Expanding an asset management reference model: bridging the gap between data and maintenance

Master Thesis Industrial Engineering & Management

Production & Logistics Management

Author
A.C. (Arjan) van Laar
a.c.vanlaar@student.utwente.nl

Supervisor Arcadis
Ir. H. (Hanno) Spoelstra

Supervisors University of Twente
Ir. R.L.A. (Rogier) Harmelink
Dr. E. (Engin) Topan

October 17, 2023

ARCADIS

UNIVERSITY OF TWENTE.

Abstract

A key aspect of a reference model for asset management is information management. Consulting firm Arcadis acknowledges this importance but does not understand the role of the concept in the domain. In this research, the answer to the role of information management in the Arcadis reference model is filled in through the development of data models and the development of data analytics methods. thereby answering the research question:

How can the Arcadis reference model used for asset management projects be improved in terms of data models, architecture, and analytics?

First, a literature review and context analysis are done to develop possible solution methods as well as generate knowledge already present at Arcadis. Through a survey FMECA and GIS were seen as the most common methods. Whilst in literature the importance of performance analysis, and the value of condition monitoring became apparent, as well as the role of a decomposition of an asset. These analyses resulted in five key points of development: asset classification and decomposition through an object type library; the methods of failure mode, effect, and criticality analysis; geographic information systems; as well as the role of performance analysis and condition monitoring and prediction. The first three points use static data (i.e., data that does not change over time), while performance analysis and condition monitoring and prediction use dynamic data (i.e., data that changes over time). The methods with static data require data less often compared to the other analyses and therefore require a different data model. Based on a use case, each of these methods is executed and analyzed for the most important data components.

To gather the data needed to do successful information management, we analyze the five key points in a use case. This use case is a movable bridge in the Netherlands. Based on the data available and used in analyses for the case, the data model is expanded and validated.

To further detail the five key points, first, we analyze the methods using static data. Asset classification is the backbone of asset management. As the same types of assets behave in similar patterns regarding failure statistics and behavior, analyses for these assets use the same information. Therefore, identifying what parts are in an asset (e.g., pumps, motors) can speed up the process. Both GIS and FMECA use data that is often readily available from the asset owner, and therefore do not require big data acquisition processes. FMECA is a relatively standardized method consisting of the same data components regardless of the project (e.g., time to failure, detectability). Therefore, a good understanding of the data through a data model helps in speeding up the process. GIS is a broad domain, only the basic data elements (e.g., location) are verified and put in a data model.

The dynamic data analyses are more complex and require more and constantly changing data. Performance analysis connects several key steps in the reference model between the strategic and operational levels. From our analysis of the use case, it became apparent that the performance of the movable bridge was unacceptable when compared to the business value framework of the asset owner. Therefore, asset information helps us validate the work and process. For condition monitoring and prediction, we found that there are several models which can predict the current condition of an asset. The machine learning models are compared based on accuracy, where a long short-term memory neural network performed best on accuracy followed by the random forest model. We then simulate using the behavior of the use case to compare the current maintenance

policies, which are preventive maintenance policies to the predictive maintenance policy. The predictive maintenance policy outperforms preventive maintenance policies, as the number of maintenance occasions and thus costs are 17% lower whilst the percentage of availability of the usable bridge is the same. From the performance analysis and condition monitoring analyses we then identify key data components such as performance measures and sensor signatures. We then create a data model for the key points involved with dynamic data.

Based on the analyses of the five key points, the reference model can be improved and expanded with a data model for information management containing the aspects of the analyses. This insight should save time in defining functions of an asset and combining this with failure, performance and sensor data. We thereby bridge the gap between data and maintenance engineering, specifically the gap for the five key points. The model can also be expanded with a more accurate description of the performance analysis and solution generation methods through condition monitoring which is currently lacking in the reference model. The research fills a gap in the literature between the value of information management and reference architecture in asset management practices specifically for movable bridges.

Contents

List of figures

List of Tables

List of Abbreviations

General

AM	Asset Management
EAM	Enterprise Asset Management
GIS	Geographic Information System
OTL	Object Type Library
RUL	Remaining Useful Life
SAMP	Strategic Asset Management Plan
SCADA	Supervisory Control And Data Acquisition
UML	Unified Modelling Language 2.0

Departments

ADM	Asset Data Management
AMIOI	Asset Management Infrastructure Objects and Installations
AMRA	Asset Management Rails
AMRO	Asset Management Roads
IM	Information Management

Methodologies

BTA	Bow Tie Analysis
CVIC	Computer Vision and Image Classification
DIS	Data Integration and Structuring
DPV	Data Preparation and Visualization
ETA	Event Tree Analysis
FMEA	Failure Mode & Effect Analysis
FMECA	Failure Mode, Effect & Criticality Analysis
FTA	Fault Tree Analysis
GIS ATR	Geographic Information System Analysis and Trend Recognition
GIS DAP	Geographic Information System Data Acquisition and Preparation
GIS V	Geographic Information System Visualization
NLP	Natural Language Processing
RAMS	RAMS(SHEEP) - Reliability, Availability, Maintainability, Safety, (Security, Health, Environment, Economics and Politics)
PC	Pareto Charts
RCA	Root Cause Analysis
RCM	Reliability Centered Maintenance
RR	Risk Register
RT	Risk Tree
TSA	Time Series Analysis

Sensors

EPS	(Electric)motor Power Sensor
CE	Cooling Efficiency
CP	Cooling Power
FS	(volume) Flow Sensor
PS	Pressure Sensor
TS	Temperature Sensor
VS	Vibration Sensor

1 Introduction

Maintenance management is becoming increasingly important in our current society. As the need for sustainability increases, the problem of dealing with existing assets becomes more important. As prolonging the life of built assets can help reduce the stress on new demand for such assets. This also holds for infrastructure assets in the Netherlands, such as railways and roads. Maintenance is an older concept that has gone through four main transitions or generations, as described by Arunraj and Maiti (2007). These generations are shown in Figure 1.

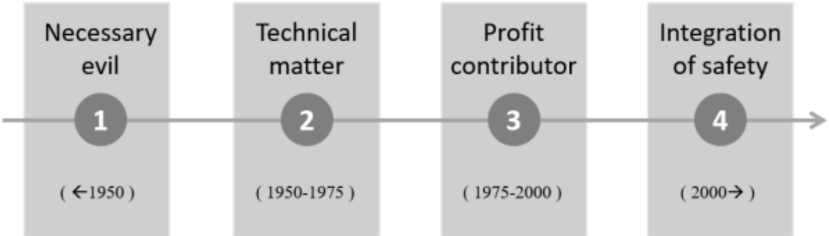


Figure 1 Generations of maintenance (Arunraj & Maiti, 2007)

Throughout the years, maintenance became more important because of the more complex systems and breakdowns of these systems. To cope with these complexities, over the years, new policies like preventive and condition-based maintenance were introduced. In the fourth and current generation, risk analysis has become the most important part of maintenance. Thereby integrating maintenance and safety. In this generation, the importance of data for maintenance is also acknowledged, as is the challenge of incorporating data analytics into the maintenance process (Arunraj & Maiti, 2007).

In this first chapter, we further introduce the thesis. Next up is the company introduction of Arcadis in Section 1.1. Research motivation and the objectives of this research are discussed in Sections 1.2 and 1.3, respectively. Next are the key definitions given in Section 1.4. The scope and limitations of the research are given in Section 1.5. Lastly, in Section 1.6, we provide the research questions.

1.1 Company introduction

This research is done at the company Arcadis. Arcadis is an engineering consulting firm. Worldwide, more than 36 000 people are working for Arcadis. The company has offices in more than 70 countries; however, the headquarters are based in the Netherlands. According to Engineering News-Record (2020), it was the ninth-biggest engineering firm in the world based on revenue generated in 2020. Arcadis wants to make a better world for everyone by focusing on sustainability in all projects they do. This is done through physical projects such as social challenges and places, but also in the digital world through data and technologies. As an engineering firm, Arcadis participates in various projects in various branches.

The structure of Arcadis is shown in the organogram in Figure 2. We are working from the Asset Data Management (ADM) team, but in close collaboration with the members of the other operate and maintain teams. The main activity of the operate and maintain department is creating company value from assets by finding the optimal balance between operational performance, financial performance, and the risks of operating and owning the asset.

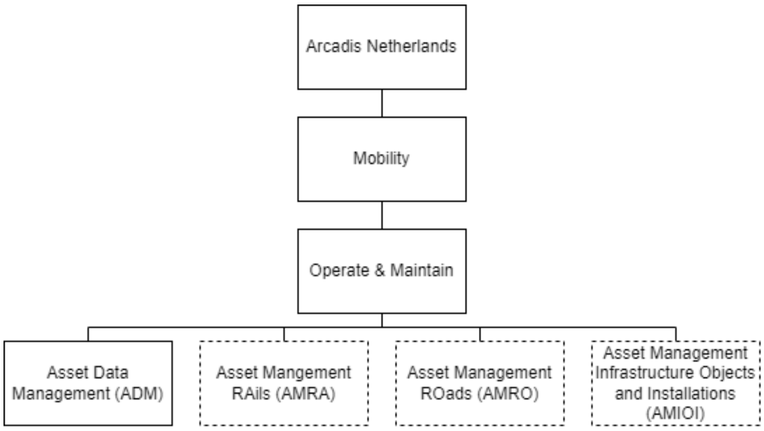


Figure 2 Organogram Arcadis Netherlands

Important for the Asset Management (AM) departments is the ISO 55000 norm. This is a standard on asset management released internationally in 2014. ISO 55000 introduces the main principles and technologies of asset management. Intricately linked are ISO 50001 and ISO 50002 which describe AM system requirements and AM system guidelines, respectively (International Organization for Standardization, 2014b).

1.2 Research motivation

The AM department is a large department within Arcadis and is expected to continue to grow as the need for more sustainable solutions rather than simple replacement arises. Therefore, asset managers might be consulted not only in the initial stages of assets but also in later stages of assets. They are mostly consulted to create structured and efficient maintenance plans in addition to long-term strategic asset management plans. With this challenge comes the increasing complexity of multiple working procedures and questions for example the various standards and types of assets.

AM is an old concept; however, only in the 1980s did the term start to be used in the private and public sectors, especially in relation to physical assets. Only recently have organizations recognized the need for asset management as a service that is much more than an extension of maintenance. This can also be seen in the ISO 55000 norm, which, compared to its previous versions, is more on a strategic level and integrates functional disciplines rather than isolated life phases, thereby moving towards the concept of lifecycle management. Additionally, assets are part of complex systems; their behavior is dynamic, and they are not only technical but also voiceless, therefore requiring monitoring (The Institute of Asset Management, 2015a; Tinga, 2013).

In current times, AM projects come with substantial amounts of data on asset performance and costs, while there are also vast amounts of data available on factors that might influence the state of the asset, such as weather conditions and human-caused events. To execute these projects in a structured way and to comply with the ISO 55000 standard, Arcadis has developed a process model. This model is used by various AM departments to maintain a high quality of projects.

The process model is developed iteratively. Current challenges lie in describing the inputs and outputs of the described methods and linking these inputs and outputs. For example, linking the outputs from performance analysis to the inputs for the next step risk analysis. This should help the methods and the people work better together. The main challenge is thus creating a common information model, containing a data model and a description of the tools and data used.

In this research, we aim to give practical relevance by generating insights on data structure (e.g., data models), data analysis methods (e.g., predictive condition monitoring), and a recommendation for future directions of the process model. We generate scientific relevance by researching information management practices in AM infrastructure and developing connections between these practices.

1.3 Research objectives

We have multiple objectives for this research. The first objective is to create insight into the information and tools used in current projects where the reference model is used. This insight should help in identifying the most important tools, information, and methods.

The most commonly used methods can be combined with a literature study for the second objective. This objective is to identify the lessons to apply to the development of information management in infrastructure AM. An addition to this objective is the secondary objective of delivering insight into the relevant standards for the project model and information model.

The third research objective of this research is to deliver an improved reference model with an accompanying information model, with which Arcadis can easily handle different projects and which can be implemented with other systems of customers and partners, thereby imposing generalizability.

The last objective of this research is to draw a conclusion on the currently used methods, their benefits, and the possible benefits of alternative methods that are less or not used currently. This conclusion should allow Arcadis to evaluate or expand the current process model with other relevant methods and tools. If these methods are new, a description of the method should be provided.

1.4 Definitions

In the field of asset management and infrastructure, a lot of different concepts and definitions are commonly used. To clarify the definition of various concepts, this section has been added. There are a lot of terms that are related or mean the same but have a different name. The first definitions are definitions as stated in the terms and definitions section of the ISO 55000 norm. In that list, various notes on further specifications are given; these are available in Appendix A (International Organization for Standardization, 2014b):

- **Asset**
  item, thing, or entity that has potential or actual value to an organization.
- **Risk**
  effect of uncertainty on objectives.
- **Asset management**
  coordinated activity of an organization to realize value from assets.
- **Nonconformity / Failure**
  non-fulfilment of a need or expectation that is stated, generally implied or obligatory.
- **Corrective action (in this report also referred to as corrective maintenance)**
  action to eliminate the cause of a nonconformity and to prevent recurrence.
- **Preventive action (in this report also referred to as preventive maintenance)**
  action to eliminate the cause of a potential nonconformity or other undesirable potential situation.
- **Predictive action (in this report also referred to as predictive maintenance)**
  action to monitor the condition of an asset and predict the need for preventive action or corrective action.

1.5 Scope and limitations

In this research, we focus on the operation and maintenance departments only. Even though this department works together with other departments. We first focus on linking the process model to the existing AM data practices. Thereby identifying the inputs and outputs of various processes and linking them. In this research, a case study is used to learn from practice and apply theory. Through the case, the general model is expanded. For further expansion, other cases (e.g., other assets other asset owners) could be used, but these are left out of the scope of this research due to the size of the case used.

1.6 Research question

We discuss the research questions in this section to give structure to the report. The research consists of one main research question and five sub-research questions. For each sub-research question, several additional questions are discussed that help solve the sub-research questions, which in turn help solve the main research question.

To solve the problem of information management in the Arcadis reference model we look at data models, which are models showing relationships between data. Data architecture is also investigated which shows the relations between data, applications and processes. Lastly, we look at data analytics which investigates existing and new methods that can be used to analyze data and streamline the process described in the reference model. We try to improve the reference model through information management, that reduces the time spent in the process. The main research question in this thesis is:

How can the Arcadis reference model used for asset management projects be improved in terms of data models, architecture, and analytics?

A clear understanding of the current reference model is needed. Based on this information, the model can be further improved. Next to this, it is important to discuss basic concepts in the process models and their implications for the research. Therefore, the first sub-research question is:

How does the reference model compare to the asset management processes currently done at Arcadis?

> *What are aspects of a reference model?*
> *What reference model is adopted by Arcadis?*
> *What methodologies and tools are used in practice for asset management at Arcadis?*

For the second question in this research, we are focused on research from the literature. As literature can provide various insights and directions for improvement, it is important to do a literature review. In this literature review, we discuss best practices in data models and infrastructure AM, as well as data management methodologies. The second sub-research question is:

What is the current state of asset management and connected domains in literature?

> *What is the current state of asset management literature?*
> *What is useful literature about information management for asset management?*
> *What can we learn from the standards for the Arcadis reference model?*
> *What are some of the maintenance policies and tools useful for infrastructure AM?*

In the third part, we focus on the actions that can be taken to improve the reference model. For this, the results from the first two sub-research questions are combined with knowledge from a use case and the knowledge of the people working at Arcadis. The third question is on the methods that use static and dynamic data. The key points using static data are not executed frequently whilst the methods using dynamic data get new data frequently and therefore need to be executed frequently. We research dynamic data next to static data methods because there are potential benefits in predictive and performance-based maintenance than more traditional methods. The resulting sub-research question is:

How can the reference model be expanded to include a data model for information management containing the key points using static and dynamic data?

What can be learned from literature to link the reference model to information management?
How can information management best practices be implemented in the reference model?
What is the data model behind the static data concepts described in the reference model?
What is the data model behind the dynamic data methods described in the reference model?
What is the combined data model?

In the fourth part of this research, we focus on validating our data model and checking whether the model delivers benefits to the Arcadis reference model. As the key points in the static data model are well known within Arcadis the focus is on the lesser-known dynamic data models. These models are validated through examining a use case and sensitivity analyses on the available data.

Does the data model validly improve the reference model?

Is the data model validated based on the information needed for the static data methods?
Is the data model validated based on the information needed in performance analysis?
Is the data model validated based on the information needed in condition monitoring and prediction?

After all the answers to the sub-research questions are generated and discussed, the main research question can be answered.

2 Current project reference model

In this chapter, we discuss the reference model currently adopted by Arcadis. This is analyzed to answer the first research question.
How does the reference model compare to the asset management processes currently done at Arcadis?

To answer this question, we first explain the aspects of an asset management reference model. Then we discuss the theoretical reference model and the processes and methodologies used in practice. We then compare the reference model and the used processes.

2.1 Aspects of an asset management reference model

A *project reference model* is a descriptive model that is used in a company to guide projects in a standardized way. A *project reference model* can have various inputs and outputs. The outputs of a project based on the asset management reference model can differ per project, as one project might result in maintenance planning while another project can result in a system that gives advice on asset performance. A reference model contains multiple aspects, as described below.

First are the *actors* in a reference model. *Actors* are the parties that perform the actions as described in the reference model or are somehow related to an action in the reference model, such as providing input or checking the output. Arcadis identifies three roles that actors can take in the reference model, as described in Figure 3. Each of these actors must deal with the consequences of processes of the other actors and are therefore related to each other. This also follows most AM norms and literature, for example, the NEN NTA8120 norm in the Netherlands (NEN, 2020).

The first role is that of the *asset owner*; this party is responsible for identifying stakeholders and requirements. This role is also responsible for setting business values, objectives, and tolerance levels for risks and budgets. This role is taken by the actor, who has paid for the asset and is responsible for its operations and maintenance. In the Netherlands, this role is often taken by government parties (e.g., Rijkswaterstaat and Prorail), and the other roles are outsourced. The second role is that of the *asset manager*, who is responsible for determining what must be done to realize the objectives set by the *asset owner*. This involves planning, analyzing, and contracting for the plans. The role also agrees on the service level with the third role, the *service provider*. The *service provider* is responsible for executing the plans created by the asset manager. Together, these three roles can execute the steps in a reference model.

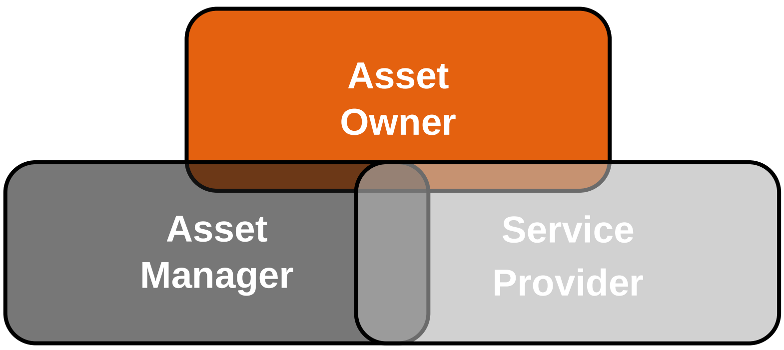


Figure 3 Roles in the asset management reference model

Arcadis uses the aspects of the model as shown in Figure 4, created by The Institute of Asset Management (2015a). In this model, there are various external influences, such as customers and legislation, that have an influence on asset management but are outside the scope of asset management.

The organizational strategic plan shapes AM strategy and planning. The Strategic Asset Management Plan (SAMP) is a long-term plan and can be the output of a project following the reference model. Asset management decision-making and the management of asset information are crucial for successful AM. A comprehensive project model should guide each life cycle stage, where both the asset manager and service provider need to coordinate accordingly. The organization and people aspect aligns culture, organization, and skills. Risk and review policies are the last part of the reference model and are involved with reviews on various aspects such as stakeholders, risks, and sustainability.

