in nature.
- *Data point origin:* Therefore, only PM data points are always usable. Collecting these data points happens during a PM visitation. When changing to a condition-based strategy, there is no longer the creation of new data points. Therefore, condition-based maintenance is only possible from data generated independently from PM activities. These relations are hard to predict from a PM perspective, it is easier with the goal behind the PM activity in mind, according to experts at Philips Healthcare.
- *Age of available data:* Logging test data is happening for years, before the introduction of Ingenia systems. Although, not all the data is this complete, like a limitation in the history, two major constraining variables are;
  - *Planned maintenance visitation:* are only known since 2016
  - *Usage data:* like the number of scans, also only known since 2016
- *The number of data points:* Frequency of PM activities happens at least 0.5 times per year up to 2 times per year. Therefore, the maximum PM data points of a unique SRN is the time since 2016, multiplied by the frequency minus one. That means in time of writing (Q4 2018), this is a maximum of 5 to 6 data points.

Therefore, learned from the interviews and data analysis, condition-based opportunities are only replacement maintenance activities, which are then controlled by remote data independent of maintenance activities or other irregular interventions.

5 Design

With the knowledge from the interviews, and experience of the data analysis, the next step is the design phase. This consists out two deliverables (1) the review of planned maintenance (PM) activities related to RQ3 and (2) design of PM activities related to RQ4 while combining these results in solving the problem via the desired process of defining PM activities.

5.1 Review of planned maintenance activities

The practical learnings from the CRISP-DM in paragraph 4.6.3, give the possibility to design a new, more focused, model based on CRISP-DM, as illustrated in Figure 17 - Data analysis PM activities. This model consists out of the main stages of CRISP-DM, but each stage has additional details relevant for PM activities at Philips Healthcare, additionally, there is a definition of the entry and exit criterions as well. This paragraph describes the motivation for each attribute defined in this model, including the practical implications for the analysis. The eventual goal of this process is to review a PM activity in a structural, reproducible way.

The methodology is large and complex, and therefore not ideal with a hands-on approach by a data scientist. Therefore, a manual is developed describing the exact activities the researcher has to do to achieve structurally reviewing PM activities. The actual manual is described in Appendix C, and paragraph 5.1.2 describes the application of this manual.

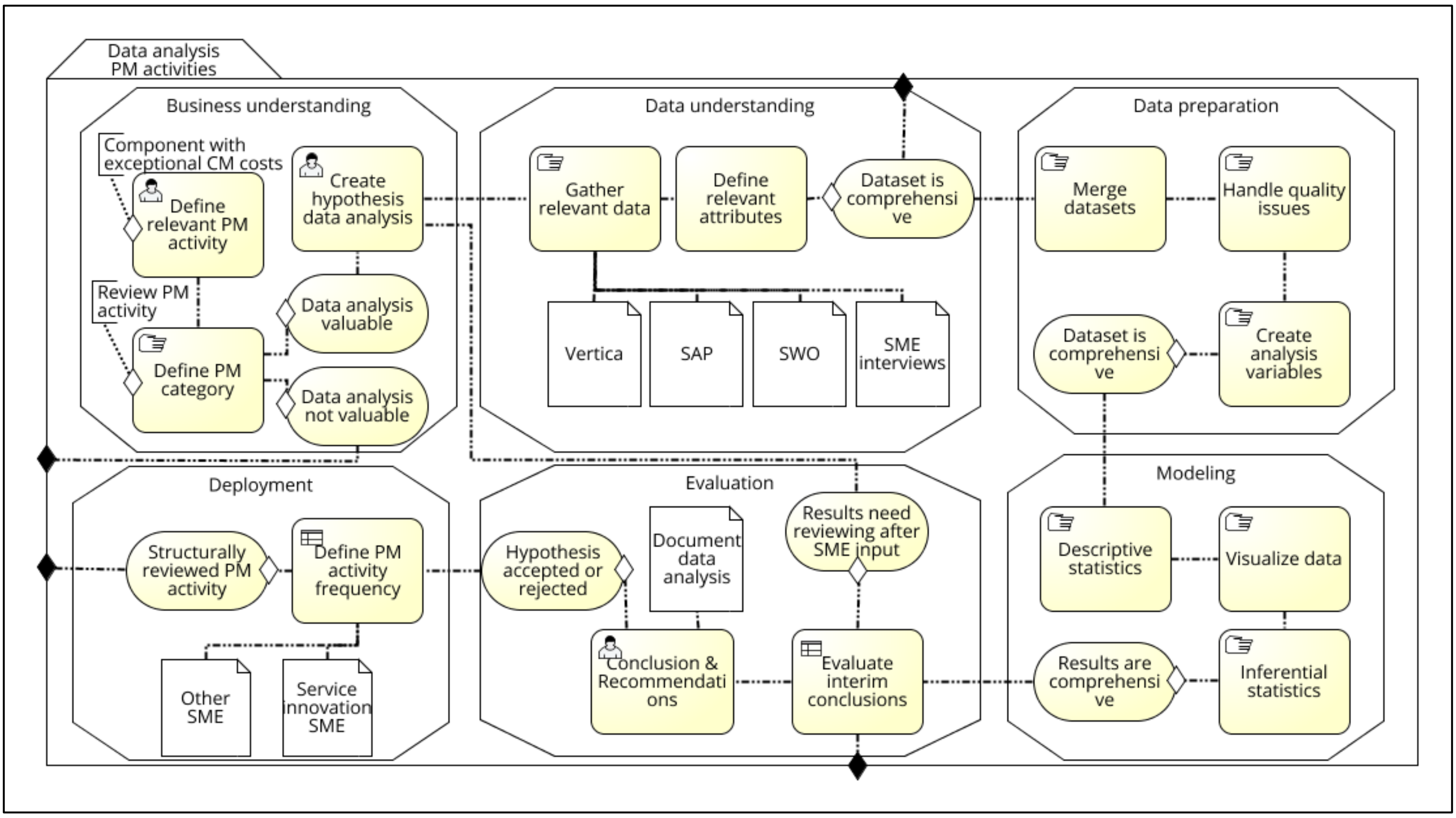

Figure 17 - Data analysis PM activities

The process starts with an entry criterion, triggering the review, initiated by (1) a component with significant CM costs triggering an analysis or (2) the choice of reviewing a PM activity. If the entry criterion (1) triggers the model, it requires the transformation from a component to *relevant PM activities*, thus which activities relate to that component.

Every PM activity has a *category* and *characteristic*. The combination of these business artifact aspects decides if the milestone *data analysis is valuable* is achieved or not, an analysis might not be valuable if (a) the category has a mandatory nature, as legislation or (b) the characteristic has no relevant activity data. In principle, all the preventive maintenance PM activities are relevant for the characteristics functional test, inspection, and calibration. Outside of these, only cleaning is a characteristic requiring a lot of time in preventive maintenance, which has no data.

When considering data analysis valuable, the business goal including a *hypothesis* has to be determined, where the analysis concludes if the PM activity is effective and in what frequency it is required. There are two hypothesis formats designed, which encapsulate these aspects. Firstly, to measure the effectivity of PM, the probability of an error during PM is often comprehensive to accept or reject the hypothesis, formulated as:

H0: The probability that the planned maintenance test ***X*** *passes is greater than* ***Y****%*

Considering a PM activity effective in terms of probability of error is not comprehensive, the other important aspect is the frequency of PM. Proving the effectivity, and determining the frequency using the data achieves the goal of creating data-driven PM activities.

Secondly, determining the required frequency makes the analysis comprehensive to reject or accept the right of existence for a PM activity. For example, with calibrations, the test could always be in the specification range for the calibration, but there could still be a drift, keeping the test valuable. It is essential to understand the test context since when there is a drift shown by the desired frequency in the second hypothesis; the effectivity hypothesis can comprehensively be accepted or rejected.

H0: The probability that the planned maintenance test ***X*** *passes consistently during a timeframe of* ***Y*** *years is greater than* ***Z****%.*

The hypothesis concludes the business understanding, thus the next step is to gather the *relevant data*; which requires domain knowledge gathered with subject matter expert (SME) interviews. Potential data sources are Vertica tables, SAP and service work orders (SWO), in collaboration with SMEs, it is required to define the relevant *attributes* of this data. Finally, there is an evaluation if the dataset is comprehensive enough to answer the hypothesis; if this is not the case, and no additional data is available, the process ends by failing to accept a hypothesis.

If the conclusion is that the milestone *"the dataset is comprehensive"* is achieved, it is time to prepare the data, firstly by *merging the datasets*, the identifier to do so is the system reference number (SRN). Additionally, datasets can be impure, therefore requiring handling quality issues as described in paragraph 4.6.3.3. Finally, by combining attributes, additional *analysis variables* are created, making the modeling easier at a later stage, at the end of this phase the analyst should achieve the milestone that the new "*dataset is comprehensive"* to accept or reject the hypothesis, and there are no preventable quality issues that weaken the conclusion.

Modeling starts with a comprehensive and qualitative dataset, where the analyst needs to start understanding the data to learn what modeling techniques are suitable, e.g. by performing a parametric test. To do so, *descriptive statistics* are the first step to understand the data, and *visualizations* enhance interpretation and are a presentation tool. Dependent on these results, the analyst can estimate what *inferential statistics* are suitable to test the hypothesis, repeating this process until the milestone *results are as comprehensive* is achieved for the hypothesis.

When the model results are as comprehensive as possible, there is an evaluation of the results in collaboration with an SME, to decide the next step in the analysis. If the conclusions are not comprehensive, the process can iterate again with an adjusted hypothesis or the SME can decide to terminate the process, concluding that data analysis cannot give a recommendation. If the results are evaluated to be comprehensive, the conclusion of the data analysis is valuable in terms of effectivity and frequency. Then it requires documentation to form input in defining the PM activities and interval, and only then, the milestone of evaluating the hypothesis is achieved.

Defining the PM activity frequency is a decision intensive process with multiple inputs; firstly, there are probabilities available to calculate the expected CM costs, while the expected PM costs are available as well. Combining these aspects with arguments outside of the data analysis, experts can decide on the PM activity frequency, if any, achieving the milestone of structurally reviewing PM, all the aspects are summarized in Figure 17 - Data analysis PM activities.

5.1.1 Validation

Although carefully created from an existing methodology and expert knowledge, designing this model happened by an individual, and even though there was testing of the model by applying it to nine data analysis, the recommendation is to validate the correctness and effectiveness of the model. Although, the researcher had limited resources to do validation, and therefore he describes potential validation methods ranked from the strongest to the weakest, namely

1. Have two groups do data analysis using PM activities, one group with, and the other without the manual and process description, and then compare the results of both groups compared to the business value of the results, and the time spent to reach a conclusion.
2. Give the manual and process description to a user, and let the user do the data analysis on a PM activity and evaluate the business value of the analysis and time of the researcher, including their feedback; the downside of this method is a lack of a baseline.
3. Evaluate the manual and process description by discussing it with SMEs, this validation does not go in depth, but it might catch simple or fundamental mistakes.

5.1.2 Example of an application of the manual

To show how the model works for an actual test, this chapter describes one example, where Appendix E describes the example the 'Calibration time-based 1' in chapter A, while other analyses are described in more detail in the other chapters, all including test values and figures.

The component related to this test had high CM costs and therefore triggering the process of the data analysis, where this calibration was one of the PM activities attempting to mitigate the high CM costs. Additionally, the SME stated that the calibration might be outdated because of a design change to a solid-state model enhancing the expected value of an analysis. Therefore, the PM activity is preventive-maintenance on a timely basis (yearly) where the characteristic is calibration. Since the motivation of this category is not mandatory and there is data for calibrations, data analysis is feasible and worthwhile to test the following hypothesis

- The probability that the PM-test Calibration time-based 1 passes, is greater than 99%
- The probability that the PM-test Calibration time-based 1 passes consistently during the service lifetime of 10 years, is greater than 95%

With the hypothesis concluding the business understanding, relevant data is gathered. For the PM tests that consisted of all PM activity data, PM visitation dates, Product models, Magnet type, number of scans, log file completeness and amplifier type. The relevant attributes for the analysis were SRN, Date, Test results, ValFloat, Specs, Log file completeness, Number of scans, magnet type, product model and amplifier type, judging this data comprehensive enough to answer the hypothesis, leading to the data preparation.

Merging the datasets happened via linking by SRN. There were quite some quality issues, therefore applying the following filters:

- Remove all SRN which had multiple product model configurations
- Remove all SRN which had multiple Amplifier type configurations
- Remove all SRN which had no known product model or amplifier type configurations
- Remove all duplicate values in one day
- Remove all SRN which had an age younger than one year
- Remove all SRN < 100 or SRN > 97000 which are possible test machines
- Remove all tests with a rare parameter ID (~50 observations)
- Remove all data points which are in the first month of the first test (installation data)

Creation of analysis variables simplifies and speeds up analysis, therefore introducing:

- PM: a Boolean if a test date was during a PM visitation for an SRN
- Age: a double, the years between the oldest test and now, estimating the system age
- DaysSinceFirstTest: an integer, the days between the oldest test and the current test
- Data points: an integer, the number of data points for the SRN + ParId combination
- Scans normalized: the number of scans divided by the log completeness (in %)

This resulted in the final dataset, suitable to answer the hypothesis. Firstly, using some descriptive statistics it was learned that for the modeling phase:

- All the attributes had statistical significant influences according to ANOVA tests, like the amplifier type and parameter IDs
- The probability of error did not differ over attributes, except for the Boolean PM; 0.01% compared to 0%
- Means and standard deviations do not explain error rates

From visualizing the data, other relevant aspects learned were:

- When looking at 'passed' values, linear regression seems suitable
- The data were normally distributed when using parametric tests, illustrated in Figure 18

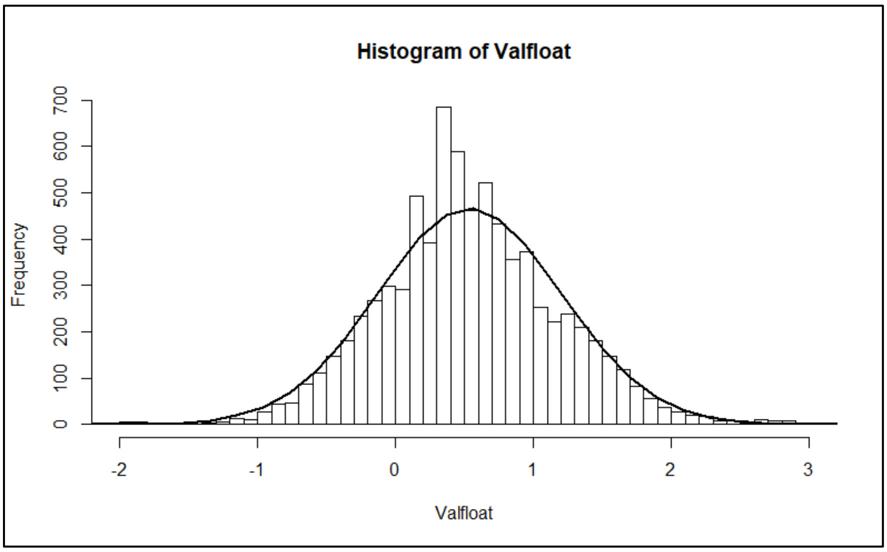

Figure 18 - Example parametric test

With a calibration, the goal is to adapt for drift (described with an alpha value), a potentially suitable method to predict this drift is linear regression, and since the float values are parametrically distributed this is feasible. The linear regressions focuses on each unique SRN, as a function of Parameter IDs. To do so, it requires a threshold defining a minimum amount of data points to do the analysis, which is set to six. This results in reducing the amount of unique SRN analyzed by 32%, with 9402 remaining observations that accounted for 19% of the total.