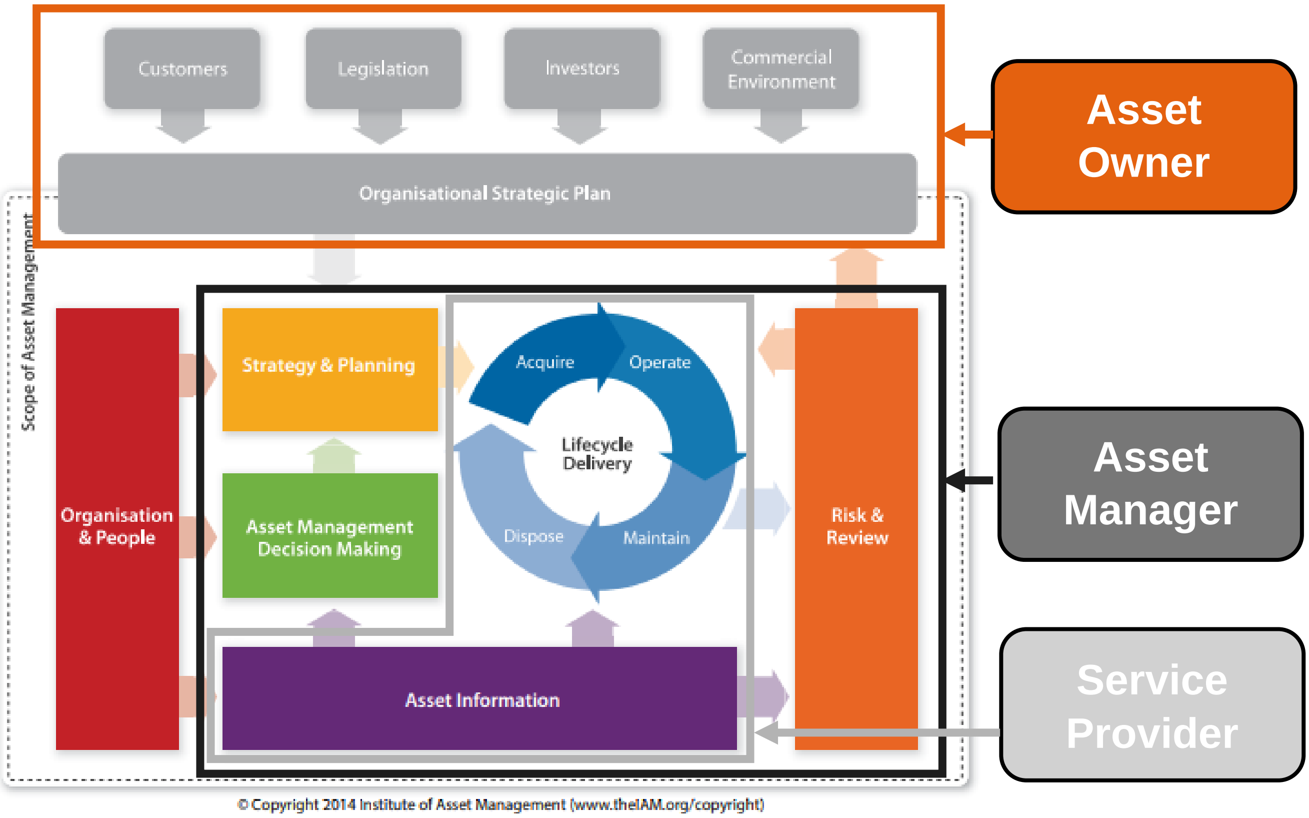


Figure 4 The IAM's conceptual asset management model from The Institute of Asset Management (2015a)

According to The Institute of Asset Management (2015b), the final output of a project reference model should be a clear approach to creating a validated asset management plan. This is documented information that specifies the activities, resources, and timescales required for an individual asset, or a grouping of assets, to achieve the organization’s AM objectives.

2.2 Reference model

In this section, the reference model that Arcadis adopts is discussed. This reference model was developed to give structure to AM projects. It uses insights from both the IAM conceptual model and the ISO 55000 norm. Arcadis also identifies the three key roles as described in Section 2.1. Most of the time, Arcadis fills the role of asset manager. Arcadis rarely fills the role of service provider; this is often outsourced to other parties by the asset owner.

2.2.1 Asset owner activities

Arcadis identifies two key activities for the asset owner. The first activity is scenario planning. This involves comparing possible future events and stakeholders. The second key activity is developing an asset management strategy. First, it is important to define the SAMP. This is followed by developing the business value framework. The goal of a business value framework is to identify the key values, assets, and activities of a business regarding asset management.

2.2.2 Asset manager activities

The asset manager is subject to five activities following the Arcadis reference model. The first activity is performance analysis. Here, the current performance of assets is analyzed based on acceptability. The next key activity for an asset manager is risk management, where risks associated with the business's values are identified and assessed. Arcadis defines risk as:

$$Risk = Consequence * probability$$

Consequences can be related to the business values described by the asset owner, and probability can be related to the occurrence of these risks. The next key activity for an asset manager is solution planning. Here, the goal is to generate solutions to mitigate the risks and optimize operational performance. The goal of the fourth step is portfolio optimization, which is to create the most company value per euro using the designed solutions over all the assets in the scope of the project. The last key activity for the asset manager is portfolio delivery management. This activity is all about preparing the portfolio of solutions for execution by the service provider.

The activities of the service provider role are not needed to understand the remainder of this research and are therefore left out of it. A complete overview of the project reference model is visible in Figure 5.

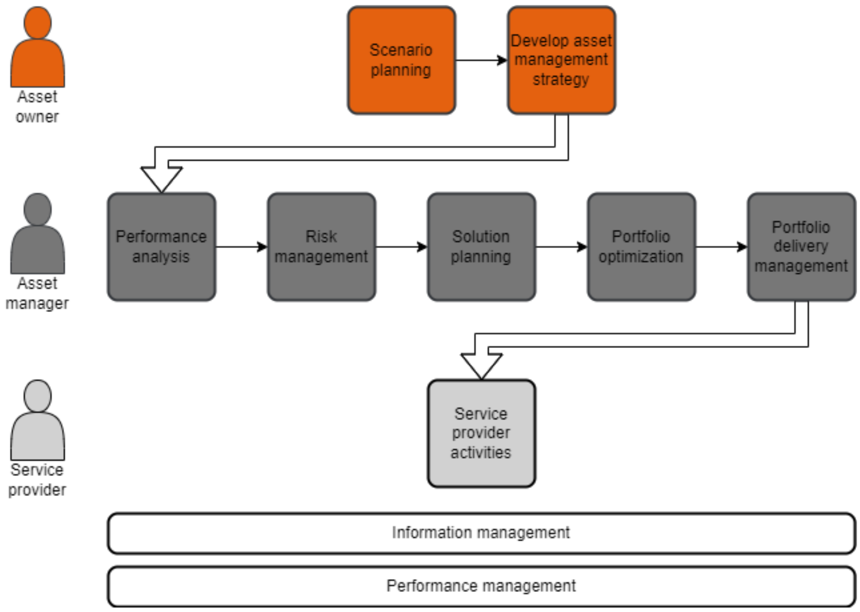


Figure 5 Reference model of Arcadis

2.3 Reference model in practice

There is a difference between the theoretical model and the activities used in practice. In this section, the results of surveys are used to identify the key methods in practice and deduce the differences between the model and real life. The setup of these surveys is first discussed. Then we discuss the information about respondents, followed by the most common methods used.

2.3.1 Setup of the surveys

To understand the information used and the methods used by Arcadis employees, a survey is used. The survey is sent to maintenance and data engineers of all operate and maintain departments; see also Figure 2. To make the survey time-efficient, two versions of the survey are created. The departments of AMIOI, AMRO, and AMRA received the survey with activities and methods related to maintenance. The ADM and Information Management (IM) departments received a survey on the methods used in data engineering. As all respondents were Dutch, the survey was also in Dutch. The exact questions for the surveys can be found in Appendix B.

For setting up the survey, we use the paper by Debell et al. (2021). They recommend grid-like questions. These are questions where a respondent answers with an option from a predetermined scale, for example, a scale from never to always. They also recommend branching for follow-up questions. Therefore, the survey uses a combination of a grid-like format, where if the respondent is deemed knowledgeable on the method, follow-up questions are provided using branching. Debell et al. (2021) also state that it is a good practice to first gather identifying data and introduce the purpose of the survey. Therefore, these questions and this information are added as well.

2.3.2 Respondent analysis

To give an overview, first, a respondent analysis is done. As stated, two surveys are created, one for data engineers and one for maintenance engineers. The number of respondents per department and per survey type is visible in Figure 6. The total number of respondents was 9 and 14 for the maintenance and data engineering surveys, respectively, totaling 23 responses.

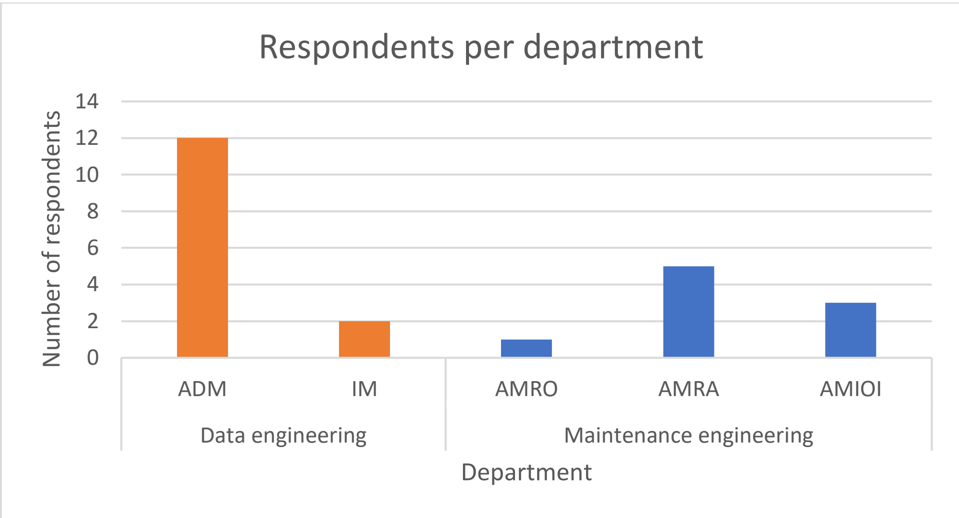


Figure 6 Respondents per department per survey type

As can be seen, most respondents come from the ADM department. This is also logical, as this is the biggest department of the considered departments, and the survey was sent to the greatest number within this department. In fact, the survey for maintenance engineering has a higher response rate than that for data engineering, with 90% and 67%, respectively.

Another interesting result from the survey can be found in the process steps of the reference model. All respondents were asked to identify which steps they were most often working on. The results are visible in Figure 7. Note that a respondent can operate in multiple process steps. In this figure, there are also percentages visible; these percentages are the ratio of respondents that acknowledge the step. For example, 67% of the AMIOI department acknowledges risk management; this means that 2 out of the 3 respondents in this department acknowledged this step.

The most prominent steps are performance analysis and information management. It is also worthy to note that all maintenance engineers acknowledged performance analysis as a key step, and almost all maintenance engineers also acknowledged risk management as an important step. From the percentages, we can see that performance analysis is also greatly supported by ADM, as 42% of the respondents acknowledged this compared to a lower percentage for other activities. Information management is mainly supported by members of the data engineering departments. Performance management is also often mentioned and is practiced by a variety of departments. Solution planning is less often practiced and is practiced by various departments. Portfolio optimization and portfolio delivery management are rarely acknowledged. Another conclusion from this graph is that the IM department is only focused on the supporting processes rather than the actual process steps.

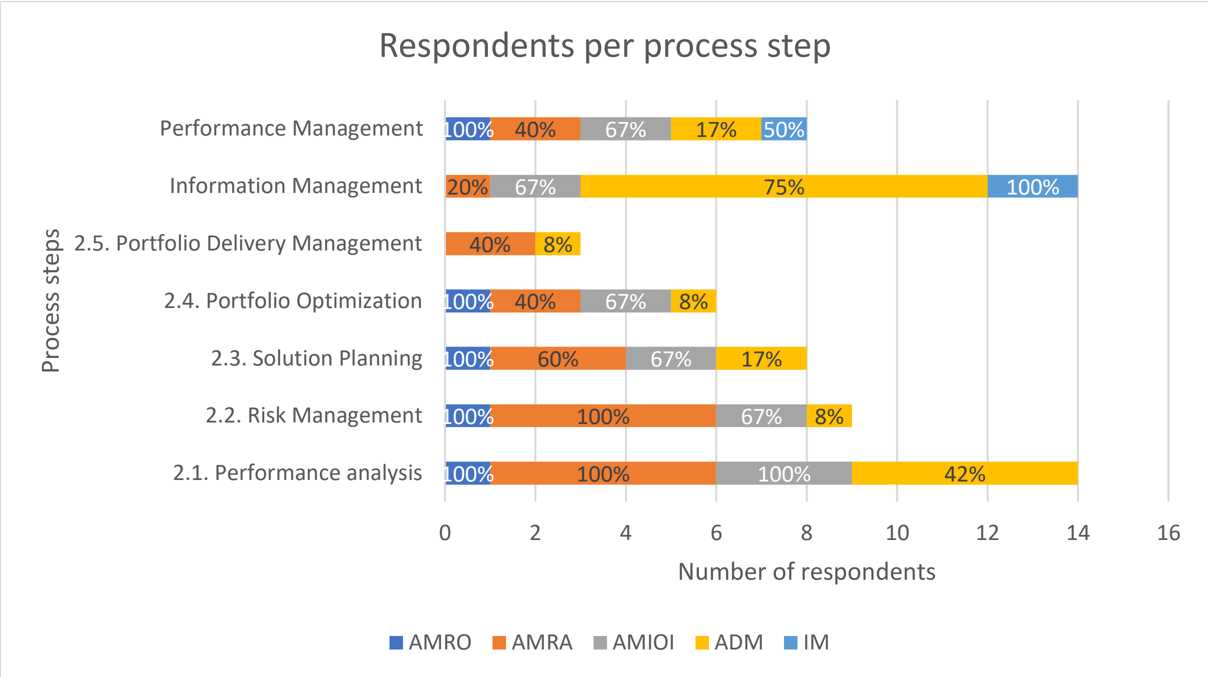


Figure 7 Respondents per process step, percentages represent the number of respondents per department which acknowledged the process step.

To conclude, the focus in the remainder of this research should be on the process steps performance analysis, risk management, solution planning, and the supporting process of information management.

2.3.3 Maintenance engineering

In this section we discuss the main methods used by the maintenance engineering departments. The maintenance engineers were asked to give a description of their usual project process. Some commonly mentioned steps are shown in Figure 8.

Figure 8 Steps mentioned by maintenance engineers

There are several methods mentioned in the reference model of Arcadis; these methods are given in Table 1. Accompanying the methods is a reference where the benefits and disadvantages of the methods are discussed.

Table 1 Maintenance methods

Maintenance method	Abbreviation	Reference
Root Cause Analysis	RCA	Tinga, 2013
Failure Mode, Effect and Criticality Analysis	FMECA	Tinga, 2013
Reliability Centered Maintenance	RCM	Braaksma, 2012
Fault Tree Analysis	FTA	Tinga, 2013
Event Tree Analysis	ETA	Khan & Abbasi, 1998
Pareto Charts	PC	Tinga, 2013
Bow Tie Analysis	BTA	Khan & Abbasi, 1998
Risk Register	RR	Tinga, 2013
Risk Tree	RT	Khan & Abbasi, 1998
RAMS(SHEEP)	RAMS	Rijkswaterstaat, 2010

For each method, the maintenance engineers were asked how often they used it. The complete results are visible in Appendix C but are also summarized in Figure 9. In this figure, the FMECA method is the most often used method; it is used in 59% of the projects, as indicated in the survey. It is followed by the groups of RR, RCA, PC, and RCM. The risk register method and the RCM method are closely related to the FMECA method; therefore, correlation is expected. Other methods are used less often and are not discussed further. FMECA is popular because of the clear structure of the method; both the client and the employees know what to expect when doing an FMECA. Next to this, FMECA has proven that it works, whereas other methods, such as ETA, do not consider risks that should have been considered.

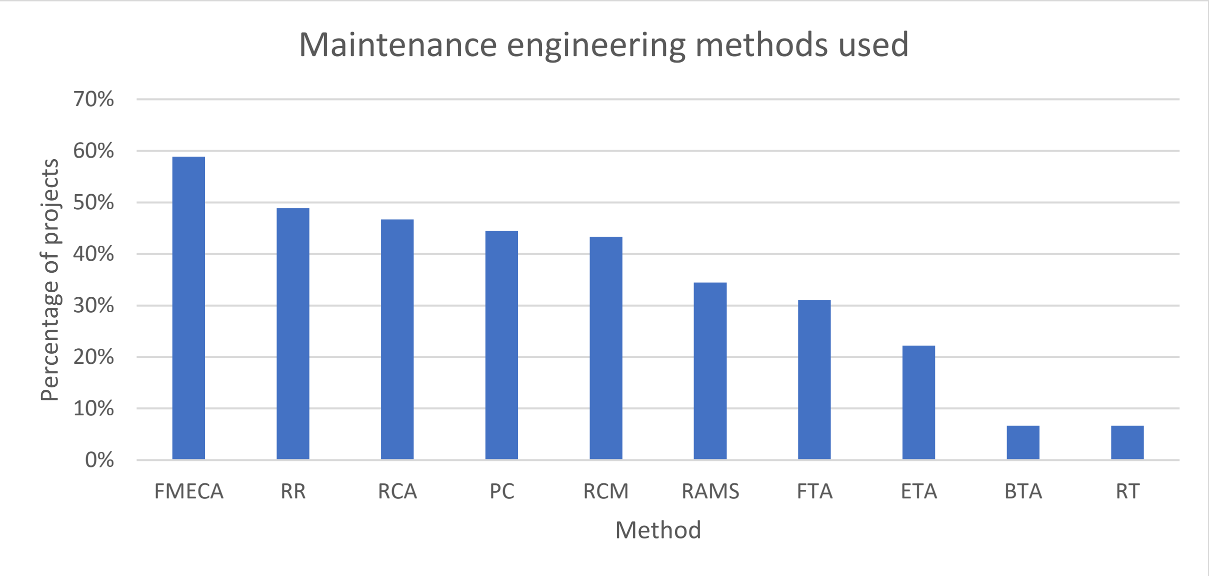


Figure 9 Use of maintenance engineering methods

In FMECA, the user tries to identify risks through failure modes and prioritize these risks. A more complete description of the method is given in the literature review in Section 3.4.2. Currently, FMECA is often done in spreadsheet programs like Excel, although more advanced tools are available, such as Clarify and RCMCost. The main input for the FMECA is the breakdown of the asset;

this can be a functional or system breakdown, also known as the asset register, as described by the Institute of Asset Management (2015a). We hereafter refer to this breakdown as the "asset register" to prevent further confusion with the concept of a system breaking down in the sense of failure. The failure modes of components of assets are delivered by the customer or the experts at Arcadis themselves. The resulting output is an overview of the most likely failures of the assets, which is often provided in a spreadsheet program as well.

In Figure 10, the use of the FMECA method, per Arcadis' reference model process step, is shown. This is the relative use, which means the percentages are normalized based on the number of respondents who acknowledged the process step. To illustrate, 2.2. Risk Management is used as an example. From Figure 7, we can see that 8 out of 9 maintenance engineers acknowledged the process step. In the survey, we asked the respondents to indicate how often, per 10 projects, the method was used. The total number of times the method was used is 47. We then divide this number by the maximum possible number of times, so $\frac{47}{(8*10)} \approx 59\%$. As can be seen, the method is most commonly used by the respondents working with portfolio optimization and performance management. Both are logical, as FMECA lays the basis for the optimization measures of the portfolio and is often a measure of the performance of the project. Third is the risk management step, which is also logical as this is the step at which the FMECA is made most often.

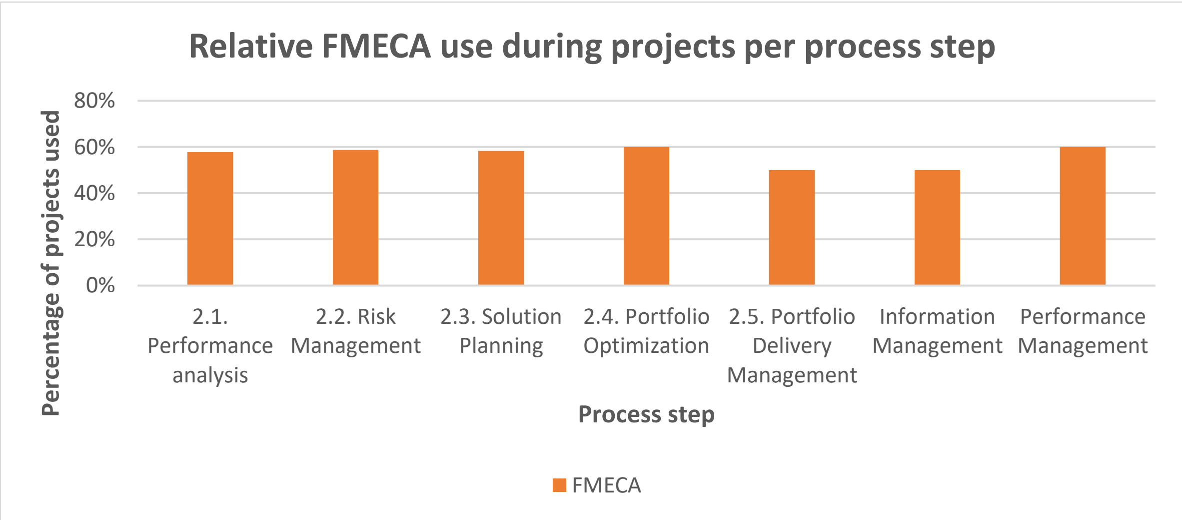


Figure 10 Use of FMECA in the process steps

The other methods are used less often and are therefore deemed less important. Some of the methods are discussed as alternatives to FMECA in Section 0. Additionally, an analysis of the use of the methods is provided in Appendix C. In the upcoming chapters, the focus is on FMECA.

We also analyze the survey results per process step. As can be seen in Figure 11, which contains the normalized numbers calculated like the example above. For the step performance analysis, FMECA is the most commonly used method. In general, the graphs of most individual process steps, follow the order of Figure 9. The analysis for the other steps is provided in Appendix C.