The aggregated results per ParId and RfType in terms of average drift, the average time to run out of spec (abbreviated with TTROOS), and number of SRN running out of spec in 10 years (abbreviated with OOS10) is illustrated in Table 14.

Table 14 – Linear analysis results per Rf amplifier type and Parameter ID with $R^2 > 80\%$

ParId\Rf	**Rf_1**	**Rf_2**	**Rf_3**	**Rf_4**
High 1	μ: 0.63 σ: 0.52 α: -0.016 (per year) β: 0.59 TTROOS: 213 year OOS10: 0	μ: 1.22 σ: 0.62 α: 0.063 (per year) β: 1.28 TTROOS: 43 year OOS10: 0	μ: 0.83 σ: 0.48 α: 0.017 (per year) β: 0.79 TTROOS: 189 year OOS10: 0	μ: 1.14 σ: 0.48 α: 0.017 (per year) β: 1.18 TTROOS: 166 year OOS10: 0
High 2	μ: 0.72 σ: 0.51 α: 0.0016 (per year) β: 0.75 TTROOS: 2000 year OOS10: 1	μ: 0.87 σ: 0.58 α: 0.11 (per year) β: 0.88 TTROOS: 28 year OOS10: 3	Not applicable	Not applicable
Low 1	μ: 0.03 σ: 0.55 α: 0.05 (per year) β: 0.046 TTROOS: 79 year OOS10: 1	μ: 1.23 σ: 0.79 α: 0.01 (per year) β: 1.23 TTROOS: 277 year OOS10: 1	μ: 0.99 σ: 0.69 α: -0.013 (per year) β: 0.95 TTROOS: 250 year OOS10: 0	Not applicable
Low 2	μ: 0.38 σ: 0.57 α: -0.04 (per year) β: 0.42 TTROOS: 90 year OOS10: 0	μ: 0.85 σ: 0.74 α: 0.12 (per year) β: 0.82 TTROOS: 27 year OOS10: 6	Not applicable	Not applicable

This table contains some important results, firstly that systems take much longer than ten years to run out of specification on average. Additionally, only 12 unique systems run out of specification in ten years, which is a percentage smaller than the 5% in the hypothesis.

For successful linear regression compared to number of scans, the aspect of log completeness becomes important. This is a different table from Vertica, keeping daily logs of all scans. Summarizing these scans gives an indication of how much an SRN scans, but these values need normalizing over log completeness. Therefore, thresholds include the amount of logs and log completeness, specifically the threshold for regression analysis is > 5 data points, > 100 logs > 70% completeness. After the log filter, the remaining unique SRN and data points gave a too small sample size to make statements or conclusions for the entire population.

The probability of error is low, only 40 errors in over 50000 measurements, this results in an on average error probability of <0.1%, and even 0% (zero errors absolute) during PM visitations. Since scan and PM data is only available since 2016, these are not suitable for trend analysis, where there are no general linear trends that predict systems running out of specification, either on fleet level or on system level, these conclusions are based on data outside of PM activities as well, while in terms of the hypothesis, the conclusions are:

- The probability that the PM-test Calibration time-based 1 passes, is greater than 99%

H0 accepted: the probability that the PM-test Calibration time-based 1 passes is 100% in the PM visitation period 2016-2018.

- The probability that the PM-test Calibration time-based 1 passes consistently during the service lifetime of 10 years, is greater than 95%

H0 accepted: the PM-test Calibration time-based 1 passes over a consecutive period of 10 years for more than 95% of the MR systems.

Before reaching this conclusion, there have been multiple iterations of the analysis. In addition, verification happened via a data analysis expert at Philips Healthcare. After this stage, there was documenting of the results, as shown in Appendix E chapter A. These conclusions reach the SMEs relevant for deciding if the activity remains in the PM manual.

This implementation was one of the nine examples of data analysis; only one is described because the analyses are comparable. This is because the process is standardized, the key aspects are all treated in this one example, and the process has become repeatable. In essence, all the milestones are verifications if the business artifact meets a certain condition in its aspects. For example, if the dataset is comprehensive depends on the aspect of data points, and the milestone if the data analysis is valuable is dependent on the motivation and characteristic aspects of the business artifact. This way, PM activities that have the same business artifact attributes in this process, should evolve towards the same decision in terms of recommendation based on the data hypothesis. Other aspects that determine what the required frequency is for a PM activity is considered out of scope since these are outside of the frame of data-driven maintenance.

5.1.3 Results of all data analysis applications

This paragraph illustrates the results from each data analysis, the implication of this sample for the whole data set, and future strategies for each characteristic group of PM activities. Detailed descriptions of the data-analysis are in Appendix E, while Table 15 - Analyzed Planned maintenance activities summarizes the analyzed activities and their results over the two hypothesis formats, as illustrated below.

*H0 effectivity: The probability that the planned maintenance test **X** passes is greater than **Y**%*

*H0 frequency: The probability that the planned maintenance test **X** passes consistently during a timeframe of **Y** years is greater than **Z**%*

Table 15 - Analyzed Planned maintenance activities

Test	**H0: Effectivity**		**H0: Frequency**		**Result**
	Goal	*Conclusion*	*Goal*	*Conclusion*	*During PM*
Calibration 1	99%	H0 accepted, 100% in the PM visitation period >2016	10 years 95%	H0 accepted, 98% of SRN for 5 years, extrapolated to >95%	Stop
Calibration 2	99%	H0 accepted, 100% in the PM visitation period >2016	10 years 95%	H0 accepted, 98% of SRN for 5 years, extrapolated to >95%	Stop
Calibration 3	95%	H0 rejected, 92% in the PM visitation period >2016	10 years 95%	H0 rejected, 79% of SRN for 5 years, thus even less in 10	Keep
Calibration 4	95%	H0 rejected, 28% in the PM visitation period >2016	10 years 95%	H0 rejected, 47% of SRN for 5 years, thus even less in 10	Keep
Calibration 5	95%	H0 accepted, >99.9% in the PM visitation period >2016	10 years 99%	H0 accepted, 99% of SRN for 5 years, extrapolated to >99%	Stop
Functional test 1	95%	H0 accepted, >97% in the PM visitation period >2016	10 years 95%	H0 rejected, 85% of SRN for 5 years, thus even less in 10	Keep
Functional test 2	95%	H0 rejected, 85% in the PM visitation period >2016	10 years 95%	H0 rejected, 63% of SRN for 5 years, thus even less in 10	Keep
Functional test 3	99%	H0 rejected, 85-95% in the PM visitation period >2016	10 years 95%	H0 rejected, 85% of SRN for 5 years, thus even less in 10	Keep
Functional test 4	99%	H0 accepted, 100% in the PM visitation period >2016	10 years 99%	H0 accepted, 99.9% for 5 years, extrapolated to >95%	Stop

When extrapolating these results, four out of nine activities of preventive maintenance are redundant, based on a sample of 5 of the 15 calibrations and 4 of the 31 functional tests. Furthermore, there is a discussion about the effectiveness of doing functional tests during PM for components with random failures. The estimate of the SMEs with extrapolating these figures and knowledge is that about 30% PM reduction is a realistic target. When considering work steps fixed, this means saving 3 hours per PM visitation, which is significant.