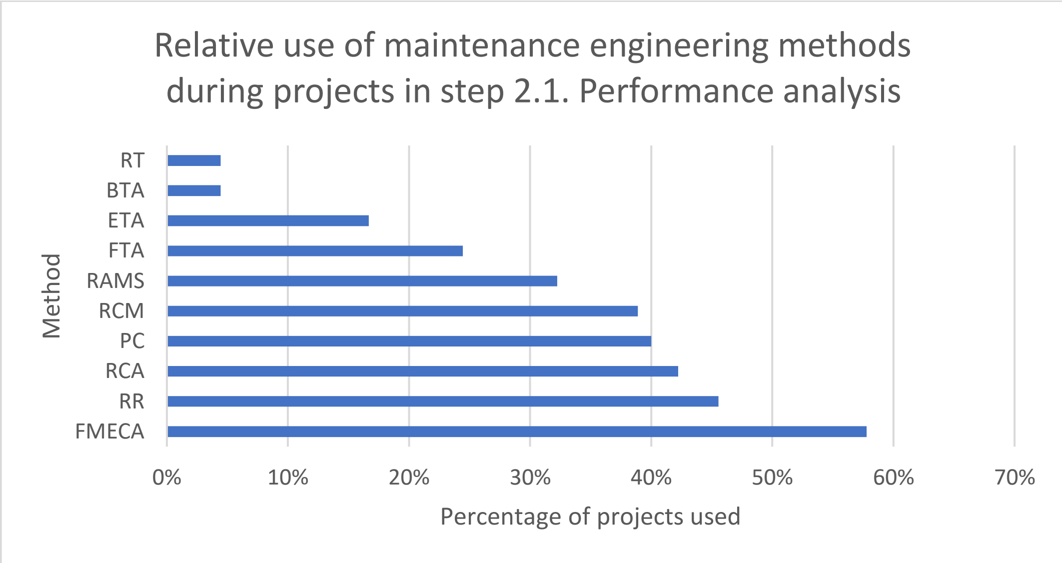


Figure 11 Relative use of the maintenance engineering methods in the process step 2.1. performance analysis

2.3.4 Data engineering

In this section, we add the data engineering methods to the maintenance engineering methods. Additionally, the closely connected concept of information engineering is also included.

There is a slight difference between "information" and "data." An explanation is given through the DIKW pyramid, shown in Figure 12. Data is a cluttered collection of elements and sources, while information is organized data placed in context. Information engineering is making data understandable, for example, by presenting the data to answer why, who, and where questions (Frické, 2009).

Figure 12 The DIKW pyramid, from DIKW Intelligence, 2023

Again, we discuss the main methods used, but now used by the data engineering departments. The data engineers were asked to give a description of their usual project process. Some commonly mentioned steps are shown in Figure 13. The process is very similar to that of the maintenance engineers.

Figure 13 Steps mentioned by maintenance engineers

There are also several data engineering methods; these methods are given in Table 2. Accompanying the methods is a reference where the benefits of the methods are discussed.

Table 2 Data engineering methods

Data engineering method	Abbreviation	Reference
GIS Data Acquisition and Preparation	GIS DAP	(Alvarado et al., 2022)
GIS Analysis and Trend Recognition	GIS ATR	(Alvarado et al., 2022)
GIS Visualization	GIS V	(Alvarado et al., 2022)
Data Integration and Structuring	DIS	(Euzenat & Shvaiko, 2013)
Time Series Analysis	TSA	(Ataspinar, 2018)
Natural Language Processing	NLP	(Einstein, 2018)
Computer Vision and Image Classification	CVIC	(Weber et al., 2022)
Data Processing and Visualization – without GIS	DPV	(Jensen et al., 2022)

For each method, the data engineers were asked how often they usually use it. The complete results are visible in Appendix C. The GIS methods are the most commonly used in projects and process steps. The results of this analysis are available in Appendix C.

In the remainder of this research, we should thus focus on FMECA and performance analysis for the maintenance engineers and focus on GIS and information management for the data engineering departments.

3 Literature review

In the previous chapter, we researched the reference model and usage in Arcadis, and the most popular methods regarding maintenance management of assets and data engineering. In this chapter, we conduct a literature review. The goal of this literature review is to learn about the various aspects of the problem. This part is included in this research to place it in the context of the connected domains. This goal is also translated into the second sub-research question.
What is the current state of asset management and connected domains in literature?

To answer this question, we follow the structure shown in Figure 14. In this figure, we specifically show the connected domains of AM. First, the domain of AM itself is discussed in Section 3.1. This domain is dependent on the information management domain, which is discussed in Section 0. AM is, like information management, also involved with a set of standards. The related sets of standards are reviewed in Section 0. In the last section, the maintenance domain is discussed. This domain enables asset management and is influenced by efficient information management.

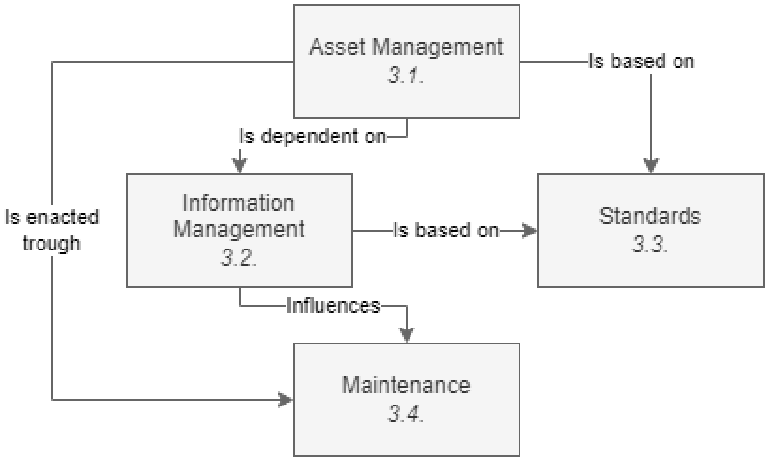


Figure 14 Structure and relations in the literature review

3.1 Asset Management

In this section, the current state of the literature on AM is discussed. As AM is a broad concept, the focus is on the literature about AM in infrastructure and specifically on process modeling. Maintenance is also excluded, as it is discussed in Section 0.

The term "asset management" is relatively new in literature. The term was first called "terotechnology," which originated in the 1970s. From the 1990s on, the term "asset management" took over. This can also be seen in Figure 15(White, 1975; Wijnia & de Croon, 2015).

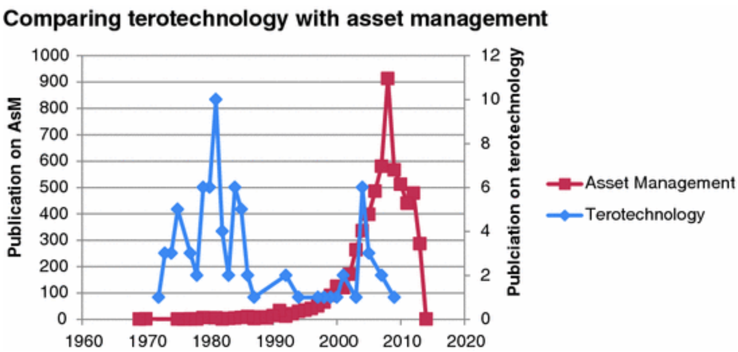


Figure 15 "Publications on asset management and terotechnology, according to Scopus (Query (TITLE-ABS-KEY("Asset Management")) versus (TITLE-ABS-KEY(terotechnology)) search date 20131220)." from Wijnia & de Croon, 2015, pp 448.

Over the years, the definition of AM has varied. In the 1990s, businesses were generally aware of the concept or some form of it. Because of this, asset management processes were developed and described, like in the International Infrastructure Management Manual and its forerunners (The Institute of Asset Management, 2002). A widely accepted definition emerged with the creation of PAS55 in 2004, which was then followed up by the ISO55000 family. Where first the focus was on physical assets, AM now also considers nonphysical assets, such as software products (Wijnia & de Croon, 2015). However, the focus of this research is on infrastructure; therefore, only physical asset management is considered.

At their core, infrastructures are networks of assets that enable connectivity between various locations. While individual assets play a crucial role in establishing these pathways, the network itself can be considered an asset (Aven, 2012). Given that multiple users often rely on these infrastructures, negotiations regarding terms of use can become complex, especially when numerous parties are involved. To streamline this process, it is often convenient to appoint an infrastructure manager who can balance the needs of all stakeholders and determine the terms of use (Herder & Wijnia, 2012).

This network of multiple parties makes AM for infrastructure different than AM for other assets. However, it is still important to manage these assets effectively, as the value of public infrastructure assets is estimated at around 400 billion euros in the Netherlands (Ruitenburg et al., 2017). AM in infrastructure is also characterized by the three key values of AM, as shown in Figure 16. It is important to find the balance between these three main values: performance, risk, and cost (Ruitenburg, 2017; The Institute of Asset Management, 2015a).

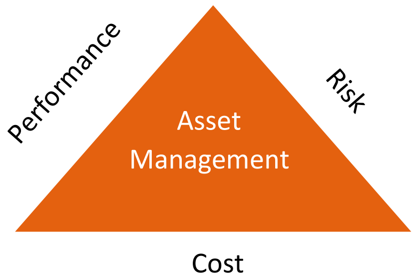


Figure 16 The values of AM

3.1.1 Performance

The first key value of asset management is performance. Performance was already introduced in the definitions in Section 1.4 as a part of AM. Performance is about the functioning of the asset in its operations. Performance can be analyzed from various perspectives. It is important to have the same perspective when analyzing assets; therefore, the two most common perspectives are discussed below (Ruitenburg et al., 2017).

The first perspective is a technical perspective, where the performance of an asset is determined based on its ability to fulfill the technical specifications. An example of this perspective is the concept of reliability. The most common indicator of this perspective is the Remaining Useful Life (RUL) (Pudney, 2010). The second perspective is the economic perspective. Here, the performance of assets is measured in terms of financial benefits and costs, including maintenance and operations. An example is the comparison between the upkeep and maintenance costs and the profits of the asset (Pudney, 2010).

It is important to note that an asset can have multiple (expected) moments of failure based on perspective. An example can, for example, fail from a technical perspective in 10 years but fail from an economical perspective in 8 years (Pudney, 2010; Ruitenburg et al., 2017).).

3.1.2 Risk

Like performance, we identify risk in Section 1.4. Arcadis defines risk as the probability of failure multiplied by the consequences of failure. Risk is therefore closely related to performance, as more risk will have an impact on the worsening performance of assets with a higher probability. Therefore, it is important to take risks into account when managing assets. The most appropriate methods for these are often quantified methods, so risk should be quantifiable (Davis, 2012).

Risk management is a crucial aspect of AM in infrastructure. It involves identifying potential risks, assessing their likelihood and impact, and developing strategies to mitigate them. (Herder & Wijnia, 2012). Risks are not only based on the assets but also on the context in which they are placed. This context can include not only other related assets in the asset system but also broader aspects like the environment and human behavior. This context is again closely related to perspectives on performance, which can also be used for perspectives on risks. It is important to take the context of assets into account very early in the development of an AM process and strategy (Pudney, 2010; Ruitenburg, 2017).

3.1.3 Cost

The third key aspect of AM is cost. While minimizing risks and increasing performance can in theory be done very efficiently, often high costs are associated with negating all the risks. Therefore, costs play a critical role in AM (Pudney, 2010).

There are various types of costs in AM. One of the most significant costs is capital expenditure, which refers to the costs associated with constructing, acquiring, and developing new infrastructure assets. Another significant cost in infrastructure AM is operational expenditure, which includes ongoing maintenance, repair, and replacement costs associated with keeping infrastructure assets in good working order. Both costs are significant for asset owners, and often lowering operational expenditure costs comes at a higher capital expenditure cost. Therefore, an optimal cost-effective balance should be found. Because of these costs, a tradeoff must be made between the tolerated risks and the expected performance (Davis, 2012; Wijnia & de Croon, 2015).

3.1.4 Processes

There are various process models for AM available in the literature, such as those described by The Institute of Asset Management (2015a) and Sherwin (2000). Various of these process models show five essential parts.

First, the strategy behind AM should be aligned with the strategy of the organization. The SAMP plays a central role in this key activity. Next, the process models show aspects of policies and decision-making. These steps are about maximizing the value of assets over their entire lifetime. As AM is carried out by humans, the third aspect is about human development. Fourth is the element of risk management, which incorporates the earlier discussed concepts into the models. The last key element is asset information and the management of this information. It is important to note that information technology is already seen as an enabler in the paper by Sherwin (2000) but is still seen as a problematic enabler in recent papers (Braaksma, 2012; Wu et al., 2021).

3.2 Information Management

Information management is a key aspect of AM; therefore, it is discussed further in this section. According to The Institute of Asset Management (2015b), organizations involved in AM should be careful with various aspects of information management regarding assets. The first aspect is collecting and acquiring information, as organizations should invest in information and data acquisition. The second aspect is receiving, using, and transforming this information. These steps involve information concepts like data quality and integrating data sources. It is important for organizations to keep track of not only technologies and software but also the people and knowledge available to execute these processes. Another aspect is maintaining information, which involves data storage and security (Wu et al., 2021).

Because information management is such a broad concept, we focus on the reference architecture and ontology below. Reference architecture is focused more on defining the structure and components of systems and processes, whereas ontologies are focused on defining information and relations.

3.2.1 Reference architecture

A reference architecture is a document or collection of documents that suggests how IT services and processes should be integrated to create a solution. The reference architecture represents widely acknowledged industry best practices. An easy-to-understand reference architecture directs the execution of sophisticated technological solutions and projects (Galster, 2015; Harmelink, 2022). Reference architectures can be very helpful in information management for AM; they play a crucial role in determining the requirements for information and information management (The Institute of Asset Management, 2015b).

Reference architectures are useful tools for organizations where multiplicity, like multi-project, multi-site, and multi-departments, is present. Reference architecture can provide various benefits, such as taxonomies and lexicons, a central vision, and modularization. Communication between the various aspects is facilitated by the shared taxonomy and lexicon. The shared vision unifies the efforts and perspectives of numerous individuals and groups (Galster, 2015; Muller, 2008).

The main goals of reference models are standardization and facilitation. Standardization is used to make management systems and processes interoperable, whereas facilitation is more of a guide to make the creation of specific use cases easier (Angelov et al., 2009; Muller, 2008).

Angelov et al. (2009) describe five optimal types of reference architecture that are successful in practice. These architectures are described in three dimensions. The first dimension of criteria is the goal dimension, in which one criterion is available. G1 asks the question, “Why it is defined?” about the reference architecture. This can be either *standardization* or *facilitation* of the design of concrete architectures.

The second dimension is the context dimension, which contains three criteria. C1 asks the question, “Where will it be used?”. This question has the answers *single organization* or *multiple organizations*. C2 asks the question “Who defines it?”. This question can be answered by various answers, dependent on C1, which can either be organizations: *software organizations, user organizations*, *research centers*, and *standardization organizations* or individuals in a single organization: *software designers*, *software users*, *software researchers*, *software/project managers*. The third criteria, C3, asks the question, “When is it defined?”. The answers to this question are *preliminary* and *classical.*

So, the reference architecture is made before or after the concrete architecture has already been implemented, respectively.

The third dimension is the design dimension, which contains four criteria. D1 is the first criteria, which corresponds to the question “What is described?”. The corresponding answers are *components, interfaces, protocols, algorithms,* and *polices and guidelines*. D2 is the second criteria corresponding with “How detailed is it described?”. The possible answers are *detailed, semi-detailed,* and *aggregated.* D3 is concerned with the question “How concrete is it described?”. The possible answers for this question are *concrete, semi-concrete* and *abstract.* The last criterion in this framework is D4, with the corresponding question “How is it represented?”. The corresponding answers are *informal, semi-formal,* and *formal.*

The five types of reference architecture proven to be successful according to Angelov et al. (2009) are:

- **Type 1.** This type of reference architecture is created for standardization within multiple organizations. The focus is on a more abstract model.
- **Type 2.** The second type of reference architecture is created for the purpose of standardization in a single organization. Again, a classical approach is used. However, the focus is now on a more practical model.
- **Type 3.** This reference architecture is created for facilitation purposes. Again, this is for multiple organizations and a classical approach. However, this time on a detailed level.
- **Type 4.** This reference architecture is also created for facilitation purposes, but for a single organization. Again, a classical approach and a detailed level are needed.
- **Type 5.** The last type of reference architecture is facilitation for multiple organizations. But now a preliminary approach is used, which means the models are developed before the services have been executed. The focus is again abstract.

There are various details and methodologies for constructing and editing a reference architecture, such as the TOGAF model. However, not one method is deemed perfect in literature; therefore, a best practice should be adopted when (re)creating reference architecture (The Institute of Asset Management, 2015b). For this, the five types described by Angelov et al. (2009) can be used as a starting point. This type can then be combined with a suitable modeling tool or language like the Unified Modeling Language 2.0 (UML) (Harmelink, 2022).

3.2.2 Ontology

A definition of ontologies can be found in ISO 19150: "a formal and explicit specification of a common conceptualization used to organize and describe knowledge is called an ontology" (International Organization for Standardization, 2012). It refers to a formal definition of a conceptualization that outlines the categories, traits, and connections that exist within a specific field of knowledge. With the help of ontology, knowledge can be expressed in a structured, methodical manner that allows for machine-readable interpretation and information sharing (Kovalev et al., 2018).

Ontologies come in a variety of forms, such as domain ontologies, task ontologies, application ontologies, and upper ontologies. A particular field of knowledge, such as medicine or finance, has ideas, relationships, and limitations that are described by a domain ontology. The concepts and connections that are pertinent to a particular job, like diagnosis or planning, are described by task ontologies. The ideas and connections that are pertinent to a particular application, like a supply chain management system, are described by an application ontology. A general framework for describing ideas and connections that are shared by several different categories, like time, space, and causality, is provided by upper-level ontology (Kokla & Guilbert, 2020).

A related concept is that of the Object Type Library (OTL). This is a collection of pre-defined object types that stand for a particular class of objects or data structures. These object types make it easier to create and handle complex data structures in programming languages and software development environments (Luiten et al., 2019). Compare this to a builder who uses blueprints to create a structure; OTLs can be thought of as instructions for creating objects and models of objects. OTLs specify an object's characteristics, methods, and events, as well as how it interacts with other objects. In the infrastructure domain, OTLs have also been playing a key role in the decomposition of assets (Beetz & Borrmann, 2018; The Institute of Asset Management, 2015b).

A well-constructed OTL is interoperable with other OTLs and applications. There have been many OTLs developed in the Dutch infrastructure sector, such as the CB-NL and the Imbor, but also OTLs by the ministries, like the Rijkswaterstaat-OTL as adapted by the Dutch national road authority. Overall, OTLs are useful tools for information communication and standardization, also in the context of infrastructure (Luiten et al., 2019).

3.3 Standards

In this section, we discuss relevant standards. A standard is a set of documents that describe exactly how a process, product, or system should be designed. The discussed standards are either international or local Dutch standards. The standards considered cover various aspects but are all connected to the reference model and processes at Arcadis. An overview of various relevant standards is visible in Figure 17. There are standards in the domains of AM, risk management, building information modeling, geographic information systems, and information communication. In the remainder of this section, we discuss the most relevant standards.

3.3.1 Asset management standards

We discuss three related AM standards: ISO 55000, ISO 55001, and ISO 55002. There are other standards in this family of standards; these are, however, not relevant. ISO 55000 offers recommendations for AM. It is the follow-up standard for PAS55. The purpose of ISO 55000 is to help organizations maximize the value of their assets throughout their complete life cycles. Making informed choices about how to manage assets in a way that satisfies organizational goals and stakeholder requirements involves every stage from acquisition to disposal (International Organization for Standardization, 2014b).

ISO 550001, like ISO 55000, offers recommendations for AM; however, the focus is now on the system in which AM is executed. It is more on the software side, while ISO 55000 is on the process side. ISO 55002 is about the implementation of the two other standards in practice. It offers examples of how organizations can use the standards in practice and offers helpful guidance on how to understand and apply the standards. The advice aims to assist businesses in enhancing their asset management procedures and to make ISO 55001 adoption and certification easier (International Organization for Standardization, 2014a; International Organization for Standardization, 2018a).

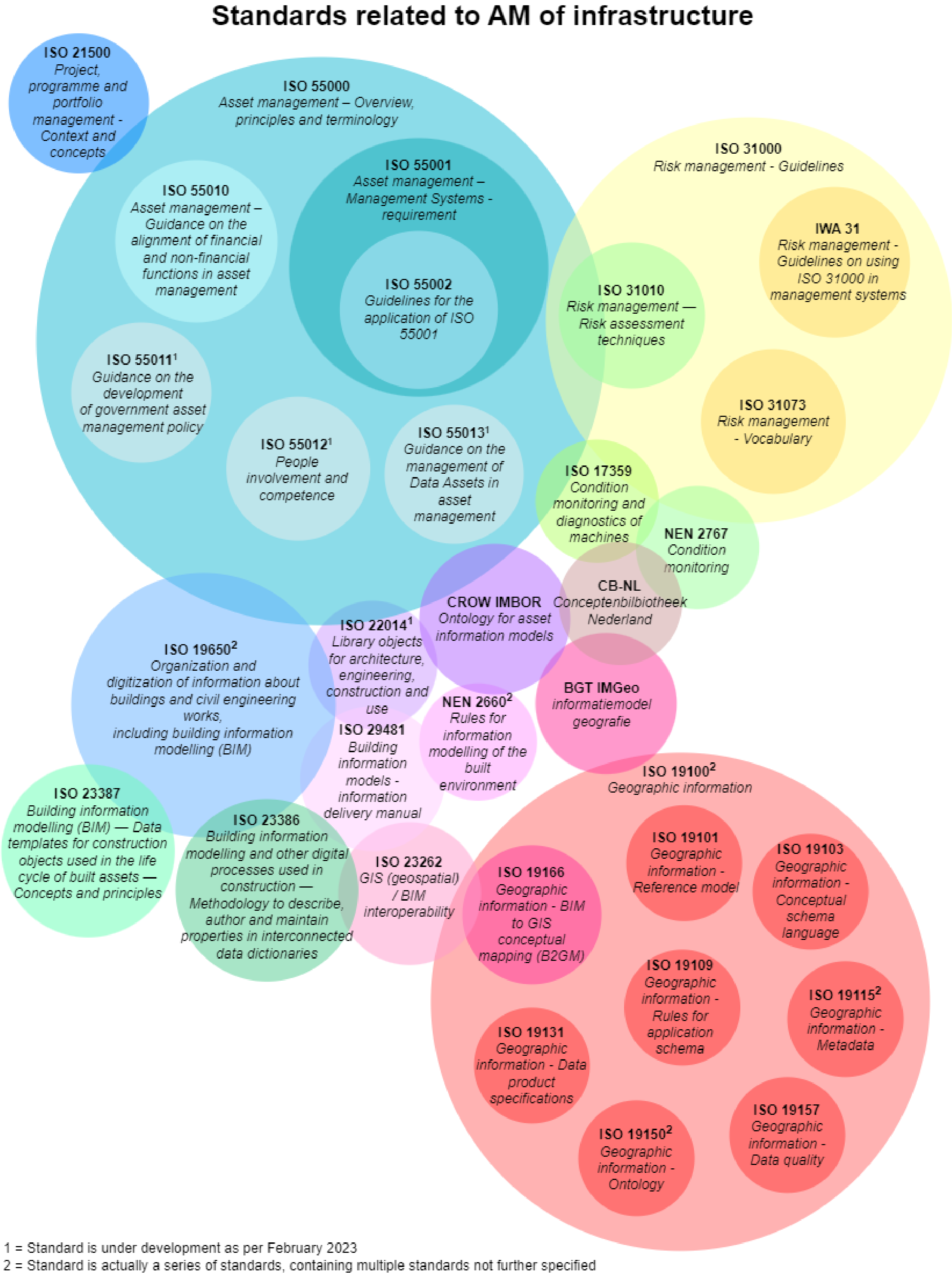


Figure 17 Standards related to Asset Management, Risk Management, Building Information Modelling and Geographical Information of infrastructure assets.

3.3.2 Risk management standards

A key aspect related to AM is risk management, as described in Section 3.1.2. A standard that offers advice on risk management is ISO 31000. The goal of ISO 31000 is to offer a structure and procedures for efficient risk management that can be used by businesses of all shapes and sizes engaged in all kinds of industries and endeavors. Figure 18 depicts the process proposed for risk management by the standard (International Organization for Standardization, 2018c).

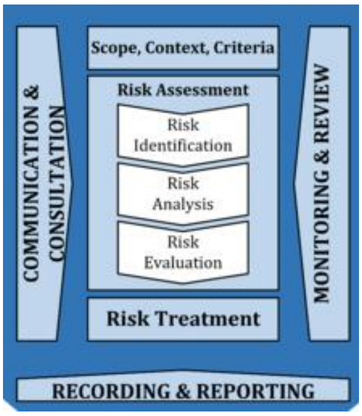


Figure 18 The process of risk management (International Organization for Standardization, 2018c)

Another relevant standard is ISO 31010, which offers advice on the choice and use of risk assessment techniques. It offers a model for evaluating risks using different methodologies, such as quantitative, semi-quantitative, and qualitative approaches. It emphasizes how crucial it is to choose the best method for a specific circumstance based on elements like the data's accessibility, the risk's complexity, and the degree of uncertainty present. Examples of the techniques mentioned are FMECA and reliability-centered maintenance (RCM) (International Organization for Standardization, 2019).

The Dutch standards body NEN created the NEN 2767 standard, which offers instructions for evaluating the state of structures and installations. Its main emphasis is on the use of visual examinations to evaluate the state of structures and installations. According to the standard, various components can be given a condition score based on how much they have deteriorated; higher scores signify more deterioration and the need for repair or replacement. The standard also offers recommendations for when it may be essential to supplement visual inspections with additional methods like non-destructive testing and laboratory analysis. Next to this, the standard also contains system breakdowns of various infrastructure assets, such as bridges. By providing a standardized approach to inspections and scoring, the standard can help identify areas in need of repair or replacement and support more effective maintenance and asset management practices (NEN, 2019).

It is related to ISO 17359, in which guidelines for the monitoring of conditions for machines are described. In the standards, the use of sensors and data for condition monitoring is described, as is the linking of data to failure modes (International Organization for Standardization, 2018b).

3.4 Maintenance management

AM on its own can be useful, but it is explicitly useful for creating maintenance schedules. Maintenance can put the plans and principles created at AM into practice. Maintenance is seen as the actions to reduce the number of failures or the mitigation of the consequences of failures (Sherwin, 2000). This section is split up into types of maintenance and methods useful for maintenance management.

3.4.1 Types of maintenance

According to NASA (2008), there are three main types of maintenance. These have also already been briefly introduced in Section 1.4. The three types are shown in Figure 19.

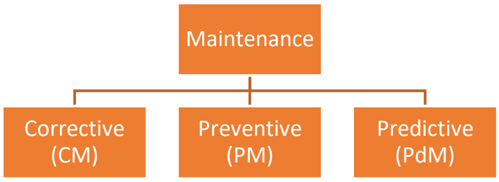


Figure 19 Types of maintenance

CM is also known as reactive, repair, fix-when-fail, or run-to-failure maintenance. When we use this maintenance type, equipment repair or replacement only takes place when a failure has occurred. Delays can occur if the object breaks down and the repair parts are unavailable, increasing costs. The benefits of CM are the absence of inspection costs (NASA, 2008; Tinga, 2013).

PM entails the routine inspection, alterations, cleaning, lubrication, replacement of parts, calibration, and repairs of machinery and components. To decrease equipment failure, PM plans routine inspection and maintenance at predetermined intervals (time, working hours, or cycles). The focus of PM is on failure rates and the intervals between failures. It is assumed that a component can be replaced before it is expected to fail and that these variables can be statistically determined. The foundation of PM is the idea that overhauling machinery through disassembly and the replacement of worn components safely returns the machine to like-new condition. The idea behind this renewal job is that new components have a lower failure rate than older ones of the same design. Benefits are the decreased number of failures; drawbacks are the increased costs of inspections and other activities (NASA, 2008; Tinga, 2013).

PdM is the third type of maintenance; it is also known as predictive testing and inspection or condition-based maintenance. It mainly evaluates the condition of machinery using nonintrusive testing methods (like sensors), visual inspection, and performance data. Instead of performing maintenance at random intervals, PdM schedules repairs only as needed by the condition of the machinery. Planning and scheduling maintenance or repairs in advance of failure is made possible by the ongoing analysis of data from the monitoring system. To assess the equipment's state and spot potential failure signs, various methods like trend analysis and statistical measures are used. The methods used can be simple, like regression, but they can also involve complex machine learning models, like neural networks. The choice of methods should be based on the complexity of the asset and the importance or budget available for the asset (NASA, 2008; Tinga, 2013). Benefits include a further decrease in failures and possibly maintenance actions; however, this comes at an increased price for monitoring tools.

3.4.2 Maintenance strategies and methods

In this section, we discuss the most popular maintenance strategies and methods following Tinga (2013). The first methodology discussed is reliability-centered maintenance (RCM). In literature, it is seen as an important methodology within the maintenance domain due to its extensive use in various branches such as aerospace engineering and the energy sector. RCM was created over a thirty-year span; its inception can be traced to a report ordered by the US Department of Defense that discussed the use of RCM in the civil aviation sector. The use of RCM can have a big impact on operational costs because it serves as the foundation for preventive maintenance actions (Braaksma, 2012).

A key part of RCM is the Failure Mode and Effect Analysis (FMEA). All potential system failures are listed in the FMEA, along with a description of the financial, safety, and practical ramifications of each failure. Since the analysis begins with potential component failures and then determines what the repercussions (on the higher system level) are, an FMEA is an inductive or bottom-up method. A team of individuals with diverse backgrounds typically works together to complete an FMEA. It is more likely that all potential failures are recognized and their effects are accurately estimated if the team includes experience from design, operation, maintenance, and finance. The relationship between RCM and FMEA can be seen in Figure 20 (Braaksma, 2012; NASA, 2008; Tinga, 2013).

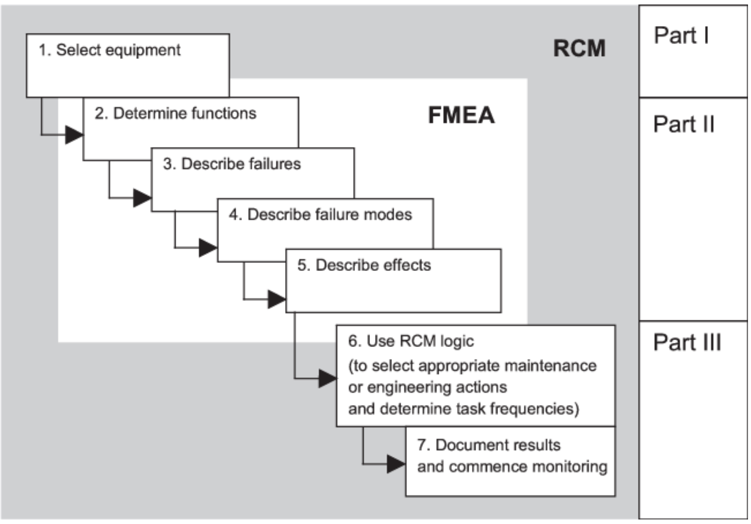


Figure 20 Relationship between RCM and FMEA, from Braaksma (2012)

An extension to FMEA is Failure Mode, Effect, and Criticality Analysis (FMECA). In this analysis, criticality is added as an important part of a failure. Thereby connecting FMEA to risk management. The method looks at three key aspects in determining failures:

- Severity / criticality (S): indicates how large the consequences of a failure are.
- Occurrence (O): indicates how often the failure is expected to happen.
- Detection (D): indicates how easy the failure is to detect.

In a regular FMECA, these attributes each get a score per failure on a scale of 1 to 10, where 1 represents the most preferable effect and 10 the least preferable effect. Often, there are preset descriptions available for the scales. These scores are then combined into the risk priority number (RPN):

$$RPN = S * O * D$$

Failures with a high priority number should be prioritized in further maintenance and mitigation plans, as these failures pose a great threat. These maintenance mitigation plans are not part of FMECA but are part of RCM (Braaksma, 2012; NASA, 2008; Tinga, 2013).

Other common approaches are fault tree analysis (FTA), root cause analysis (RCA), and pareto analysis (PA). FTA is a logical or top-down approach, in contrast to FMECA. All potential underlying events and failures of subsystems or components are found starting from a system failure, which is referred to as the "top event." The outcome is a set of fundamental events that may have contributed to the top event's happening (Tinga, 2013).

Whereas FMECA and FTA can be done before a failure occurs, RCA is done after a failure occurs. Although RCA is not formally a method with a clear definition, it truly encompasses all structured approaches looking for the root cause of a failure, accident, or event. The fundamental tenet of the

RCA is that problems can only be solved by addressing their underlying causes as opposed to their symptoms to prevent recurrence (Tinga, 2013).

For complex systems, PA can be used to prioritize these improvement attempts. Complex systems typically exhibit a wide variety of failures, though not all these failures are similarly detrimental to the system's ability to function. To eliminate the most significant failures, a structured approach is offered by PA. It is founded on the finding that 20% of failures account for 80% of maintenance expenses, or 80% of the system's overall downtime. Therefore, to improve the system, it should focus on the top 20% of failures since fixing them results in a substantial decrease in costs or an increase in uptime.

4 Setting up the data model

From the context analysis and the literature review, we can conclude that a connection between data engineering and maintenance engineering needs to be made. This connection should be made to serve as an interface between the processes and the underlying data used. This can help complete the reference model and help understand the connections between processes. The reference model contains methods that use static data, such as FMECA, but also methods that use dynamic data, such as performance analysis. In this chapter, we first focus on the static data which is followed by the dynamic data, thereby answering the third sub-research question.
How can the reference model be expanded to include a data model for information management containing the key points using static and dynamic data?

As a first step to creating the data models, we apply the lessons learned from Angelov et al. (2009) by comparing a working reference architecture to the current Arcadis reference model, thereby trying to identify the missing elements. Following the lessons learned by Wu et al. (2021), another key step towards successful AM is to develop a unified asset classification model. Therefore, in the second part, we focus on identifying the types of asset classification and asset information that are present. In the third part, we further discuss the model currently in development at Arcadis for asset classification: an Object Type Library (OTL). Additionally, we introduce a case to validate the methods and data in practice. In the last parts of this chapter, we create the data models for the OTL, the GIS methods, and the FMECA method and the data models for performance analysis and condition monitoring and prediction.

4.1 Reference architecture type

To further develop the reference model, we determine the reference architecture type of the model based on the criteria of Angelov et al. (2009) see Section 3.2.1. The three dimensions containing criteria are also described in Section 3.2.1. For further explanation about the dimensions or criteria, the paper by Angelov et al. (2009) should be used.

The values for the criteria for the reference model for Arcadis are shown in Table 3. As the model is created for the purpose of streamlining projects by making project steps and methodologies compatible, it is a *standardization* model for question Goal 1 (G1).

The model is only implemented at Arcadis; it is seen as a *single organization* for question Context 1 (C1). The reference architecture is not necessarily oriented toward creating software products but rather toward streamlining projects. The model is created by these *managers* with the input of some *users* involved in the process for question Context 2 (C2). The model was made after several projects have been executed this way and is therefore *classical*, as is an option for question Context 3 (C3).

The model contains *components,* as several key activities have been worked out, and contains an order and advice on how to do these activities, thereby providing *policies and guidelines*. Some *interfaces* are present through communication with the customer, which is important. But as *interfaces* between steps are not defined, this part is absent. This creates the answer to the question of dimension Design 1 (D1) The model is split into two levels: an *aggregated* level and a *semi-detailed* level. The level is not yet detailed because exact methods to use have not been written down. The model uses *semi-concrete* levels of abstraction as the methodologies are left open, although some guidelines for the methodologies are provided. Lastly, the model is written *semi-formally* as no formal modeling language is used, although the model is structured throughout the entire model. Providing the answers to Design 2, 3 and 4 respectively.

Table 3 Reference architecture types of Arcadis and type 2 architecture from Angelov et al. (2009)

Dimension	Question	Values Arcadis reference model	Values Type 2 from Angelov et al. (2009)
Goal 1 (G1)	Why?	*Standardization*	*Standardization*
Context 1 (C1)	Where?	*Single organization*	*Single organization*
Context 2 (C2)	Who?	*Users and managers from the organization*	*Software users, designers and managers from the organization*
Context 3 (C3)	When?	*Classical*	*Classical*
Design 1 (D1)	What?	*Components, policies and guidelines*	*Components, interfaces, policies and guidelines*
Design 2 (D2)	How detailed?	*Aggregated, semi-detailed*	*Aggregated, semi-detailed, detailed*
Design 3 (D3)	How abstract?	*Semi-concrete*	*Semi-concrete, concrete*
Design 4 (D4)	How formal?	*Semi-formal*	*Semi-formal*

The type 2 reference architecture values of Angelov et al. (2009) are also shown in Table 3. There is often an exact match between the type 2 and the Arcadis models. There are, however, some exceptions. The first exception is at C2, where the main difference is the word "*software*." The Arcadis reference model is not necessarily oriented toward developing software; however, the users of the model do use software. Additionally, the *designer* is absent, as this role is filled by the managers. A second exception can be found at D1, where *interfaces* are absent in the Arcadis reference model. According to Angelov et al. (2009), *interfaces* are the connections between components of the reference model and the connections between reference models. They can take many forms, such as data models and user interfaces. As already mentioned, *interfaces* are present for the customer but absent between applications. The other exceptions are in D2 and D3, where *detailed* and *concrete* are absent, respectively. Both terms are optional in the type 2 reference architecture and would enrich the model. Therefore, this is a slight deviation from the type 2 model.

According to Angelov et al. (2009), the reference architecture has a greater chance of succeeding if it matches one of the five types of reference architecture, based on successful and less successful reference models. Each of these types has set answers to the questions above. Currently, the Arcadis model matches closer to the type 4 architecture, which is oriented on *facilitation* rather than *standardization*, to further decrease the time spent in the processes, and to make this easier to do. Arcadis is trying to move from *facilitation* to *standardization*, so this is logical. Arcadis should expand the model with *interfaces*; additionally, *software engineers* might need to be included for the successful design of a standardization reference model.

4.2 Asset classification and information

Now that we know interfaces are needed and this can be done through a data model, we need to determine the type of data that should be put in the model. In this section, we discuss asset classification, also known as asset configuration. Asset information and the related asset information exchange are also discussed. There is a difference between the classification and information of assets. Classification of assets is about organizing the asset into a type or class, whereas asset information is the information specific to the asset.

There are various classification methods for infrastructure assets, but commonly used are classifications based on groups and types, often grouped by function, or secondly, classifications based on the decomposition of an asset. The first type of classification is known as a taxonomy, whereas the second type of classification is known as a decomposition or asset register. The relations between classes in a taxonomy are often related to each other using the parent-child relationship, whereas the register uses a part-of relationship. A parent-child relationship means the child has the same attributes as the parent but differentiates from other children on a new attribute.

Asset information is a broader concept; it is the information needed to effectively operate the asset. There are various types of information that can be used for infrastructure assets; these are shown in Table 4 (Wu et al., 2021).

Table 4 Asset information data types (Wu et al., 2021)

Data type	Related question	Example(s)
Physical asset data	What assets are owned/operated and what are their technical characteristics?	Asset name, building year
Location and spatial links	Where is the asset and how does it relate to other assets?	Coordinates, closeness, system data
Performance data	How does this asset contribute to serviceability targets?	RAMS, performance perspectives from Section 3.1.1
Condition data	What is the current (estimated) state or condition of and around the asset?	Vibration, sound, wind speed
Work management data	What work has been / will be performed on this asset?	Date, time, type of maintenance activities
Failure data	When and with what criticality did failures occur?	Date, time, duration and cause of failure
Cost data	How much does the asset cost to buy and operate?	Costs of maintenance and upkeep, costs of failure, purchasing costs

As can be seen from the table, asset information is a broad concept involved in many disciplines. It is also good to note that there is a large distinction between dynamic and static data. Where static data types do not change constantly over time, dynamic data does change over time. It does, however, not mean static data can never change; however, this is often under specific circumstances such as relocating, expansion, or organizational changes. Examples of static data are the failure modes of an asset, which will not change over time as most failure modes are known beforehand. An example of dynamic data is performance data such as the opening duration of a moveable bridge, which indicates how often traffic can flow freely or not. This changes with each passing moment and is therefore dynamic.

The frequency, or number of changes in data over a period, might vary per data type. A high frequency means the data changes often, whereas a low frequency means the data changes less often. High-frequency data is therefore often larger in size, and the times at which it is received are often also greater. So high-frequency dynamic data types require better data management systems. Additionally, data types are also largely influenced by the source of the data. The source of data can vary per specific asset; however, some typical sources are available from experience at Arcadis. The SCADA and EAM systems are common sources of dynamic data. Supervisory control and data acquisition (SCADA) systems are software applications for operating assets and logging operating actions such as operator commands. Enterprise asset management (EAM) systems are integrated systems that handle logistics and maintenance planning and connect these to SCADA systems. The dynamicity, frequency, and typical source are shown in Table 5.

Table 5 Dynamicity, frequency and typical source of data types

Data type	**Dynamicity**	**Frequency**	**Typical source**
Physical asset data	Static	Once	Building plans, Asset register
Location and spatial links	Static	Once	GIS
Performance data	Dynamic	High	SCADA systems
Condition data	Dynamic	High	SCADA systems, sensors, inspections
Work management data	Dynamic	Low	EAM
Failure data	Dynamic	Low	Log systems
Cost data	Dynamic	Low	EAM, financial records

The data type "location and spatial links" is especially highlighted in infrastructure assets by Wu et al. (2021). It is highlighted because there is a complete software system that allows for the linking and use of geospatial data; this is the earlier explained GIS. These systems are useful for analyzing networks of assets and the relations between assets and asset components.

Another aspect that is also a development action at Arcadis is the exchange of information. In the standards, this information exchange is also represented by the libraries IMBOR, CB-NL, and BGT IMGeo. Information exchange is a major focus of the development of the OTL at Arcadis. However, it is not in the scope of this research, so we do not explain this further.