5.2 Design of planned maintenance activities

The learning is that the component design determines which PM activities are required, therefore instead of formulating PM activities from an existing design, optimization for effectivity and frequency requires integrating PM in the design choices. The goal of this process model is not to catch the full process of component design, since this is a complicated process of R&D, and therefore the focus is rather on the decision-making during the component design.

Because the focus is limited to the decision-making, the model is not extensive; it does not contain all the aspects related to component design but just the ones influencing the design choice. While a design process may be complex, the design choices are a consideration between attributes and costs. The design has some attributes that are mandatory and some attributes that are nice to have, for the latter there is a consideration between costs and having the attribute. Therefore, decision-making in the design process is a trade-off between component attributes and total component costs.

This decision-making happens in design teams of R&D, who make considerations in different design choices. For example, they can make the consideration between a more expensive component that does not require maintenance and a cheaper component that does. Currently, the bill of material costs, so component costs, measures the design performance of R&D rather than the total lifetime or maintenance costs, resulting in a preference for cheaper components. In the desired situation, this consideration happens in a holistic way to guarantee the best decision for Philips Healthcare, making PM an integral part of the decision-making in component design.

Because the focus is on the decision-making, a different modeling standard than CMMN is more suitable, namely the Decision Model and Notation (DMN) also published by the Object Management Group. One of the advantages is that DMN is usable for a standalone process, but also complementary to CMMN, and therefore our data analysis model. This model has the goal to illustrate the decision-process in a simple readable way, making decisions repeatable.

The first aspect in the DMN is the determined characteristics and motivation in applying these ensures that engineers consider all aspects of PM when they think about a component so that they do not forget certain aspects. These characteristics and motivations support in formulating the *required PM activities* in a comprehensive way.

Outside of the required PM activities, other attributes than PM have an essential role in this decision making as well, examples acquired from the interviews are manufacturability, reliability, serviceability and functionality. These *other component attributes* play a significant role in the component design, and it is not to say if these are of higher or lower importance than PM. Therefore, in the decision process of the component design engineers needs to balance the best choice for Philips Healthcare considering the whole component lifetime and related costs.

Although the goal of the proposed process is to balance the decision variables, costs remain an important aspect nonetheless. This is because it might be possible to justify a cheap component choice and accept lower performance in other attributes. For example, cheaper fans have wear and tear and require more replacing in the lifespan of a machine, while the alternative is designing fans with ceramic bearings that are expensive. However, since replacing the cheaper fans correctively over the course of the lifetime is cheaper than using one fan with ceramic bearings, this justifies the choice for using the cheaper component.

Therefore, the final *component design* is dependent on multiple *component attributes*, where one of these is the *required PM activities* derived from the PM characteristics and motivations. The choice of the *component design* is a holistic consideration between all component attributes, and their respective cost (savings) compared to the component costs of the design, with the priority of making the best choice, summarized in Figure 19 - Decision-making process component design.

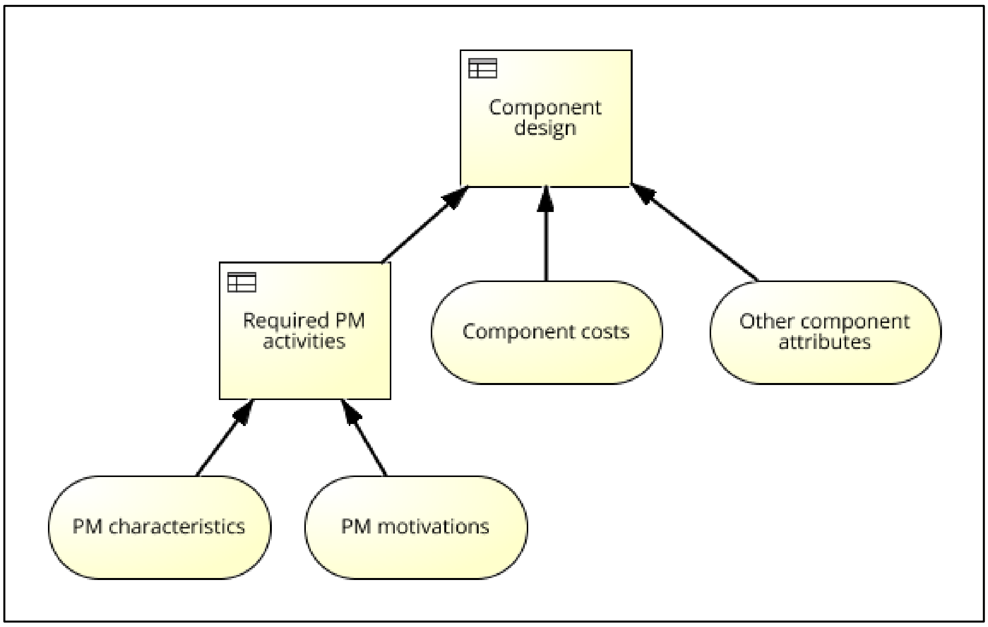

Figure 19 - Decision-making process component design

5.3 Defining planned maintenance activities

The new desired process of designing PM activities as illustrated in chapter 1 consisted of two deliverables. Firstly, there is a design in CMMN of a methodology to do the "review of PM activities", described in detail in paragraph 5.1. This methodology uses the available *PM activity data* at Philips Healthcare to evaluate PM activities on effectivity and required frequency to create input for the process of redesigning the existing PM activities, updating the existing *PM activity set* documented in the *PM Manual*.

Secondly, there was a "methodological design of PM activities", the key learning is that the design of PM activities is directly dependent on the design choices made, and therefore the process of designing PM activities is not independent, but rather an integral part of the *component design*. As described in paragraph 5.2, in the updated process *component design*, making decisions happen in a holistic way considering all desired component attributes and costs over the entire lifetime, and PM is one of these attributes. To support analysis of PM, the defined characteristics and motivations for PM are usable.

In the proposed Business Process Model and Notation (BPMN) of designing PM activities in general, the PM activities have become an integral part of *component design* using *characteristics and motivations* of PM activities as input for this decision-making modelled in DMN. Additionally, there is a CMMN model with a methodology on how to *review PM activities* to keep evaluating their effectivity and frequency using the *PM activity data*. This new process of designing PM activities is summarized in Figure 20 - Desired process of defining PM activities, which effectively uses the triple crown method, by combining BPMN, DMN and CMMN (OMG Group, 2016).