The three aspects of information management—asset classification, asset information, and information exchange—are shown in Figure 21 (Wu et al., 2021). In this figure, it is shown that all three aspects are related to each other and to the asset itself. As the asset is the main source of information, the links between the asset and information aspects are logical. The links between information exchange and the other aspects are explained by the fact that both asset information and asset classification need to be communicated between stakeholders, especially since the infrastructure domain is project-based rather than process based. Lastly, the relation between asset classification and asset information can be explained by the fact that most of the information is related to a specific class of assets, whereas for other classes of assets, this information might not be relevant. Additionally, information can be specified for class levels.

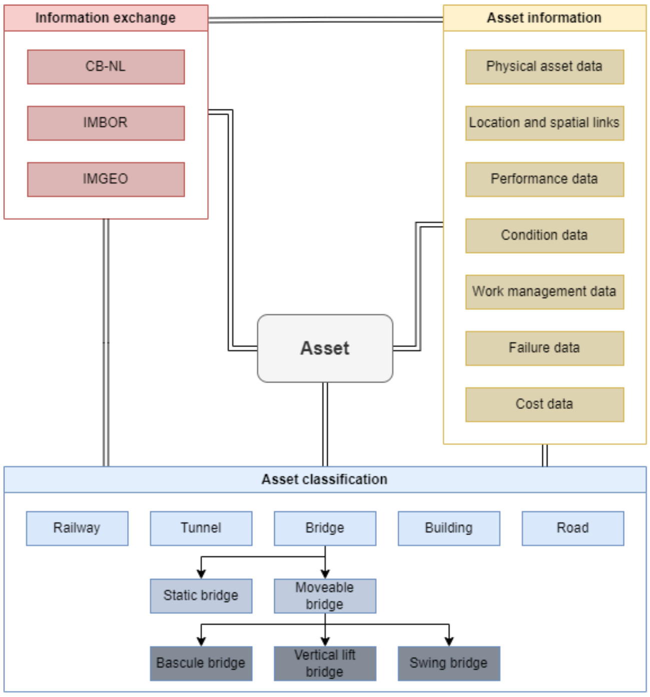


Figure 21 High level overview information management of an asset

4.3 Reference model, OTL and use case

To create the data model based on the research by Angelov et al. (2009), we identified asset classification as an important data type. Asset classification is already being standardized within Arcadis in the form of an Object Type Library (OTL). To further develop the model, we investigate the available OTL created by Arcadis and prepare this OTL for implementation with the reference model. Currently, not all infrastructure types are available in the OTL. However, the bridge is available, and as there is a movable bridge available for analysis in the use case, we focus on the OTL for bridges. The bridge in the use case shall not be made publicly available. First, we explain the basic concepts surrounding bridge infrastructure. We follow this up with an analysis of the current OTL. We then discuss the concept of a movable bridge. Lastly, we compare the OTL to the movable bridge in the fourth section, and we compare the OTL to AM practices in the fifth section.

4.3.1 An introduction to bridges

Bridges are a common sight in infrastructure. Bridges help infrastructure like roads and rails cross an area without obstructing the traversability of the area underneath. The bridge considered in this research crosses a waterway while being connected to a roadway.

Bridges are complex civil infrastructures consisting of various parts. Various bridges in the Netherlands have been built from the 1950s to the 1980s. These bridges are now of substantial age, causing various AM dilemmas in different parts of the bridges. Most bridges consist of a superstructure and a substructure. The superstructure consists of the components that actually span the obstacle the bridge is intended to cross. The substructure consists of all the parts that support the superstructure. An example of the superstructure is the deck of the road, while an example of the substructure is the concrete pillars supporting a bridge. Bridges can consist of various materials, such as concrete and steel, all of which have various failure modes, such as concrete rot and fatigue. There are various other components, such as the bridge deck, sidewalks, supporting arches, and drainage systems (Snijder & Hesselink, 2017). This is not discussed in further detail, as this is outside the scope of this research; the focus is on the mechanical components of a movable bridge.

One of the most common ways to classify a bridge is based on its movability. A movable bridge, unlike a static bridge, has a part of the bridge deck that can be opened, allowing floating vehicles of greater size to pass the bridge. Further information on movable bridges is given in Section 4.3.3.

4.3.2 OTL

As already stated, Arcadis is currently developing an OTL to stimulate standardization and understanding of concepts across the organization. As stated in Section 3.2.2, an OTL is a collection of pre-defined object types that stand for a particular class of objects or data structures. Currently, the OTL is being developed for the design and engineering departments of Arcadis. However, to fully standardize across domains and the whole company, other stages in the lifecycle of infrastructure should also be considered. Therefore, the OTL will be expanded to include the operate and maintain stages as well.

The OTL can be of great help in elevating the reference model. The OTL can fulfill the need for asset classification as mentioned in Section 4.2. Additionally, the OTL is also a reference architecture type; therefore, it would be wise to align these reference architectures to stimulate compatibility.

The existing OTL consists of three libraries. The first library is the taxonomy, the second is a decomposition, and the third library contains all the materials. The material library is a simple library containing various materials, such as asphalt and bricks. The materials contain definitions and several attributes like density, mass, and identification. The library is, however, not used in current processes and is still under development; therefore, it is not used in the remainder of this research. The other libraries are quite similar and are also linked to each other. However, to fully understand the libraries, first the aspects of both libraries are discussed.

First are the actual classes in the libraries. These classes form the libraries and are linked using relationships. Each class can also contain some attributes. Examples of attributes are identification, mass, length, RUL, and carbon emissions. These can be quantifiable attributes as well as qualitative ones. Lastly, the classes are often linked to the materials library, thereby consisting of a range of materials.

The aspects of an OTL are also shown in Table 6. As stated, classes in a taxonomy describe the type, whereas the decomposition consists of components. Also, the relationships differ. Both types of OTL contain attributes and relationships with the library of materials.

Table 6 OTL: taxonomy and decomposition

OTL aspect	Taxonomy	Decomposition
Class	Group / type	Components
Relationships between classes	Parent - child	Part-of
Attributes	✓	✓
Materials	✓	✓

In Figure 22, the taxonomy and decomposition coming from the Arcadis OTL are partially displayed. As can be seen, both OTLs contain the bridge class; these are also linked and are the same object. However, the relationships and subclasses are different. The taxonomy splits into bridge types, such as movable bridges and static bridges. whereas the decomposition delves deeper into the components of a bridge.

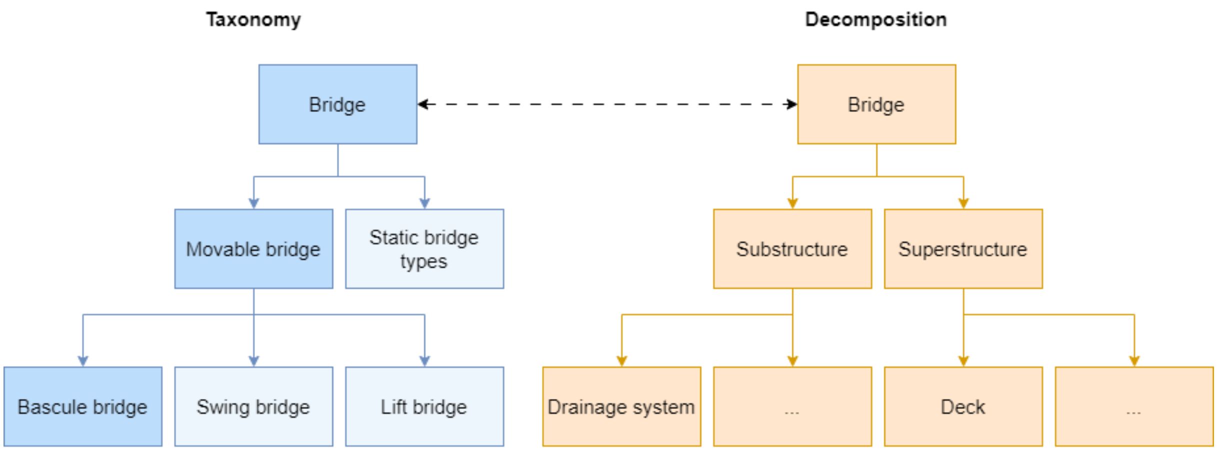


Figure 22 Taxonomy and decomposition of a bridge in the OTL

4.3.3 Movable bridge

A movable bridge in the taxonomy is a subclass of the bridge type; this means that a movable bridge contains all the aspects a bridge has and differs in some aspect from the other bridge types. The movable bridges differ in the fact that the bridge deck can be opened, allowing larger water vehicles to pass the structure.
There is also a differentiation in classes on the movable bridge, as can be seen in Figure 22. This differentiation is based on the method of opening the bridge deck. All three bridge types contain various parts that need to be lifted; this is often done by mechanical components and a power unit. As these power units are often the cause of failure, movable bridges tend to have more failures than static bridges. The bridge in the use case is analyzed based on the moving components rather than the static components of the bridge, as these tend to be the cause of bridge failures.

There are two main types of failures for the movable bridge. First, there is a failure in the function to let roadway travel pass. So, if the bridge is opened, roadways can't pass the bridge; however, this is not necessarily a failure, as opening can also be the desired effect. The second type of failure is if the bridge does not allow waterway traffic to pass. So, a failure can also occur when the bridge cannot be opened intentionally, thereby not allowing vessels to pass the bridge. It is possible that both failures occur at the same time, when, for example, safety mechanisms fail and travel underneath or on top of a bridge cannot be guaranteed to be safe.

The decomposition of bridges in the Netherlands is often done following the NEN 2767 standard, as mentioned in Section 3.2.2. In this standard, an asset is split into three levels, from highest to lowest: asset, element, building, or installation component. The relationships between these levels are part-of. The asset movable bridge consists of various elements. These elements can come from a static bridge or are elements specified for movable bridges. For example, a bascule bridge has the element bridge deck from the bridge but also the element power unit for moving the leaf. An example of a breakdown following the guidelines of NEN 2767 is shown in Figure 23.

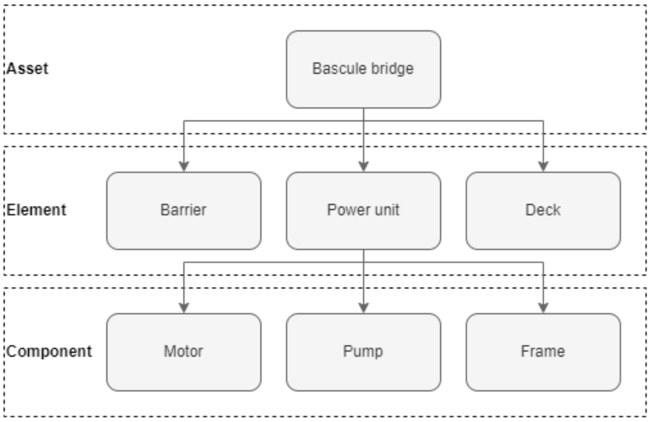


Figure 23 Breakdown of bascule bridge following the NEN 2767 standard

4.3.4 OTL and movable bridge

In this section, a comparison is made between the available OTL of bridges and the components of a movable bridge. The two methodologies use different base models and are therefore different in structure. First, the differences between the base models are explained. Then the comparison is made.

The OTL model is based on practices from clients and other ontologies like the CB-NL and IMBOR for the taxonomy and past experiences, and the taxonomy for the system decomposition. While the movable bridge decomposition is based on NEN 2767, as NEN 2767 is also based on system decomposition, the OTL on system decomposition is compared to the movable bridge.

A first comparison is made between the classes in both models. Both models consider components to be the class of the model. Also, the relationship is the same. However, the NEN 2767 decomposition does not contain any attributes, whereas the OTL does. Also, the movable bridge's decomposition is not linked to any materials. An overview of the differences and similarities is given in Table 7.

Table 7 Movable bridge decomposition and OTL: decomposition

Aspect	Movable bridge	OTL decomposition
Class	Components	Components
Relationships between classes	Part-of	Part-of
Attributes	×	✓
Materials	×	✓

A second comparison is made between the actual components of the two models. The main difference between the two bridge types follows from the taxonomy, which distinguishes movable bridges on the movable component from other bridges. Therefore, the components responsible for moving the bridge are compared to the Arcadis OTL. Examples of parts responsible for moving a bridge are power units, which are responsible for generating and delivering power to lift or rotate the bridge, and attachment systems, which are responsible for attaching and detaching the movable part of the bridge to the static parts of the bridge.

The examples mentioned above are both classified as elements according to the NEN 2767 standard. These elements contain several components. These components are compared to components in the Arcadis OTL decomposition. Part of the comparison for the two power units in a bridge can be seen in Figure 24. The full figure can be found in Appendix D. Matches are made based on name similarity using a maximum Levenshtein distance of 3 (Euzenat & Shvaiko, 2013). As can already be seen from Figure 24, there are few matches between the OTL and decomposition. Only the valve matches; however, these have a different parent class. This can be a problem with the specific needs; however, if well-defined, this should not be a problem. In the full analysis, there are only three matches out of the 33 components analyzed. so, resulting in a 9% match. There are matches between the power unit types, like the cover, as can be seen in the figure below.

From these few comparisons we can conclude that the OTL currently does not fit the movable bridge decomposition, although the structure of the OTL is intended to work as the backbone. Therefore, this OTL should be developed further to also include the movable bridge decomposition.

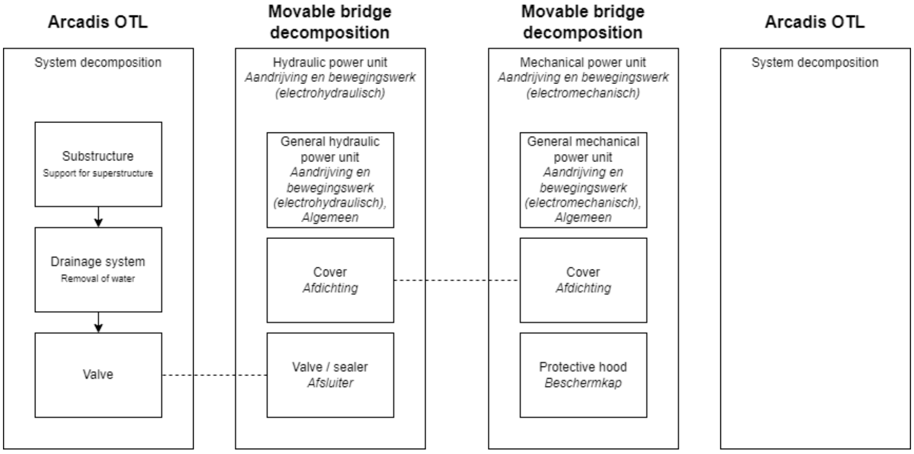


Figure 24 Arcadis OTL and movable bridge comparison containing power units (partial, full figure available in Appendix D)

4.3.5 OTL and AM

In this section, the OTL developed by Arcadis is compared to the AM processes. The OTL is developed for the design and engineering parts of projects; however, it would yield benefits if expanded towards AM. The nature of the OTL is useful for AM, as AM and specifically processes such as FMECA make use of the decomposition of an asset for further analysis.

Some attributes currently available in the OTL are also useful for AM. Examples are:

- Annual average daily traffic
- Deflection
- Design working life
- Redundancy

The full overview of all attributes from the OTL useful for AM is available in Appendix E. Some attributes of AM can be represented by these attributes, but various attributes are still missing. To further expand the OTL for use in AM, examples of attributes that should be added are:

- Type of maintenance
- Failure mode
- RUL

Implementing these attributes on various levels in the OTL will help enable standardized AM not only across bridges but also across multiple structures. However, as the OTL is still under development, this conclusion is left as a recommendation to Arcadis and is not investigated or implemented further.

4.4 Data model static data

As stated at the start of this chapter, we try to create a data model as an interface. This data model contains asset classification and asset information. Now that we have discussed the classification information through OTL, we can create a simple data model. This classification can be combined with the static data types that are used by GIS methods and that are used by FMECA, as these data types are described in the current Arcadis reference model. We first discuss the need for a data model. This is followed by a more in-depth explanation of the three initial components: asset classification, asset information through GIS, and asset information through FMECA.

4.4.1 Need for data model

After the current model analysis and literature review, we realized the need for an interface. An interface can have different forms, but as an OTL containing the asset classification is currently in development at Arcadis, a data model is most compatible. This can also be seen from the connected domains to AM, as seen in Figure 25, which can be connected through a data model. From the reference model, we get the steps and order of the data; from the survey, we get the most important steps and methods to be in the first version of the data model; and from the asset classification, we can incorporate the OTL as the classification or backbone of the data model.

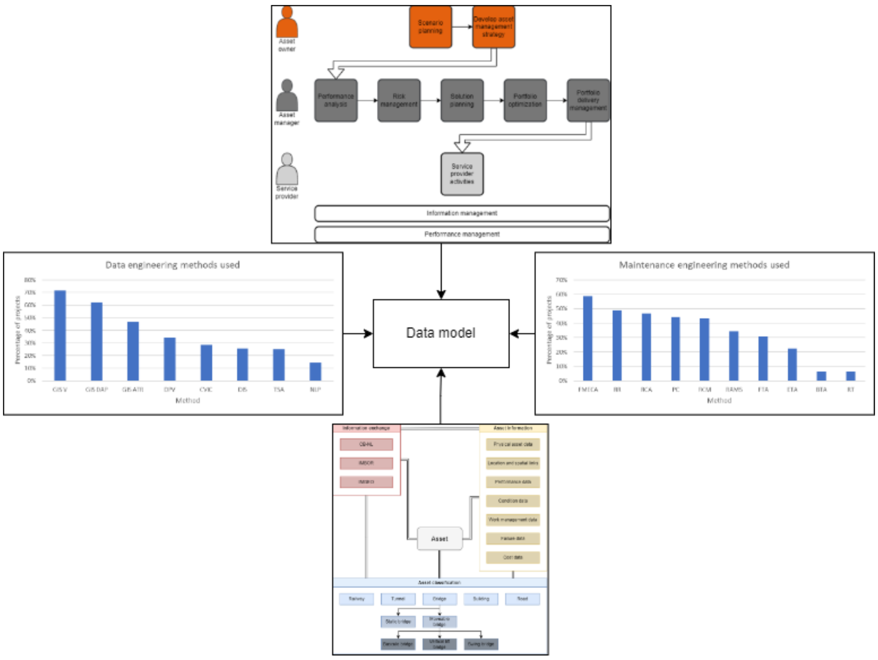


Figure 25 The connection between the reference model, data engineering methods, maintenance engineering methods and asset classification and information; connected through a data model.

A data model is a model in which a simplified version of all data and information elements is represented and connected using text and symbols. In data models, data flows and types are shown. There are various levels of abstraction for data models (SAP, 2023):

- Conceptual data model
  The highest level, containing all elements of the model but no details, shows the structure of logical data models.
- Logical data model
  The second level, specific to a part of the conceptual model. Contains details but not application or database specific.
- Physical data model
  The most detailed level contains database and application specific information.

The conceptual model for the Arcadis reference model is shown in Figure 26. In the model, three key steps from the Arcadis reference model are shown. Each of these steps relates to an aspect of the data model. The conceptual data model is simple and shows the context in which the data model is placed. The aspects in blue are the static data methods and are well described at Arcadis; there is a lot of information readily available. For aspects of performance analysis and predictive maintenance, the data needed is not known for Arcadis.

As can be seen in the conceptual model, the OTL representing the asset classification is connected to all other aspects. This can also be deduced from Figure 21 by the link between classification and information. This also shows the importance of a well-designed OTL for asset classification.

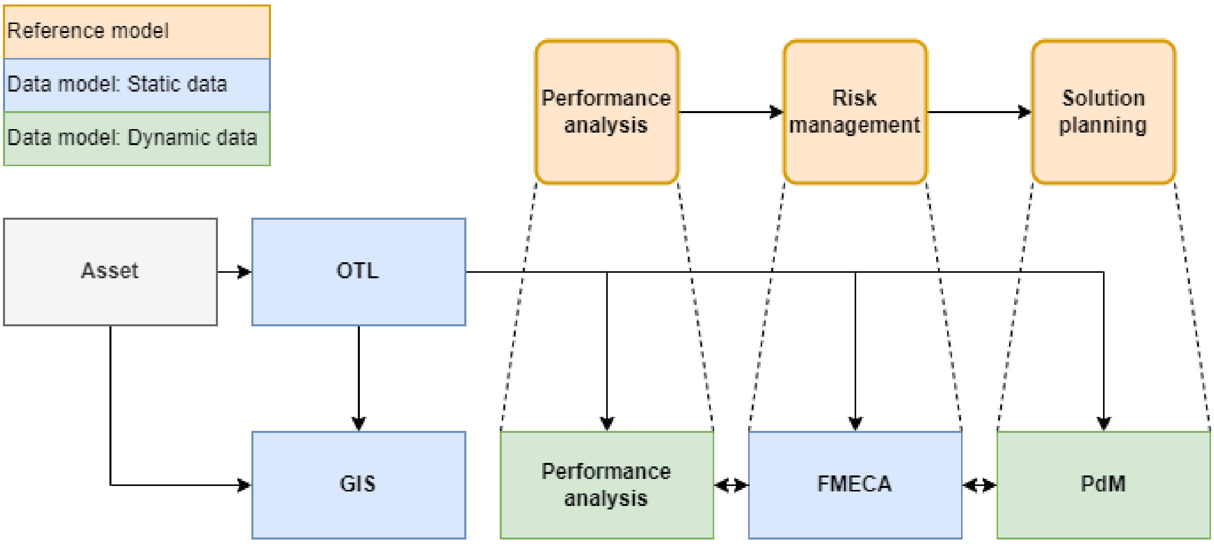


Figure 26 Conceptual data model and the context

For the logical data models, UML class diagrams are used. Further explanation of these class diagrams is available in Appendix F.