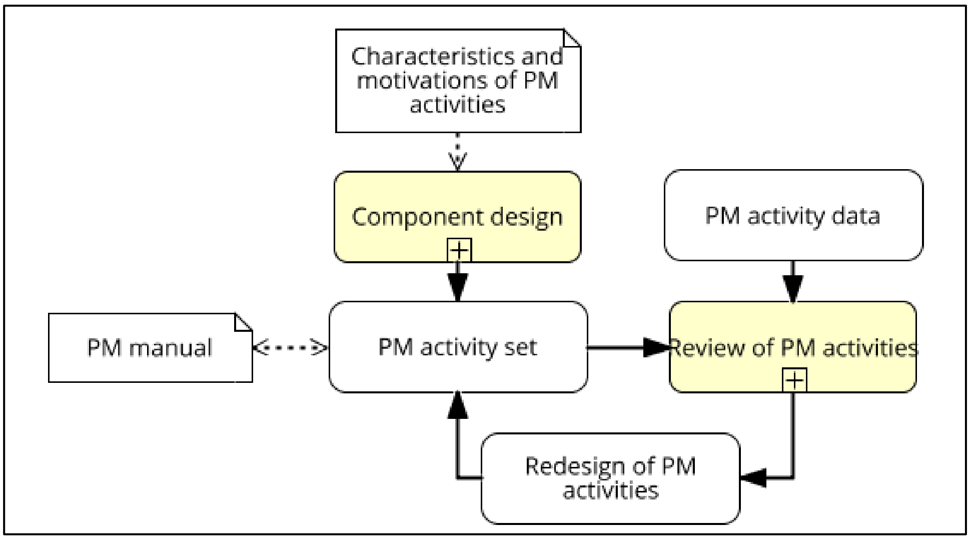

Figure 20 - Desired process of defining PM activities

6 Conclusion

The goal of the research was to design a process for defining planned maintenance (PM) activities that optimize the process of designing and reviewing PM activities for effectivity and frequency, using historical PM activity data for this optimization. In the proposed design, there were two deliverables, (1) a process of reviewing existing PM activities using historical data and (2) an optimized design for defining PM activities. To achieve a grounded design for these deliverables, four research questions required answering, where the research method to do so consisted of thematically analyzing interviews and data analysis using CRISP-DM.

Firstly, there was explorative research *on what the current process was of designing planned maintenance (PM) activities (RQ1)*. It became apparent that there are two departments relevant for PM activities, research & development (R&D) for the component design and thus the required PM activities, and service innovation (SI) who are responsible for prescribing the PM activities. Currently, they have different goals in relation to PM since measuring performance happens on component costs at R&D, while for SI the maintenance costs are leading. In the desired situation, they have a common business goal, enabling a holistic evaluation considering PM during the component design, resulting in the best decision for Philips Healthcare.

Secondly, there was the required input of data, related to the question *What PM activity data is available to analyze current planned maintenance activities (RQ2)*. The largest source is the new Vertica database system, which enhanced the accessibility of machine data at Philips Healthcare. Having nearly the entire installed base connected, uploading machine log files centrally, the relevant data in Vertica includes all the maintenance activity data, product models and configurations. Additionally, there is data available like the PM visitation dates, making it possible to have a good estimation if a maintenance activity was part of PM. Although this data is only available since 2016, this decreases the sample size because of the smaller timeframe.

The question of *how the selected data should be used to review decision making for PM activities (RQ3)* treats the reviewing aspect of the activities. It uses some of the available big data in Vertica, mainly the full history of all PM activities ever done, and in addition, the PM visitation dates known in SAP. With the selected data, a sample of 9 PM activities are analyzed. This research question focuses on defining the key business artifacts as well, including their aspects which manipulations have a determining effect on the course of the process.

By applying the CRISP-DM methodology for the analysis, and making the model more situation specific using the learnings from the data analysis, it is possible to propose a specific methodology for analyzing PM. This methodology is based on the general hypothesis on effectivity and required frequency of PM activities, analyzing these, in the same manner, to accept or reject them. The methodology is based on business artifacts, where the relevant attributes that manipulate the model are identified in the data analysis and interviews, making the process reproducible and repeatable, which is a prerequisite of decision-intensive processes.

The sample of the existing PM activities resulted in a conclusion that four of the nine activities are redundant; proving that evaluating the PM activities is worthwhile in the process. The results of this sample can be extrapolated by their characteristics, this way it is possible to make an estimate of the potential value of doing the analysis, which is a time reduction of about 30%.

Designing components requires many decisions considering multiple aspects, like costs and required maintenance, and the question of *how the process of designing planned maintenance activities should look like (RQ4)* treats this decision-making. With the newly formulated motivations and characteristics for PM activities, the new process has the tools to include PM in the design decision-making. Effectively answering the problem statement, since it is a proposal for a new *process of designing and reviewing planned maintenance activities, optimized for effectivity and frequency by using historic planned maintenance activity data.*

The learning is that the component design determines which PM activities are required, therefore instead of formulating PM activities from an existing design, optimization for effectivity and frequency requires integrating PM in the design choices. The final process only contains the decision-making in the component design and is therefore modeled in the Decision Model and Notation. In this model, there are multiple component attributes, where one of these is what the required PM activities are, derived from the PM characteristics and motivations. In the proposed design, design choices happen in a holistic way to improve the performance in terms of total lifecycle costs and customer satisfaction for Philips Healthcare as an organization.

In the new process of designing PM activities, defining the PM activities has become an integral part of the component design, and reviewing PM activities happen in a structural way. These deliverables are combined into the proposed process of designing PM activities, using the triple crown method, by combining BPMN, DMN and CMMN (OMG Group, 2016). This new process optimizes the design of existing PM activities for effectivity and frequency.

6.1 Limitations

Although the research has interesting results, improvements are possible. Firstly, the current sample of PM activities is nine of around forty PM activities that have data aspects. Additionally, the current analysis is limited to calibrations and functional tests, the results are based on a sample of five of the fifteen calibrations and four of the thirty-one functional tests. These two characteristics represent 30% of the total PM time, and with the addition of inspections and cleaning this percentage can increase up to 80%.

Secondly, an individual did the whole research. In such a situation, there is an inherent bias to the research because of the researchers' individual perceptions (Atieno, 2009). The usage of (objective) methods, the amount of verification by other researchers, and the reproducibility of the research can minimize this bias. Especially the validation of the designed models and manual is disputable because there was no objective evaluation on effectivity and validity. Although, to mitigate this limitation, the basis of the designs are from a renowned model (CRISP-DM), and the model is applied on nine PM activity cases, but this is not the most rigid method of validation.

The environment of the research is the magnetic resonance customer service department at Philips Healthcare in the Netherlands. Therefore, the research is adapted to this environment, for the business application of the client, although the researcher attempts to keep the methodologies and results as general as possible.

Currently, the data-driven aspect of the research is limited to historical maintenance data to evaluate on effectivity and frequency of PM activities, but there is also potential in using available data, especially remote data that is available on a constant or daily basis. Additionally, the research identifies opportunities for using this data for condition-based maintenance, which has the potential to be even more efficient than an optimized time-heuristic maintenance schedule.

An additional solution direction found in the literature uses the theory of condition-based maintenance (CBM), which has the potential to be even more efficient than an optimized time-heuristic maintenance schedule. The data-driven aspect of the research is limited to historical maintenance data to evaluate effectivity and frequency of PM activities, but SMEs recognize opportunities for CBM. It is noted that a design change implementing additional sensors can take up to six years to be effective.

6.2 Recommendations for future research

Derived from the conclusions, findings and limitations in this research, there are recommendations for improvement and future work of the research, and on how to apply the research at Philips Healthcare, to gain business value from the research.