4.4.2 Asset classification: OTL

We have discussed the OTL in Section 4.3.2. In this section, the logical data model of the OTL is discussed. The logical data model of the OTL is visible in Figure 27.

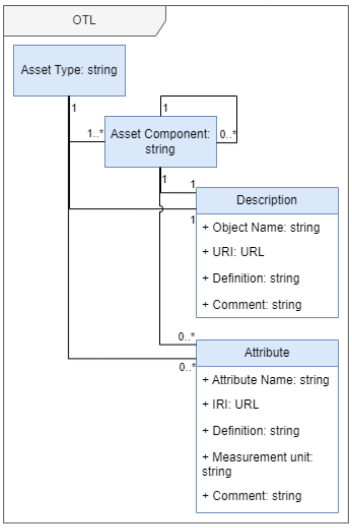


Figure 27 Logical data model of OTL

The data in the OTL is mainly focused on classifying the asset. It contains a description of each part of an asset. It can also contain attributes of components. An example of an attribute is the lifespan of the asset component's drainage system. Notable is that each asset component can have other asset components as subcomponents.

4.4.3 Asset information: GIS

GIS is another important aspect for AM of infrastructure assets. GIS is a broad domain, which could include a broad data model of itself; however, we only limit ourselves to the basic information for GIS. The parts included in the model are from the use case of the movable bridge. The logical data model for GIS is available in Figure 28.

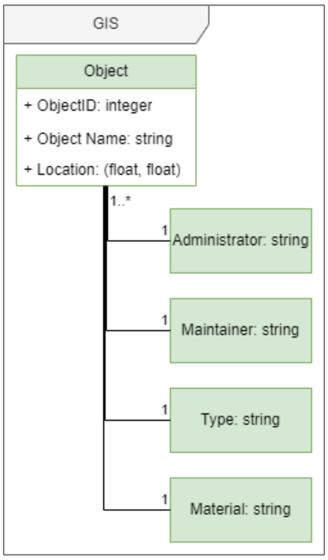


Figure 28 Logical data model of GIS

The data in GIS is about identifying parts of an asset, in this case objects. Important is the location. Additionally, there is some information for AM available, such as the maintainer and administrator.

4.4.4 Asset information: FMECA

The last part of this section is the logical data model for asset information from FMECA. FMECA is the most commonly used method by maintenance engineers, following Section 2.3.3. FMECA is also discussed in the literature review. From there, it follows that there are various common methods and aspects to FMECA. Arcadis also implements various of these aspects. The reference model of Arcadis also proposes creating a business value framework. This business value framework is then used for the FMECA. Therefore, the business value framework is included in the logical data model. FMECA is astatic data method because it focuses on possible failure modes, which do not change over time rather than actual failures. The model is visible in Figure 29.

FMECA is also dependent on components of an asset, here called an element, just like GIS. Then various functions' failure modes and causes are determined and scored. FMECA also contains so-called operations in occurrence and criticality. These operations use the descriptions of the business value framework to determine the accompanying score.

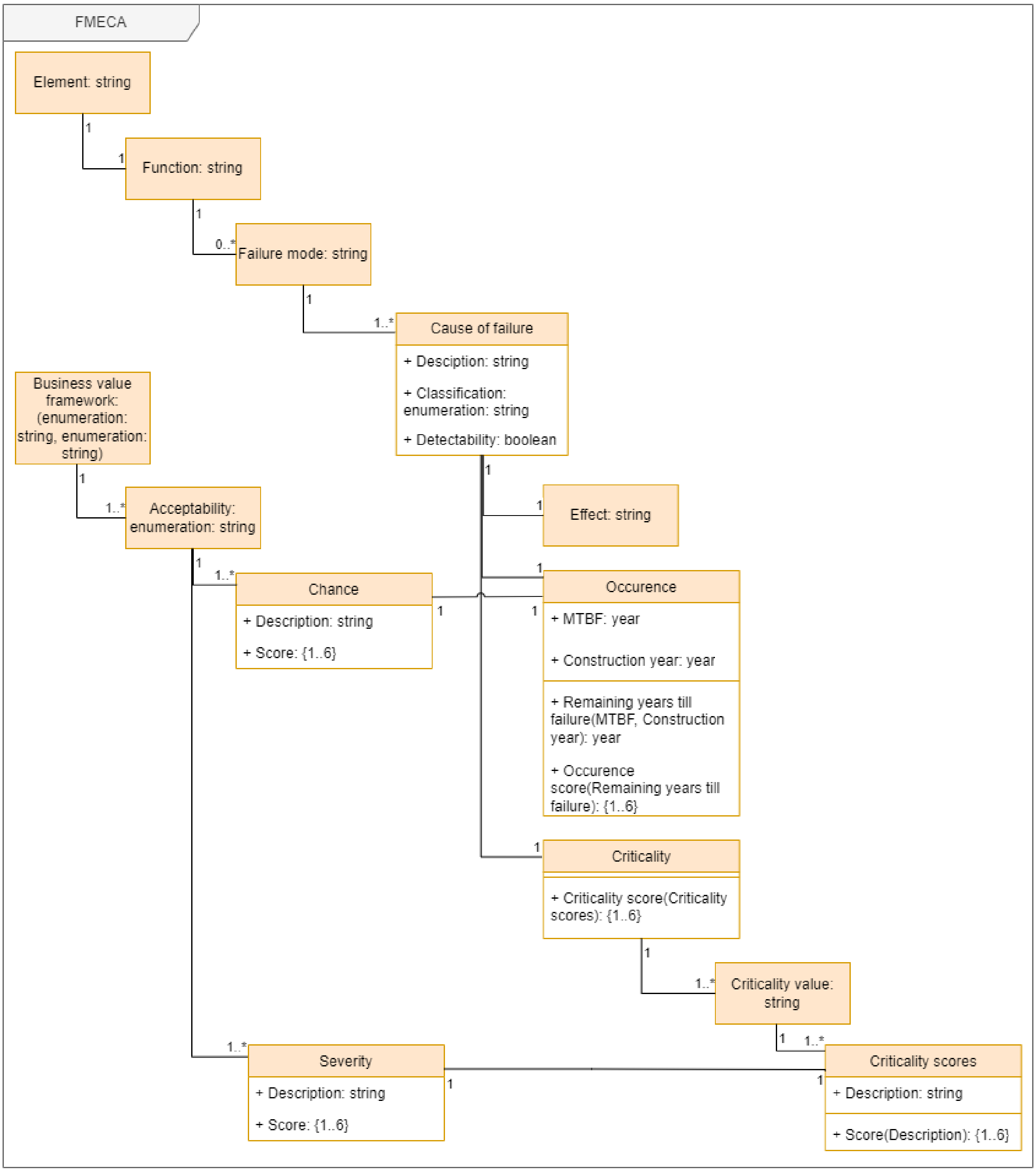


Figure 29 Logical data model for FMECA

4.4.5 Combined logical data models

Now that the three logical models have been created, they can be combined into a single logical model. Following the conceptual model from Figure 26, each of the data models contains some type of asset component; therefore, the models are combined on this aspect. The full combined data model is available in Appendix G. A simplified version is visible in Figure 30.

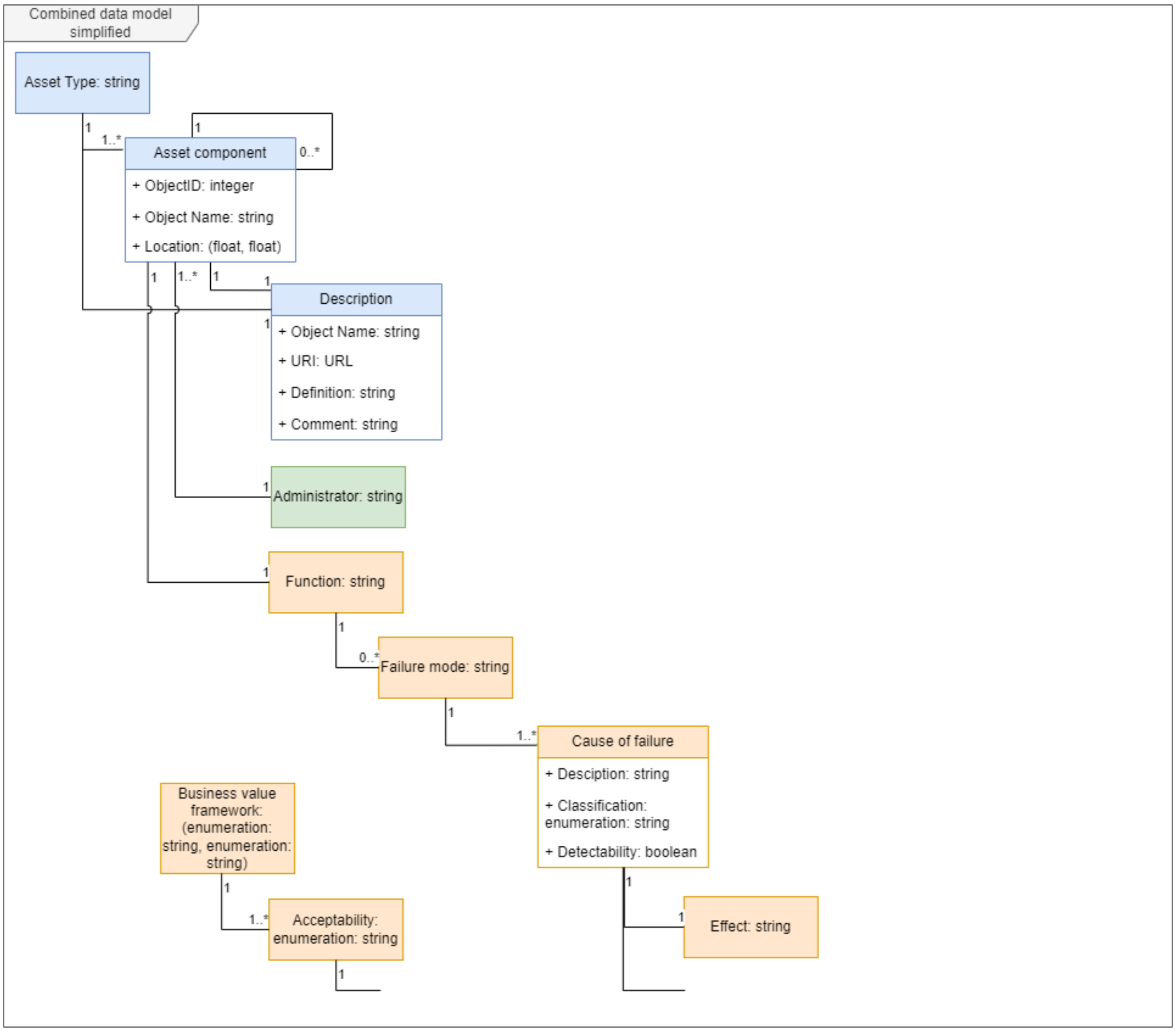


Figure 30 Logical data model of OTL, GIS and FMECA combined – simplified

4.5 Data model dynamic data

The last part of this chapter is about combining the data model from Section 4.4 with the performance analysis, and condition monitoring and prediction analyses. Therefore, the logical models of these two models are discussed first. This is then followed by a discussion on the combined and final data model.

4.5.1 Performance analysis

Performance analysis should be included in the data model as it is one of the most performed steps following the survey in Section 2.3. The data elements that are important for this analysis are not described in the reference model. Therefore, the data model for performance analysis is validated in Section 5.2. We expect that having measurements of at least one key performance indicator is needed for providing feedback on the performance of an asset. This can then be linked to other data sources to determine the performance of the asset. If linked to the strategy of the asset owner through the risk matrix, we can determine the acceptability of the current performance of the asset. The logical data model of the performance analysis is visible in Figure 31.

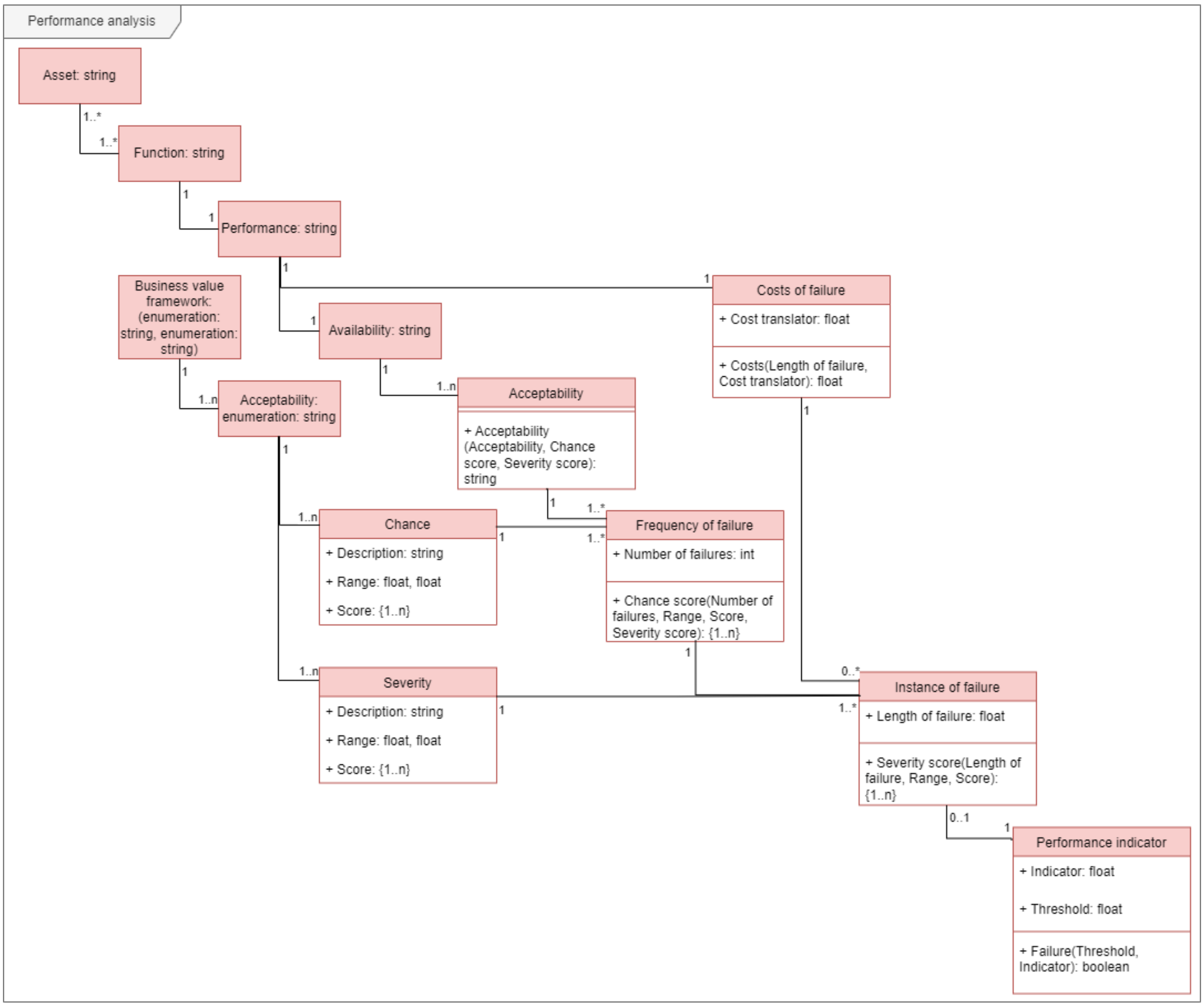


Figure 31 Logical data model of performance analysis

4.5.2 Condition monitoring and prediction

Condition monitoring and prediction are the basis for PdM. PdM is becoming important in the solution planning steps and therefore should be included in the reference model. But also here, the data elements that are important for this analysis are not described in the reference model. Therefore, the condition monitoring and prediction analysis are also validated explicitly in Section 5.3. We expect that having an indicator for a failure mode is important to predict the condition. When we use machine learning models on these features, we can possibly predict the current state of a failure mode. The logical data model of the condition monitoring and prediction analysis is visible in Figure 32.

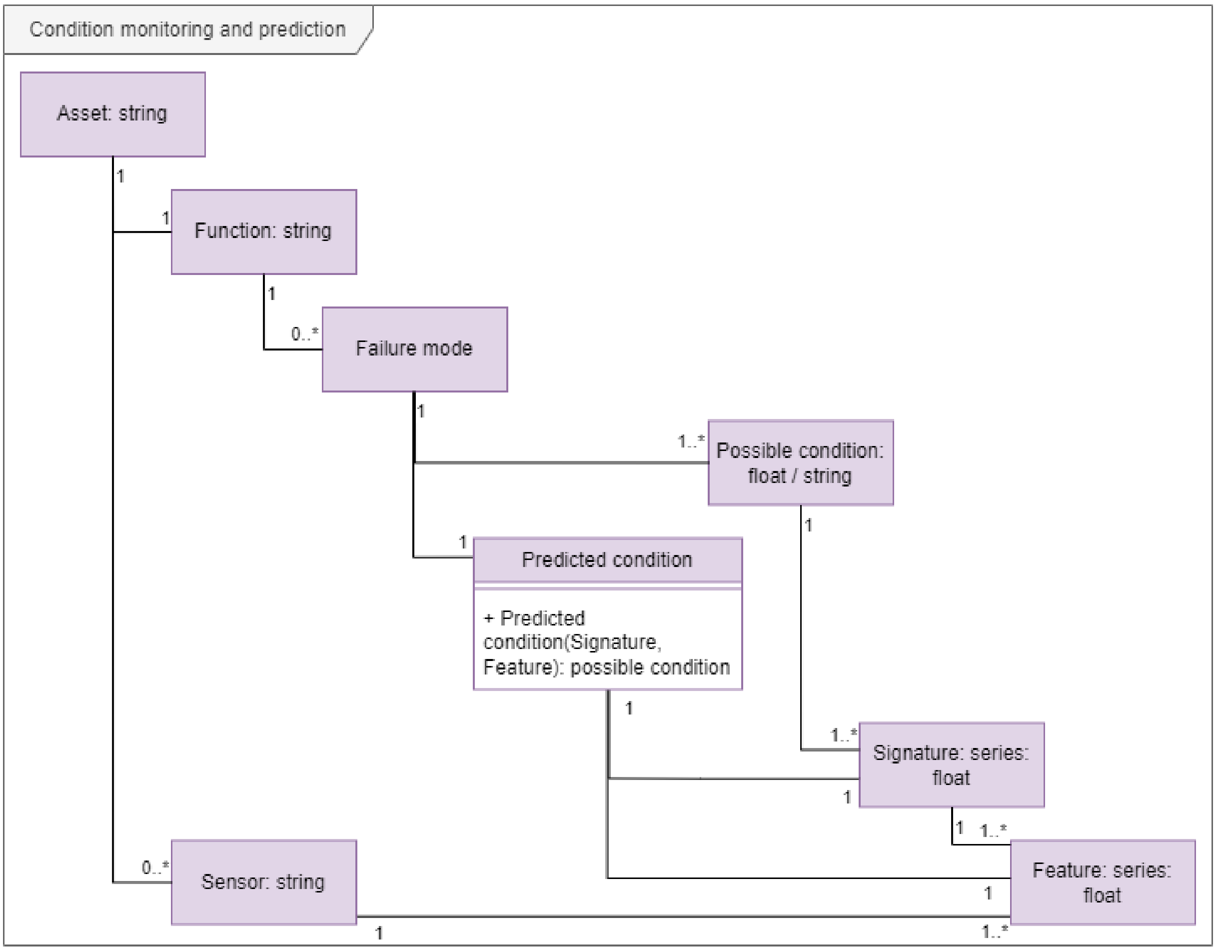


Figure 32 Logical data model of condition monitoring and prediction

4.5.3 Combined model

Now that we have the combined data model of the static data methods and the logical data models of the performance analysis and condition monitoring and prediction analysis, we can combine these into a single logical model. Following the conceptual model from Figure 26, each of the data models contains some type of asset component; therefore, the models are combined on this aspect. There are also multiple methods that use functions and the risk matrix. The full combined data model is available in Figure 33; an enlarged image is given in Appendix G.

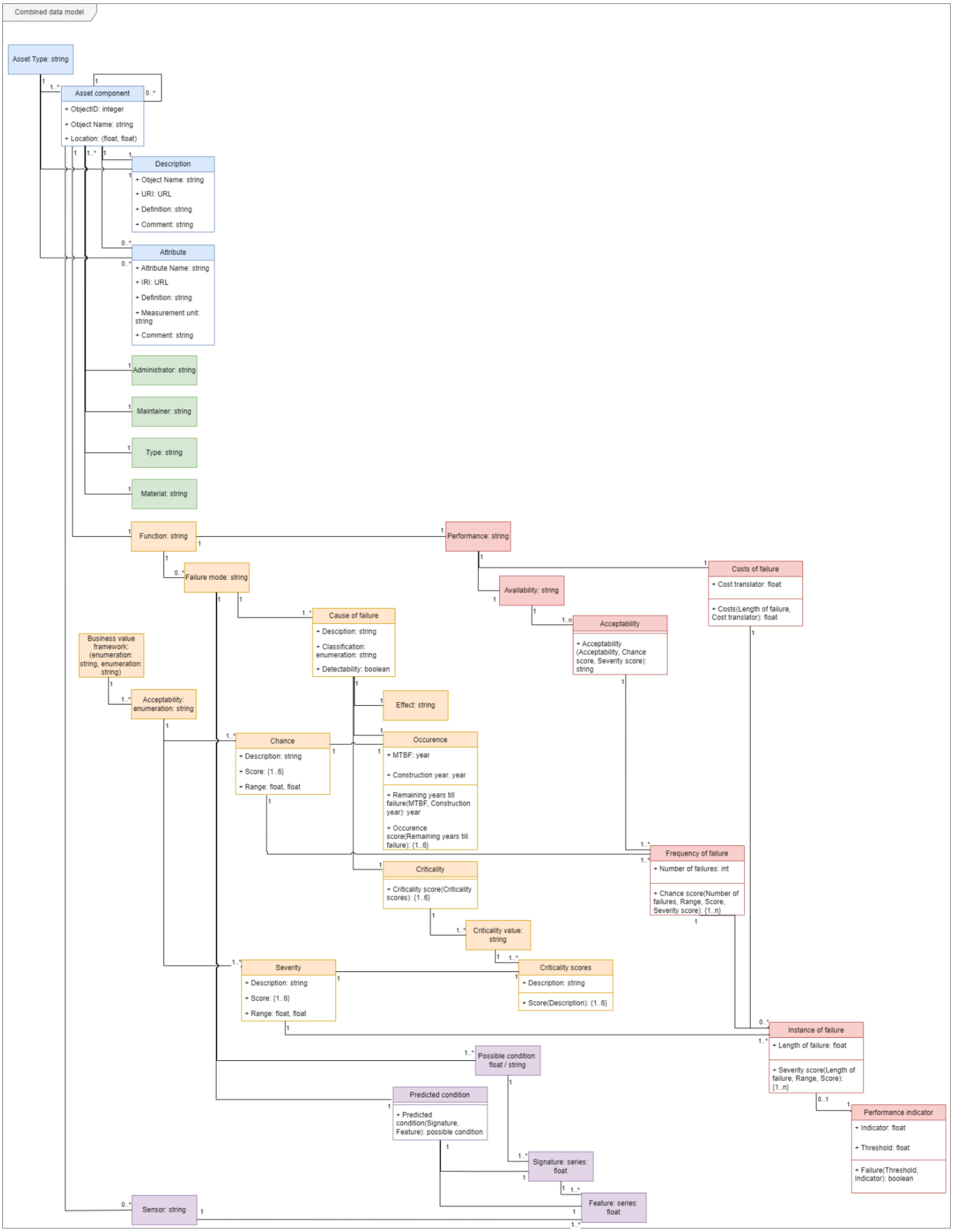


Figure 33 Logical data model of all analyses

5 Validating the data model

Now that we have set up a data model using methods that use static and dynamic data, we validate this data model through a use case. The methods for dynamic data are not yet described in the reference model due to the lack of knowledge on how to do it. Therefore, in addition to validating the data model we also develop these methods further.

5.1 Static data methods

As the methods and concepts using static data are well established in the Arcadis reference model, and the fact that the data models are based on these extensive descriptions, limited validation is necessary. We analyze the use case of a movable bridge, already briefly introduced in Section 4.3.3. For the OTL the validation is already partly done in Section 4.3.4, where the results show compatibility between the model and practice, although there is a large difference between naming. However, the structure of the OTL is compatible with the breakdown of the movable bridge in the use case, therefore the data model is suitable for use in the case, on the point of OTL.

For GIS the data model is also based on the use case. Therefore, all the elements present in the data model are also present in GIS. The model is further validated based on the GIS models of other movable bridges, where each of the elements is also present, even though locations and maintainers are different.

The same holds for FMECA, where in addition to the use in FMECA, functions and failure modes are also used in the validation steps of performance analysis and condition monitoring and prediction. FMECA is also validated on multiple projects, where FMECA is made in Excel. These FMECAs both follow the same template and therefore have the same data elements therefore the data model is compliant both with the use case and the other similar cases.

5.2 Performance analysis

The first step, as can be done by the asset manager according to the reference model, is the performance analysis. Within the reference model, there are no clear methods defined for how performance analysis should be done. From interviews, it is also apparent that there is no common understanding of performance analysis. This is possibly due to the varying measures of performance per project. Therefore, performance analysis is first discussed using the case study of movable bridges, and the lessons learned are then fed back into the data model. Performance analysis is an analysis based on datasets from the asset; therefore, first the available datasets for the use case are discussed. Next, KPIs availability and costs are discussed, followed by a sensitivity analysis of the assumptions made in this analysis. Based on this sensitivity analysis we can conclude whether the data elements in the data model for performance analysis are valid.

5.2.1 Data description

There are three datasets readily available for the asset. An overview of the datasets is available in Table 8. Each dataset is discussed in more detail below.

Table 8 Description of datasets of performance analysis

Dataset	Description	Data source
Openings	Contains data on the openings of the bridge including opening duration and the number of vessels passing	Sensor data from asset owner (https://prpnh.maptm.nl/)
Business Value Framework	Contains the filled-in risk matrix	Expert opinions from asset owner
Intensities	Contains the intensities of roadway traffic per day	National data portal road traffic (https://www.ndw.nu/)

The first dataset is a set containing data about the passing of water vessels per opening of the bridge. This is a dataset that contains various rows per passing ship under the bridge. Each passage has a date and time and a length of being open, also known as the opening duration. The dataset is available for all passages between January 2017 and April 2023. The data preparation process and descriptive statistics are available in Appendix H. The graph in Figure 34 summarizes the findings from the appendix. From here, in 2021 and 2022, the situation changed from the other years, as the total opening duration stayed roughly the same while the number of openings reduced. The number and duration of openings for 2023 are still low because only the data from the first 4 months is available.

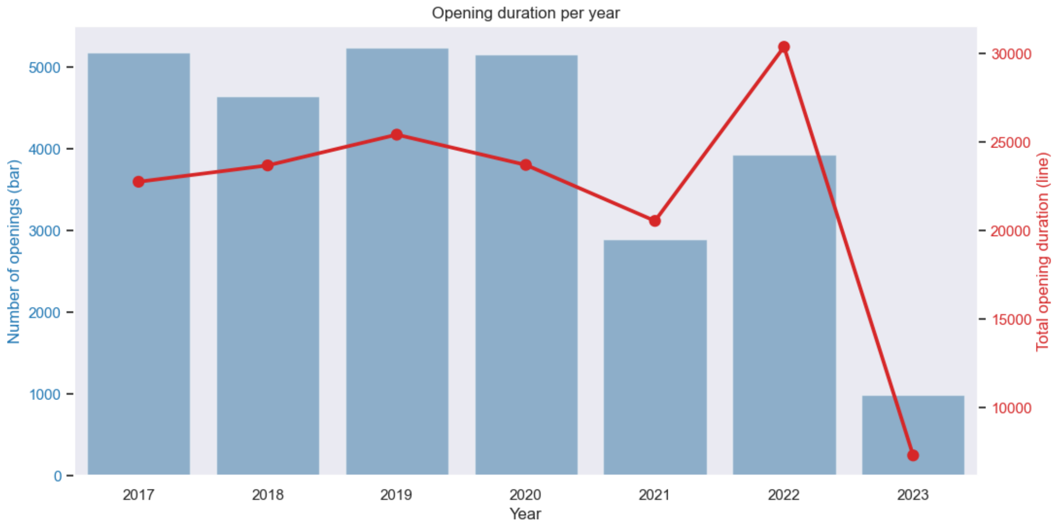


Figure 34 Opening durations and number of openings per year

The second dataset is the business value framework for the use case. This is a matrix coming from the asset owner. A business value framework, also known as a risk matrix, contains the severity of failures and the chance of failures happening. The risk matrix is shown in Figure 35. Through this matrix, we can link the strategic goals of the owner to the operational performance of the asset.

The matrix contains quantifiable values for both severity and chance. The quantifiable and thus added part by the asset owner is shown in gray in the figure. As the bridge connects two important road sections, negligible risks do not exist. If a failure occurs, it is at least a minor risk.

Severity		Chance					
		Rare	Unlikely	Possible	Likely	Very likely	Almost certain
		< once in 30 years	≥ once in 30 years	≥ once in 10 years	≥ once in 5 years	≥ once a year	≥ once a week
Catastrpohic	> 1 week not availble	to be mitigated	not acceptable	not acceptable	not acceptable	not acceptable	not acceptable
Severe	> 1 day not availble	acceptable	to be mitigated	not acceptable	not acceptable	not acceptable	not acceptable
Major	> 4 hours not availble	acceptable	acceptable	to be mitigated	not acceptable	not acceptable	not acceptable
Moderate	> 1 hour not availble	acceptable	acceptable	acceptable	to be mitigated	not acceptable	not acceptable
Minor	≤ 1 hour not availble	acceptable	acceptable	acceptable	acceptable	to be mitigated	not acceptable
Negligible		acceptable	acceptable	acceptable	acceptable	acceptable	to be mitigated

Figure 35 Risk matrix

The third dataset is on the intensities of traffic on the road over the bridge for each day in the period 2017–2023. Given the intensities, we can calculate the amount of traffic that has to wait due to bridge openings and bridge failures; thereby proposing data for the performance indicator costs.

5.2.2 Performance

To determine the performance of a bridge, we must set indicators to measure performance. There are various indicators that we can use. An example is RAMS, which consists of the indicators reliability, availability, maintainability, and safety. Another example comes from the risk matrix of the asset owner, which contains indicators on traffic flow, livability, safety, image, and costs. Based on the available data, we choose two indicators: availability (known as traffic flow in the risk matrix) and costs. To calculate the performance on these indicators, we need to determine the statistics around openings of the movable bridge. To determine these statistics, we need some assumptions. These assumptions are discussed first. Later, a sensitivity analysis is done on some of these assumptions to test their validity. All assumptions are also validated by asset managers at Arcadis.

The assumptions are:

- To determine if an opening is a failure, we use thresholds. A formula for calculating a threshold is set up with Arcadis' maintenance engineers:
  $$\begin{aligned} Threshold = 0{,}5 * MAX(mode[Timeperiod]) \\ + 0{,}5 * \#PassingVessels * \mathrm{MAX}\,(mode[Timeperiod]) + Margin \end{aligned}$$
  Here, we take the mode as the basis for the threshold. The mode can be determined by various methods over a variable time period. An example is the mode per year, but the mode per year per month is also possible. Half the mode is assumed to be because of the opening and closing of the bridge. The other half of the mode is assumed to be because of vessels passing and is therefore multiplied by the number of passing vessels. A margin is also added to the formula as opening durations slightly above the mode should also be accepted.
- Openings taking longer than 1000 minutes are invalid, based on news stories (allesinkaagenbraassem, 2021).

- Failures follow a linear extrapolation. So, if a failure occurs once every 5 years, it occurs six times every 30 years.
- The business value framework can be filed based on flow only, excluding livability, safety, etc.

Using these assumptions, the availability of the bridge can be determined. Availability is the time the bridge operates without failure as a percentage of the total operating time of the bridge. As an example, in January 2019, there were 635 minutes the bridge was open in total. Of these 635 minutes, 31 minutes are credited as failure; therefore, the availability in January 2019 is

$$100\% - \frac{31}{635} \approx 95\%$$

Most months the bridge operates fine with availabilities between 90% and 100%; however, there are some exceptions; most notably, there is no operation of the bridge in March 2021 because of the data cleaning process, where new sensors were installed and therefore data was removed.

Using the flow, assumptions, and failure statistics of the bridge, the business value framework can be filled in. The filled-in risk matrix can be seen in Figure 36. As can be seen, major and moderate failures occur more often than is acceptable, while minor failures are to be mitigated. Therefore, the bridge is currently underperforming. Notable is that due to the cleaning of the data, it is not possible that a failure has a severity of "severe" or "catastrophic", as these require failures of a longer opening duration than 1000, which are filtered.

Another measure of performance is cost. Part of the costs can be calculated based on lost time. Using this variable, we can calculate the costs for the public without including costs for maintenance or operation. This can be done by multiplying the amount of vehicle loss hours, based on intensity from the third dataset, with the costs per vehicle loss hour, which are approximated by €10.- per lost hour (Voerknecht, 2017). The costs per year for failures and total openings can be found in Figure 37.

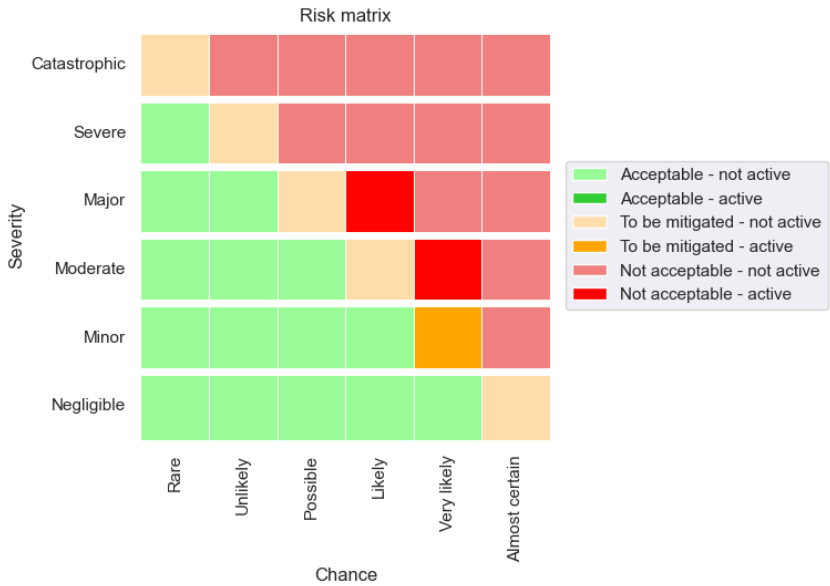


Figure 36 Risk matrix of movable bridge over period 2016 – 2023

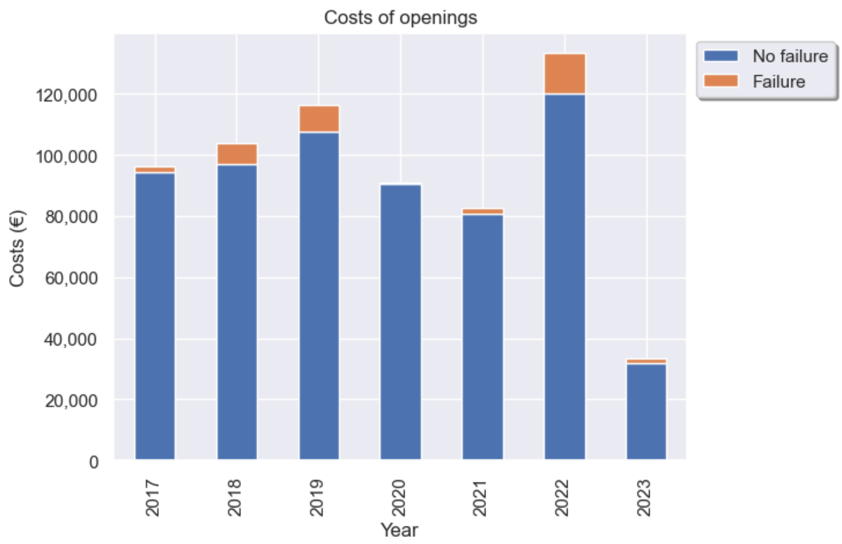


Figure 37 Costs of opening the bridge based on vehicle loss hours

5.2.3 Sensitivity analysis

To test the validity of the analysis further, a sensitivity analysis is done. In a sensitivity analysis, input parameters are changed, and the effect on the output is analyzed. This can help to investigate which parameters have a large effect on the outcomes. The three inputs that are subject to analysis are shown in Table 9, as output availability is measured. We do this to test our analysis's robustness, especially the robustness of the assumptions.

Table 9 Parameters of sensitivity analysis

Input parameter	Range	Default value
Method of determining the threshold	(constant) Threshold: 5, 7, 9, 11 Margin: 0, 2, 4, 6	Margin 4
Time period of mode	Year, Month, Weekday, Year – Month, Year – Weekday, Month – Weekday, Year – Month - Weekday	Year - Month
Data cleaning	Cleaned, Not cleaned	Cleaned

The first analysis is on the method of determining the threshold for classifying openings as failing or not failing. We can replace the formula with a constant number, such as 9 minutes. We can also change the value we take for the margin in the formula. The results of the analysis are visible in Figure 38. From here, it can be seen that until 2020, changes are small compared to afterwards. This is because measurements of opening duration are more accurate from March 2021 onward. It can also be seen that availability for the constant threshold rules is lower except for the highest threshold. Notable is also the fact that both methods follow the same pattern for all values. Especially the low values show a largely varying pattern compared to the higher values. There is a slight exception in 2021 where threshold 11, margin 2, margin 4, and margin 6 follow a somewhat higher curve than threshold 5, threshold 7, threshold 9, and margin 0, which follow a lower curve.

As availability ranges from 0% to 98% for threshold 5 and margin 6 in 2023, respectively, it can be concluded that the method of threshold determination has a large impact on availability.

Based on the 97.5th quantile between opening duration and mode, which is 4, and experience from Arcadis employees, the margin of 4 is chosen as the default option.

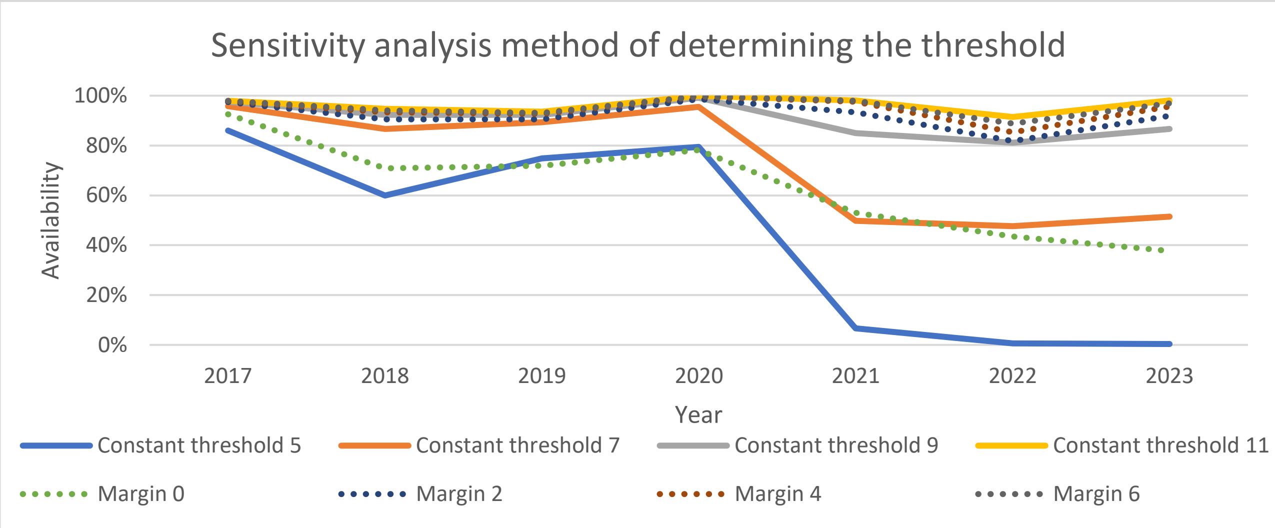


Figure 38 Sensitivity analysis method of determining the threshold

The second analysis is on the time period used for determining the mode. The results of the analysis are available in Figure 39. Again, there is a larger sensitivity in the analysis after March 2021. Notably, the single dimension options year, month, and weekday follow the same availability until 2021. Month and weekday follow the same availability for all years, and logically, the same availability is followed by the combination month - weekday. These options follow the same availability, as measurements before 2021 are based on whole digits, and therefore the mode is always, the same for each, weekday, month and year. Year – month and year – month – weekday follow the most stable patterns, showing the least sensitivity. Due to this reason, year – month is chosen as the default value for the other analyses.

Also, the time period shows a large variance, which can be seen from the largely varying availability in 2023, even though it is a bit smaller than the influence of threshold or margin.