6.2.1 Improving current work

Firstly, a larger sample set will increase the confidence of the current conclusions that required extrapolating the results of this sample, e.g. in terms of the total amount of expected activities to be redundant. Additionally, the current sample set exists solely of functional tests and calibrations, and there is a high expectation from SMEs that the characteristic inspections have a large potential as well, but because of time constraints, these are not analyzed in this research.

Secondly, there is the issue with an individual doing the research, creating a potential researchers bias. To prevent this, objective validation methods can check or improve the current research. An example would be re-coding the interviews with the given codebook by another individual, reproducing a part of the research. If this coding is comparable to the original, it partly validates the interviews, although different interpretations remain a possibility. To measure the definition of ‘comparable’, intercoder reliability or intercoder agreement can be used (Campbell, Quincy, Osserman, & Pedersen, 2013).

Thirdly, there is the validation of the designs, and especially the data analysis methodology and manual. One method of validating the effectivity of this methodology is by performing the data analysis with two groups, an experimental group and control group. The experimental group has the methodology and the manual to perform data analysis, while the control group does not. The difference in effectiveness per group is measurable when they perform the same data analysis if the experimental group returns better or the same conclusions in less time, the methodology and manual proof to enhance performance.

Finally, the researcher attempts to keep the research as general as possible. Especially for other business units in Philips Healthcare, this is an important aspect if the research is reproducible for e.g. computed tomography, enhancing the business value greatly since the deliverables are applicable in the whole of Philips Healthcare, instead of being limited to magnetic resonance. There is the desire of Philips Healthcare in reproducing the current research in the business unit of image-guided therapy, a larger business unit with comparable goals as magnetic resonance.

6.2.2 Future academic research

In terms of extending this research on academic grounds, the choice of using historical maintenance data resulted in an unsuitable environment for condition-based maintenance (CBM), but there is potential for CBM using daily available data. Such data is available at Philips Healthcare because machine log files are uploaded daily containing potentially rich information. There is already some recognition of condition-based opportunities that may be suitable for future work, although the recommendation remains to evaluate the maintenance activity on effectivity first, before determining the moment of maintenance by a CBM method.

Furthermore, the current research successfully used artifact-centric process modeling for doing data analysis, but it would be interesting to explore the use of artifact-centric process modeling on condition-based maintenance, especially in combination with decision-intensive processes. This is feasible for corrective maintenance activities at Philips Healthcare using the described opportunities for condition-based maintenance that have daily available data.

6.3 Recommendations for Philips Healthcare

With this research, Philips Healthcare can start analyzing all PM activities with the designed methodology. The analyzed sample shows that a group of activities may be redundant, where the extrapolated saved time by not having to do PM activities is large enough to justify the required effort, the estimated amount of reduction is 30% according to SMEs.

The saved time is valuable since it reduces the total PM time spent in the field, reducing labor costs. Additionally, currently the PM activities take up eight hours per visitation, and this puts the field service engineer under stress to finalize the PM in these eight hours, to prevent having costs in labor overtime. Finally, when reducing the eight hours, the downtime of the equipment reduces as well, thus making scheduling easier since PM in just the afternoon becomes a feasible option, increasing customer satisfaction as well. Therefore, the recommendation is to analyze the PM for effectivity and frequency using the designed data analysis methodology.

During the research, discussions with multiple SME resulted in opportunities for condition-based maintenance (CBM). Some of these with currently available data, others by implementing other sensors. The most effective maintenance method is CBM (Al-Najjar & Alsyouf, 2003), and therefore evaluating these opportunities is the recommendation, but only after analyzing all the PM activities on effectiveness. The CBM opportunities became apparent with discussions that did not necessarily have the goal to identify these; they were rather a collateral result. Therefore, the expectation is that by having discussions focused on finding CBM opportunities, e.g. in a workshop setting, there is potential for discovering even more CBM opportunities. The assumption is that CBM is more efficient, since the usage per MRI-Scanner varies widely, and since there are activities implemented this way already, the expectation is that CBM is useful.

To summarize, a basic analysis is already valuable, and there is an opportunity for condition-based maintenance. Therefore, the recommendation is to analyze all the activities with the basic methodology reducing PM with about 30%, and then consider CBM for the remaining PM activities, starting with the found opportunities and explorative research for other opportunities.

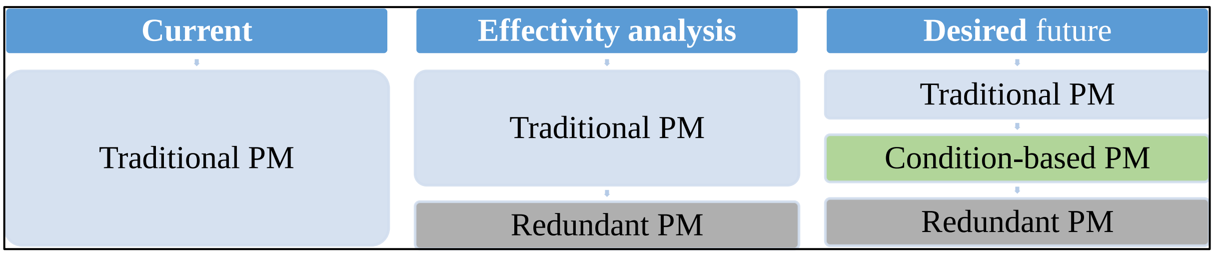

Figure 21 - Evaluation PM Activities after data analysis

6.3.1 Strategy per characteristic

When deciding on the planned maintenance activities to analyze, typical activities were preferred. The goal of this paragraph is to describe strategies to reduce the required maintenance time per group of PM activities, dividing them by characteristic. The choice of dividing them per characteristic, rather than motivation, is because the motivation has a strict influence on the conclusion. If the motivation is from a mandatory standard, there is no possibility of reconsidering the spent time on the activity; it has to stay exactly as it is, making the motivation more of a prerequisite before applying a strategy rather than a strategy categorization.

Analyzed tests consisted of functional tests and calibrations. Unfortunately, there was no time to analyze any inspections, and a quantitative analysis for the characteristic cleaning was not possible due to a lack of data. The other group of characteristics are relatively small, still, there were learnings for these as well based on the SME interviews.

Learnings from the data analysis show that there is an opportunity in reducing the number of functional tests and calibrations, while an analysis is relatively quick and easy. The expectation of SMEs is that some inspections might be ineffective, and they suspect that the results are comparable with those of the functional test. Additionally, learnings from interviews give recommendations for the other activities as well, Table 16 - Recommended strategy per characteristic summarizes these learnings in strategies.

Table 16 - Recommended strategy per characteristic

Characteristic	Time	Recommended strategy
Functional test	24%	Analysis of effectivity and frequency, the expectation of redundant activities since errors are expected to be random in nature, making the tests redundant unless they are mandatory in risk management
Inspection	25%	Analysis of effectivity and frequency, the expectation of redundant activities according to SMEs, unless they are in risk management
Calibration	6%	Analysis of effectivity and frequency, the expectation of many redundant calibrations due to sample and SME expectations
Cleaning	25%	No data currently available, condition-based maintenance opportunity by installing temperature sensors or acquiring data from current sensors not remotely connected (e.g. CPU sensor)
Software rev.	3%	Research if the current PM activity is feasible to be done remotely
Lubrication	1%	Diagnosed as unique for certain SRN, no general strategy required
Replacement	0%	Currently non-existent, there is potential to analyze if preventively replacing components actually reduces maintenance costs
Workflow	16%	None – Dependent variable

References

Aken, J. Van, & Andriessen, D. (2011). Ontwerpgericht wetenschappelijk onderzoek. In *Handboek Ontwerpgericht Wetenschappelijk Onderzoek*.

Al-Najjar, B., & Alsyouf, I. (2003). Selecting the most efficient maintenance approach using fuzzy multiple criteria decision making. *International Journal of Production Economics, 84*(11), 85–100.

Arif-Uz-Zaman, K., Cholette, M. E., Ma, L., & Karim, A. (2017). Extracting failure time data from industrial maintenance records using text mining. *Advanced Engineering Informatics*, *33*, 388–396. https://doi.org/10.1016/j.aei.2016.11.004

Asmai, S. A., Hussin, B., & Mohd Yusof, M. (2010). A framework of an intelligent maintenance prognosis tool. *2nd International Conference on Computer Research and Development, ICCRD 2010*, 241–245. https://doi.org/10.1109/ICCRD.2010.69

Atieno, O. P. (2009). AN ANALYSIS OF THE STRENGTHS AND LIMITATION OF QUALITATIVE AND QUANTITATIVE RESEARCH PARADIGMS, *13*, 13–18.

Ayo-Imoru, R. M., & Cilliers, A. C. (2018). A survey of the state of condition-based maintenance (CBM) in the nuclear power industry. *Annals of Nuclear Energy*, *112*, 177–188. https://doi.org/10.1016/j.anucene.2017.10.010

Baptista, M., Sankararaman, S., de Medeiros, I. P., Nascimento, C., Prendinger, H., & Henriques, E. M. P. (2018). Forecasting fault events for predictive maintenance using data-driven techniques and ARMA modeling. *Computers and Industrial Engineering*, *115*(February 2017), 41–53. https://doi.org/10.1016/j.cie.2017.10.033

Benatallah, B., Bestavros, A., Catania, B., Haller, A., Manolopoulos, Y., Vakali, A., & Zhang, Y. (2015). A Survey on Approaches to Modeling Artifact-Centric Business Processes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *9051*, 117–132. https://doi.org/10.1007/978-3-319-20370-6

Bousdekis, A., Magoutas, B., Apostolou, D., & Mentzas, G. (2015). A proactive decision making framework for condition-based maintenance.

Bousdekis, A., Papageorgiou, N., Magoutas, B., Apostolou, D., & Mentzas, G. (2017). A Proactive Event-driven Decision Model for Joint Equipment Predictive Maintenance and Spare Parts Inventory Optimization. *Procedia CIRP*, *59*(TESConf 2016), 184–189. https://doi.org/10.1016/j.procir.2016.09.015

Bromberg, D. (2007a). BPM for Knowledge Workers: The Structural Foundations of Decision Intensive Processes (DIPs). *BPTrends*, (January), 1–7.

Bromberg, D. (2007b). BPM for Knowledge Workers Inside Decision Intensive Processes (DIPs): Knowledge, Practice, Context, and Characteristics. *BPTrends*, (February), 1–10.

Bruno, G. (2013). Coordination issues in artifact-centric business process models. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8055 LNCS*(PART 1), 209–223. https://doi.org/10.1007/978-3-642-40285-2_19

Buchmann, A. K. F. F. (2014). Modeling and execution of event stream processing in business processes. *Information Systems*, *46*, 140–156.

Bumblauskas, D., Gemmill, D., Igou, A., & Anzengruber, J. (2017). Smart Maintenance Decision Support Systems (SMDSS) based on corporate big data analytics. *Expert Systems with Applications*, *90*, 303–317. https://doi.org/10.1016/j.eswa.2017.08.025

Campbell, J. L., Quincy, C., Osserman, J., & Pedersen, O. K. (2013). Coding In-depth Semistructured Interviews : Problems of Unitization and Intercoder Reliability and Agreement. https://doi.org/10.1177/0049124113500475

Catelani, M., Ciani, L., & Venzi, M. (2017). Condition-based maintenance and Markov modelling for avionics devices. *4th IEEE International Workshop on Metrology for AeroSpace, MetroAeroSpace 2017 - Proceedings*, 335–339. https://doi.org/10.1109/MetroAeroSpace.2017.7999592

Cipollini, F., Oneto, L., Coraddu, A., Murphy, A. J., & Anguita, D. (2018). Condition-Based Maintenance of Naval Propulsion Systems with supervised Data Analysis. *Ocean Engineering*, *149*(August 2017), 268–278. https://doi.org/10.1016/j.oceaneng.2017.12.002

Cohn, D., & Hull, R. (2009). Business artifacts: A data-centric approach to modeling business operations and processes. *IEEE Data Eng. Bull*, *32*(3), 3–9. https://doi.org/10.1007/978-3-540-88873-4_17

Common Statistical Tests. (2018).

EN 13306:2010 standard. (2010). Brussels.

Feng, Q., Li, S., & Sun, B. (2012). An Intelligent Fleet Condition-Based Maintenance Decision Making Method Based on Multi-Agent, 1–11.

Galar, D., Thaduri, A., Catelani, M., & Ciani, L. (2015). Context awareness for maintenance decision making: A diagnosis and prognosis approach. *Measurement: Journal of the International Measurement Confederation*, *67*, 137–150. https://doi.org/10.1016/j.measurement.2015.01.015

Gallardo-Echenique, E. E. (2015). An Integrative Review of Literature on Learners in the Digital Era, (December 2014). https://doi.org/10.5817/SP2014-4-8

Guan, T., Kuang, Y. C., Ooi, M. P. L., Cheah, X. G., Tan, Y. S., & Demidenko, S. (2011). Data-driven condition-based maintenance of test handlers in semiconductor manufacturing. *Proceedings - 2011 6th IEEE International Symposium on Electronic Design, Test and Application, DELTA 2011*, 189–194. https://doi.org/10.1109/DELTA.2011.42

Hsu, C. C., & Chen, M. S. (2016). Intelligent maintenance prediction system for LED wafer testing machine. *Journal of Intelligent Manufacturing*, *27*(2), 335–342. https://doi.org/10.1007/s10845-013-0866-3

Huiguo, Z., Rui, K., & Pecht, M. (2009). A hybrid prognostics and health management approach for condition-based maintenance. *Industrial Engineering and Engineering Management, 2009. IEEM 2009. IEEE International Conference On*, 1165–1169. https://doi.org/10.1109/IEEM.2009.5372976

Hull, R. (2008). *Artifact-Centric Business Process Models: Brief Survey of Research Results and Challenges*. *On the Move to Meaningful Internet Systems: OMT 2008 Part 2*.

Hull, R. (2011). Towards flexible service interoperation using business artifacts. *Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOC*, 20–21.

https://doi.org/10.1109/EDOC.2011.27

Hull, R. (2018). Rick Hull Profile. Retrieved from https://researcher.watson.ibm.com/researcher/view.php?person=us-hull

IBM. (1996). CRISP-DM help overview. Retrieved from https://www.ibm.com/support/knowledgecenter/de/SS3RA7_17.1.0/modeler_crispdm_ddita/clementine/crisp_help/crisp_overview.html

Jardine, A. K. S., Lin, D., & Banjevic, D. (2006). A review on machinery diagnostics and prognostics implementing condition-based maintenance. *Mechanical Systems and Signal Processing*, *20*(7), 1483–1510. https://doi.org/10.1016/j.ymssp.2005.09.012

Kinghorst, J., Geramifard, O., Luo, M., Chan, H.-L., Yong, K., Folmer, J., … Vogel-Heuser, B. (2017). Hidden Markov Model-Based Predictive Maintenance in Semiconductor Manufacturing: A Genetic Algorithm Approach.

Kisi, E., Durovic, Z., Kovacevic, B., & Petrovic, V. (2015). Application of T2 Control Charts and Hidden Markov Models in Condition-Based Maintenance at Thermoelectric Power Plants. *Shock and Vibration*, *2015*. https://doi.org/http://dx.doi.org/10.1155/2015/960349

Köpke, J. (2016). Towards Quality-Aware Translations of Activity-Centric Processes to Guard Stage Milestone, *9850*, 308–325. https://doi.org/10.1007/978-3-319-45348-4

Koutsos, A., & Vianu, V. (2017). Process-centric views of data-driven business artifacts. *Journal of Computer and System Sciences*, *86*, 82–107. https://doi.org/10.1016/j.jcss.2016.11.012

Last, M., Sinaiski, A., & Subramania, H. S. (2011). Condition-based Maintenance with Multi-Target Classi- fication Models, *29*(New Generation Computing, Ohmsha, Ltd. and Springer), 245–260.

Liao, W., Wang, Y., & Pan, E. (2012). Single-machine-based predictive maintenance model considering intelligent machinery prognostics. *International Journal of Advanced Manufacturing Technology*, *63*(1–4), 51–63. https://doi.org/10.1007/s00170-011-3884-3

Limonad, L., Boaz, D., Hull, R., Vaculin, R., & Heath, F. (2012). A generic business artifacts based authorization framework for cross-enterprise collaboration. *Annual SRII Global*

Conference, SRII, 70–79. https://doi.org/10.1109/SRII.2012.19

Marrella, A., Mecella, M., Russo, A., Steinau, S., Andrews, K., & Reichert, M. (2015). Data in Business Process Models, A Preliminary Empirical Study (Short Paper). *2015 IEEE 8th International Conference on Service-Oriented Computing and Applications (SOCA)*, 116–122. https://doi.org/10.1109/SOCA.2015.19

Meeker, W. Q., & Hong, Y. (2014). Reliability meets big data: Opportunities and challenges. *Quality Engineering*, *26*(1), 102–116. https://doi.org/10.1080/08982112.2014.846119

Meyer, A., & Weske, M. (2013). Activity-centric and Artifact-centric Process Model Roundtrip.

Mosaddar, D., & Shojaie, A. A. (2013). A data mining model to identify inefficient maintenance activities. *International Journal of Systems Assurance Engineering and Management*, *4*(2), 182–192. https://doi.org/10.1007/s13198-013-0148-7

Nigam, A., & Caswell, N. S. (2003). Business artifacts: An approach to operational specification. *IBM Systems Journal*, *42*(3), 428–445. https://doi.org/10.1147/sj.423.0428

Niu, G., & Pecht, M. (2009). A framework for cost-effective and accurate maintenance combining CBM RCM and data fusion. *Proceedings of 2009 8th International Conference on Reliability, Maintainability and Safety, ICRMS 2009*, 605–611. https://doi.org/10.1109/ICRMS.2009.5270119

OMG Group. (2016). BPMN, CMMN AND DMN SPECIFICATIONS AT OMG.

Patil, M. A., Patil, R. B., Krishnamoorthy, P., & John, J. (2016). A machine learning framework for auto classification of imaging system exams in hospital setting for utilization optimization. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, *2016–Octob*, 2423–2426. https://doi.org/10.1109/EMBC.2016.7591219

Patil, R. B., Patil, M. A., Ravi, V., & Naik, S. (2017). Predictive modeling for corrective maintenance of imaging devices from machine logs. *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 1676–1679. https://doi.org/10.1109/EMBC.2017.8037163

Provost, F. (2013). *Data Science for Business*.

Russo, A., Mecella, M., Patrizi, F., & Montali, M. (2013). Implementing and running data-centric dynamic systems. *Proceedings - IEEE 6th International Conference on Service-Oriented Computing and Applications, SOCA 2013*, 225–232. https://doi.org/10.1109/SOCA.2013.37

Schabenberger, O. (2016). Modern Machine Learning. Retrieved from https://www.linkedin.com/pulse/difference-between-statistical-modeling-machine-i-see-schabenberger/

Schmidt, B., Wang, L., & Galar, D. (2017). Semantic Framework for Predictive Maintenance in a Cloud Environment. *Procedia CIRP*, *62*(Cm), 583–588. https://doi.org/10.1016/j.procir.2016.06.047

Sipos, R., Fradkin, D., Moerchen, F., & Wang, Z. (2014). Log-based predictive maintenance. *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '14*, (April), 1867–1876. https://doi.org/10.1145/2623330.2623340

Solomakhin, D., Montali, M., Tessaris, S., & De Masellis, R. (2013). Verification of artifact-centric systems: Decidability and modeling issues. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8274 LNCS*, 252–266. https://doi.org/10.1007/978-3-642-45005-1_18

Sun, Y., Hull, R., & Vaculín, R. (2012). Parallel processing for business artifacts with declarative lifecycles. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7565 LNCS*(PART 1), 433–443. https://doi.org/10.1007/978-3-642-33606-5_27

Tobon-Mejia, D. A., Medjaher, K., Zerhouni, N., & Tripot, G. (2012). A data-driven failure prognostics method based on mixture of Gaussian hidden Markov models. *IEEE Transactions on Reliability*, *61*(2), 491–503. https://doi.org/10.1109/TR.2012.2194177

Trunzer, E., Weiss, I., Folmer, J., Schrufer, C., Vogel-Heuser, B., Erben, S., … Vermum, C. (2017). Failure Mode Classification for Control Valves for Supporting Data-Driven Fault Detection. *2017 Ieee International Conference on Industrial Engineering and Engineering*

Management (Ieem), 2346–2350.

Vaculín, R., Hull, R., Heath, T., Cochran, C., Nigam, A., & Sukaviriya, P. (2011). Declarative business artifact centric modeling of decision and knowledge intensive business processes. *Proceedings - IEEE International Enterprise Distributed Object Computing Workshop, EDOC*, 151–160. https://doi.org/10.1109/EDOC.2011.36

Wang, J., Li, C., Han, S., Sarkar, S., & Zhou, X. (2017). Predictive maintenance based on event-log analysis: A case study. *IBM Journal of Research and Development*, *61*(1), 11:121-11:132. https://doi.org/10.1147/JRD.2017.2648298

Wang, L., Qian, Y., Li, Y., & Liu, Y. (2017). Research on CBM information system architecture based on multi-dimensional operation and maintenance data. *2017 IEEE International Conference on Prognostics and Health Management, ICPHM 2017*, 167–172. https://doi.org/10.1109/ICPHM.2017.7998323

Wood, B. (2003). *Building care*. Wiley-Blackwell.

Yongchareon, S., Liu, C., & Zhao, X. (2011). An artifact-centric view-based approach to modeling inter-organizational business processes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *6997 LNCS*, 273–281. https://doi.org/10.1007/978-3-642-24434-6_22

Yongchareon, S., Liu, C., & Zhao, X. (2012). A framework for behavior-consistent specialization of artifact-centric business processes. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *7481 LNCS*, 285–301. https://doi.org/10.1007/978-3-642-32885-5_23

Zulfigar, A., & Bhaskar, B. S. (2016). Basic statistical tools in research and data analysis. *Indian J Anaesth*, *September*(60), 662–669.