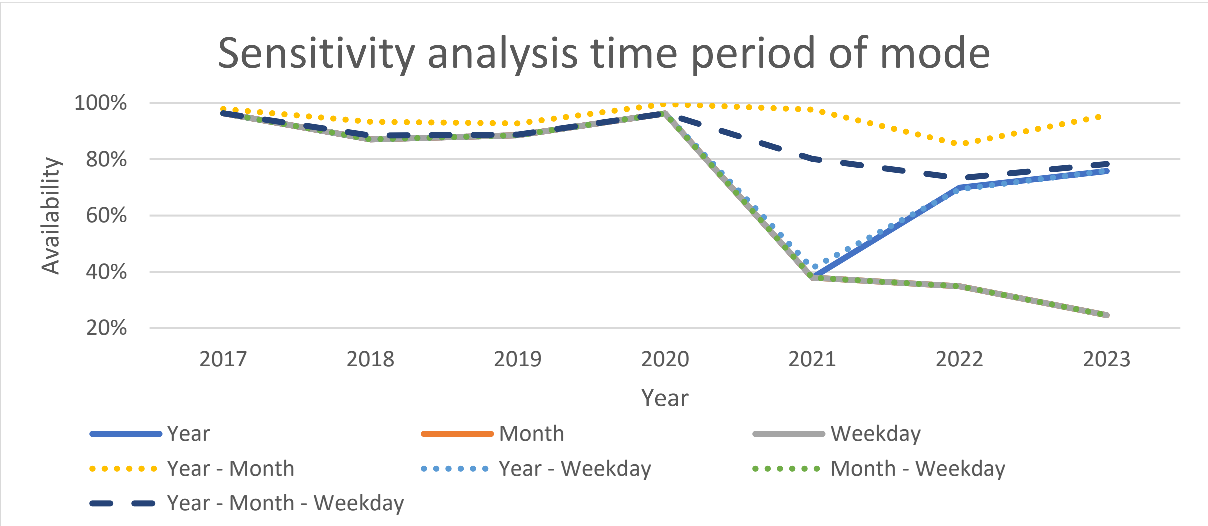


Figure 39 Sensitivity analysis time period of mode

The third sensitivity analysis is on the influence of data cleaning. The results are shown in Figure 40. Here it can be seen that the influence of cleaning is not as great as in the previous two analyses, except for 2021, where large outliers are removed. The interval of availabilities is also smaller than for the other inputs. There are also cases where the cleaned data shows lower availability than the not-cleaned data because of removing the duplicate openings.

It is expected that cleaned data better represents the actual real world as disruptions caused by sensor defects are removed while the bridge is still working correctly, as shown by news outlet posts (allesinkaagenbraassem, 2021). Therefore, the cleaned data is chosen as a default for the other analyses.

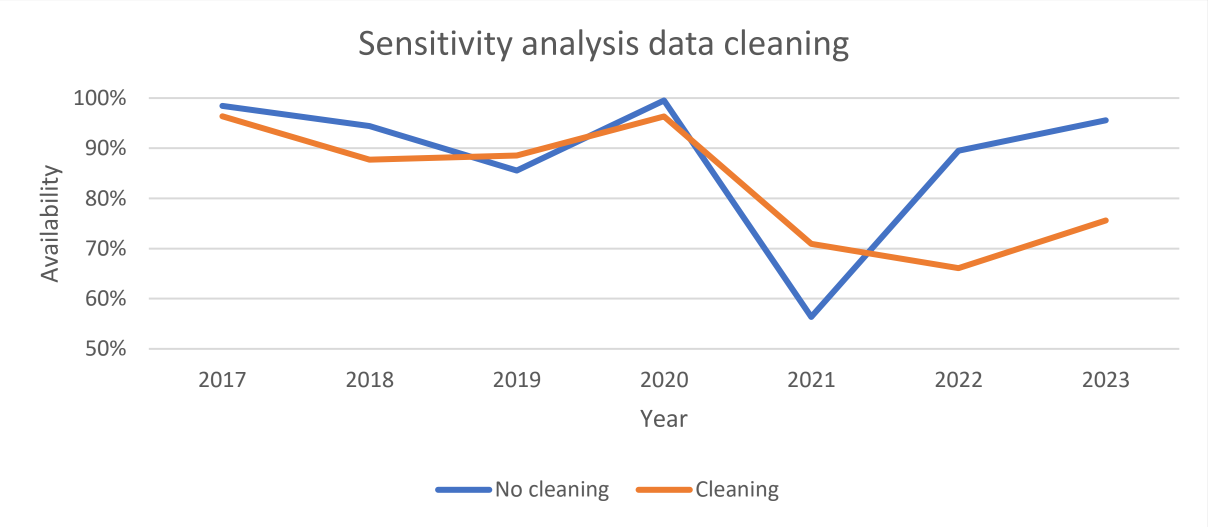


Figure 40 Sensitivity analysis data cleaning

Based on the sensitivity analysis and the execution of the performance analysis we can see that all elements present in the data model are also used in practice. The performance indicator is the opening duration in the use case. Based on a threshold we can then determine the failures. Using these failures, we can calculate the performance on availability and costs, and plot these in the business value framework.

5.3 Condition monitoring and prediction

Another step in the reference model of Arcadis is solution planning. A possible solution-planning methodology is condition monitoring. To prepare the reference model for compatibility with predictive maintenance policies, we include condition monitoring as part of the data model. Predictive maintenance (PdM) can improve the AM of infrastructure assets if done correctly (ITF, 2021); therefore, implementing this approach in the data model can help in providing better information management for AM projects following the reference model. There are also other policies, such as preventive maintenance (PM), but as PdM requires more information and data, this method is chosen for implementation in the data model. Information from PdM can also be used for PM. Currently, the movable bridge uses a combination of PM and corrective maintenance policies; we test whether PdM policies can prevent predictable failures. There are no clear, predefined methods for condition monitoring available within Arcadis. Therefore, first a condition monitoring analysis is done in this part, and then the data requirements are compared to the data model. Condition monitoring is the process of evaluating data coming from an asset and calculating the condition of its components with this data. Based on current and previous values, a prediction of the condition of the asset can be made. This is the base for PdM.

First the data is introduced and described. Then various predictive models are trained and tested. Lastly the influence of individual data inputs is tested on the outputs of the predictive models.

5.3.1 Data description

To do predictive maintenance, a valid dataset of an asset is needed. The bridge used in previous analyses is planned to have sensor data. Unfortunately, this data was not available at the time of writing. Therefore, a synthetic public dataset is used, created by Helwig et al. (2018). This dataset is created by a hydraulic test rig consisting of a hydraulic pump, cooler, valve, and accumulator. All these components are present in the movable bridge as well. Additionally, the pump also operates cyclically and not constantly. There are also some differences; for example, the duration of usage differs for each operation of the movable bridge, whereas the duration is 60 seconds for the test rig. These differences can hinder the implementation of the models in practice.

Coming with the dataset are four failure methods. These are visible in Table 10. Each failure mode has a condition score, also known as a state. The condition of the failure mode indicates the criticality of the current state of the failure mode, as given in the description. As can be seen for all failure modes, the dataset is balanced evenly; this is important as an imbalanced dataset tends to predict in favor of the frequently occurring condition even if otherwise measured. The development of the failure methods over the past 2205 cycles is shown in Figure 41. If we take, for example, cooler failure, we can see that it occurs in approximately a third of the cycles in condition 3, a third in condition 20, and the last third in condition 100. Again, in real life this data would be hard to acquire, but having an estimate based on data is better than no data at all.

Table 10 Description of failure modes

Failure mode	Condition	Description	# cycles
Cooler failure	3 %	Close to total failure	732
	20 %	Reduced efficiency	732
	100 %	Full efficiency	741
Valve failure	73 %	Close to total failure	360
	80 %	Severe lag	360
	90 %	Small lag	360
	100 %	Optimal switching behavior	1,125
Pump leakage	2	Severe leakage	492
	1	Weak leakage	492
	0	No leakage	1,221
Accumulator failure	90 bar	Close to total failure	808
	100 bar	Severely reduced pressure	399
	115 bar	Slightly reduced pressure	399
	130 bar	Optimal pressure	599
	0	conditions were stable	1,449

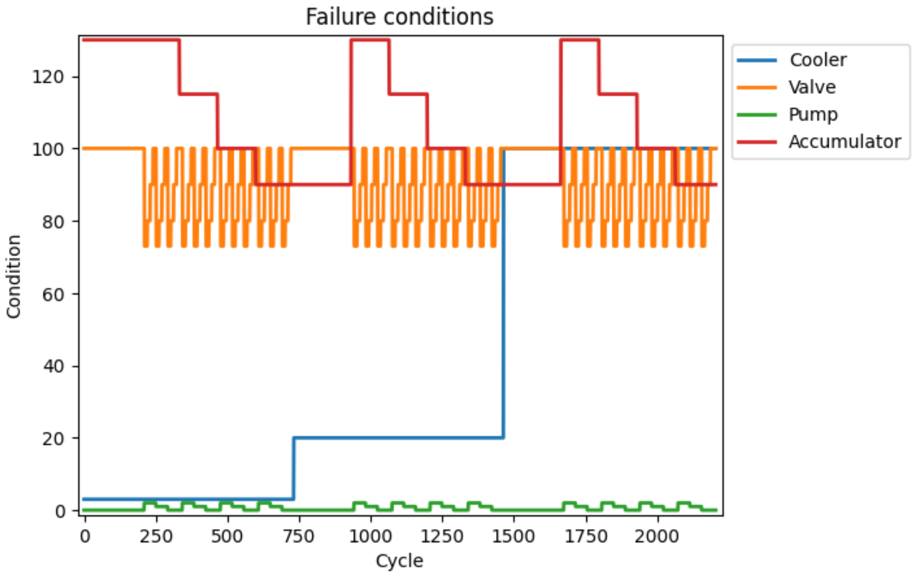


Figure 41 Failure conditions over the cycles

For predicting the condition, 16 sensors are available. The sensors are divided into sensor groups, and some characteristics are shown in Table 11, including whether the sensor group is also available in the movable bridge use case. Each of the sensors is related to some of the failure methods; the relationships between the sensors and failure modes are presented in Appendix I.

Table 11 Description of sensor groups

Sensor group	# of sensors	Measurement unit	Frequency (hz)	Abbreviation	Present at movable bridge
Pressure	6	bar	100	PS	Yes
Motor power	1	W	100	EPS	Yes
Volume flow	2	l/min	10	FS	No
Temperature	4	°C	1	TS	Yes
Vibration	1	mm/s	1	VS	No
Cooling efficiency	1	%	1	CE	No
Cooling power	1	kW	1	CP	No

5.3.2 Predictive models

Now that we have presented the data, we can use machine learning models to make predictions on the current condition of the machine. Based on this current condition, a trend analysis can predict future conditions. To do this, classification algorithms are used. A classification algorithm uses the values from sensors to classify the current state of the machine. So, in the case of the test rig, it tries to classify the current state of a failure method based on the sensor values.

There are various models that are capable of classifying based on sensor data. In this research, six models are presented and tested in this section. All the models are supervised machine learning methods, which implies the models are first trained and then tested on classified data. The models are briefly explained in Table 12. The first 5 models do not require parameter tuning but do need balanced structured tabular data. The neural network is a deep learning model and requires parameter tuning, which is done in Appendix J. Additionally, the reasons the models are considered are discussed in the table. Based on the performance calculated below, the best models are chosen for further research.

Table 12 Machine learning models explained

Model	Description	Reason to be considered
Decision Tree	This model classifies through splitting classes based on nodes, commonly known as leaves, with threshold values. The leaves in this case are the sensor groups.	A generally well-known model with high explainability.
Random Forest	This model uses multiple decision trees and classifies them based on the average classification of these decision trees.	Well known, outperforms individual decision trees in general. Moderately explainable.
(Linear) Support Vector Machines	This model uses planes or vectors to split measurements. A section divided by planes corresponds to a class.	Can handle many sensor inputs and perform well on low amounts of data.
K-Nearest Neighbor	This model uses the difference or distance between measurements to classify. It uses K points with the lowest distance, then chooses the most frequently occurring classification.	The model is well-known, explainable, and has low computing costs.
Naïve Bayes	This algorithm uses Bayes' theorem to determine the probability of the states of classes. To determine this probability, it assumes independence between inputs to make calculations easier.	The model is fast and can handle low amounts of data.
(recurrent) Neural Network	This is a model that is complicated. It is based on a long short-term memory network. This means that data points in series are processed rather than individual points and previous series are also considered. The network is build using 1 input node and then a certain number of layers of LSTM nodes. Each LSTM layer is followed by a dropout layer. Lastly the layers are combined in one output layer with a single node. This network is greatly impacted by hyperparameter tuning. The tuning is available in Appendix J.	The models are known to produce very accurate predictions, although these results are often difficult to explain (Wahid et al., 2022).

To measure the performance of the algorithms, accuracy is used as a prediction measure. Accuracy is the percentage of correct predictions over all predictions. For example, if we take five measures where the state of the cooler condition is 100% and the model predicts that four classes are 100% and 1 class is 20%, then the accuracy is $\frac{4}{5} = 80\%$. Average accuracy is the average of the accuracies of the four failure modes. For all the models, 5-fold cross-validation is used, which means that the data is split into five parts, and in five iterations, each part is used as test data once, while the other four parts are used as training data. Five folds are chosen rather ten folds, or another number, as higher numbers drastically increase computation time whilst lower number increase the influence of bias. The accuracy is then the average accuracy over these five iterations. The accuracy of the methods for predicting failure modes is shown in Table 13.

As can be seen, cooler failure is well predicted by all methods except support vector machines. Valve failure is, however, poorly predicted by most methods except the neural network. Pump leakage and accumulator failure are both somewhat decently predicted by the random forest, decision tree, and neural network, but not by the other methods. Overall, the neural network performs best on accuracy, followed by the random forest and decision tree. The support vector machines perform worst, possibly due to having too much data. For the further analyses, only the random forest and neural network are used, again using 5-fold cross-validation.

Table 13 Accuracy of prediction algorithms over the failure modes

Accuracy (%)	**Cooler failure**	**Valve failure**	**Pump leakage**	**Accumulator failure**	**Average accuracy**
Decision Tree	96.37%	47.39%	70.52%	68.93%	70.81%
Random Forest	98.19%	51.93%	78.91%	73.02%	75.51%
Support Vector Machines	64.44%	51.02%	55.37%	36.62%	52.35%
K-Nearest Neighbor	97.05%	49.21%	54.88%	41.04%	60.54%
Naïve Bayes	88.89%	47.39%	46.94%	38.32%	55.39%
Neural Network	98.87%	83.67%	77.82%	70.93%	82.82%
Average accuracy	85.00%	55.56%	63.95%	55.40%	64.98%
Average accuracy Random Forest and Neural Network	98.53%	67.80%	78.37%	71.97%	79.17%

5.3.3 Input data analysis

Now that we know that the neural network and random forest can classify the failure mode conditions well, we look at the impact of the input data. With this analysis, we want to identify the most important sensors for the models; thereby identifying the sensors that should be implemented in the data model. To do this, two methods are used: the first is based on scenarios where we leave out groups of sensors, and the second is based on the importance of each sensor individually in the random forest.

First, the five scenarios below are compared:

- All sensors: a scenario where all sensors from the dataset are considered
- Bridge: a scenario where only the sensors that are to be placed on the bridge are considered. These sensors are pressure, motor power, and temperature sensors.
- No pressure: a scenario where the sensors of the bridge scenario are considered without the pressure sensors.
- No motor power: a scenario where the sensors of the bridge scenario without the motor power sensors are considered.
- No temperature: a scenario where the sensors of the bridge scenario without the temperature sensors are considered.

The first scenario was used for the analysis of the algorithms in Section 5.3.2. The resulting accuracy for each model can be found in Table 13. As discussed earlier, the best-performing methods are the neural network and the random forest; therefore, the scenarios are analyzed based on the average accuracy of these methods. It is to be expected that the predictions based on all sensors are the most accurate, as a lot of data is available.

The second scenario is the one that is modeled according to the movable bridge use case. This bridge contains the three sensor types: pressure, motor power, and temperature. Additionally, the bridge will also contain other sensors; however, these sensors are not present in the data set and are therefore left out of the analysis. The resulting accuracy of this scenario for the methods is shown in Table 14. As can be seen, the performance drops drastically for the random forest, while the neural network performs almost the same. Cooler failure is still very predictable, but the other three, especially accumulator failure, drop drastically in predictability.

Table 14 Accuracy of prediction algorithms over the failure modes for the bridge scenario

Accuracy	**Cooler failure**	**Valve failure**	**Pump leakage**	**Accumulator failure**	**Average accuracy**
Random Forest	97.73%	38.82%	56.15%	44.49%	59.30%
Neural Network	98.05%	83.67%	77.37%	70.88%	82.49%
Average accuracy	97.89%	61.25%	66.76%	57.69%	70.90%

The third scenario is the scenario where the pressure sensors are absent. The results are shown in Table 15. As can be seen, there is a low prediction loss compared to the bridge scenario in the neural network. Valve failure even has an increased accuracy at the random forest, which can be due to randomness or the negative influence of pressure sensors on the prediction for this failure mode.

Table 15 Accuracy of prediction algorithms over the failure modes for the no pressure scenario

Accuracy	**Cooler failure**	**Valve failure**	**Pump leakage**	**Accumulator failure**	**Average accuracy**
Random Forest	97,28%	41,90%	53,88%	48,66%	60,43%
Neural Network	97,69%	83,67%	77,69%	63,36%	80,60%
Average accuracy	97,48%	62,79%	65,78%	56,01%	70,52%

The fourth scenario is the scenario of all bridge sensors except for the motor power sensors. The accuracy of the predictive models can be seen in Table 16. Pump leakage drops slightly in accuracy compared to the bridge scenario. The pattern also compares better to the bridge scenario than the scenario without pressure.

Table 16 Accuracy of prediction algorithms over the failure modes for the no motor power scenario

Accuracy	**Cooler failure**	**Valve failure**	**Pump leakage**	**Accumulator failure**	**Average accuracy**
Random Forest	97.64%	37.73%	51.20%	43.85%	57.61%
Neural Network	97.51%	83.67%	77.73%	70.93%	82.46%
Average accuracy	97.57%	60.70%	64.47%	57.39%	70.03%

The last scenario is the bridge scenario without the temperature sensors. The accuracy of the models can be found in Table 17. Cooler failure and accumulator failure predictions become less accurate for the random forest, indicating that these sensors are important for these failure modes' predictions. The neural network still performs fine.

Table 17 Accuracy of prediction algorithms over the failure modes for the no temperature scenario

Accuracy	**Cooler failure**	**Valve failure**	**Pump leakage**	**Accumulator failure**	**Average accuracy**
Random Forest	92.06%	42.36%	61.04%	30.84%	56.58%
Neural Network	98.46%	83.67%	78.55%	68.12%	82.20%
Average accuracy	95.26%	63.02%	69.80%	49.48%	69.39%

The comparison of the average accuracy of the random forest and neural network models over the failure modes is visible in Figure 42. From here, the expected loss of accuracy can be seen, especially the loss between all sensors and the bridge scenario. Cooler failure remains constant throughout the scenario. Especially accumulator failure drops accuracy drastically.

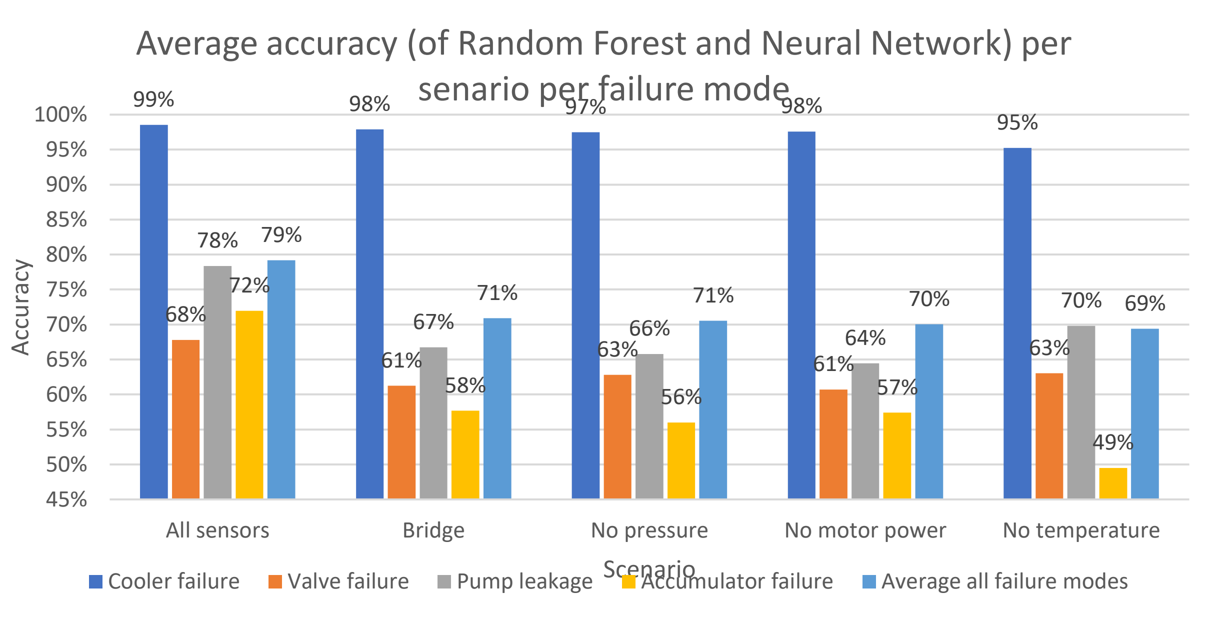


Figure 42 Average accuracy of Neural Network and Random Forest, compared over all scenarios for each failure mode

In Appendix K, three additional scenarios are discussed, representing each of the sensor groups as the only input. A conclusive graph is shown in Figure 43. From there, it can be concluded that the temperature sensors are critical for the predictive models. Pressure sensors are also important; however, the motor power sensor shows low accuracy predictions.

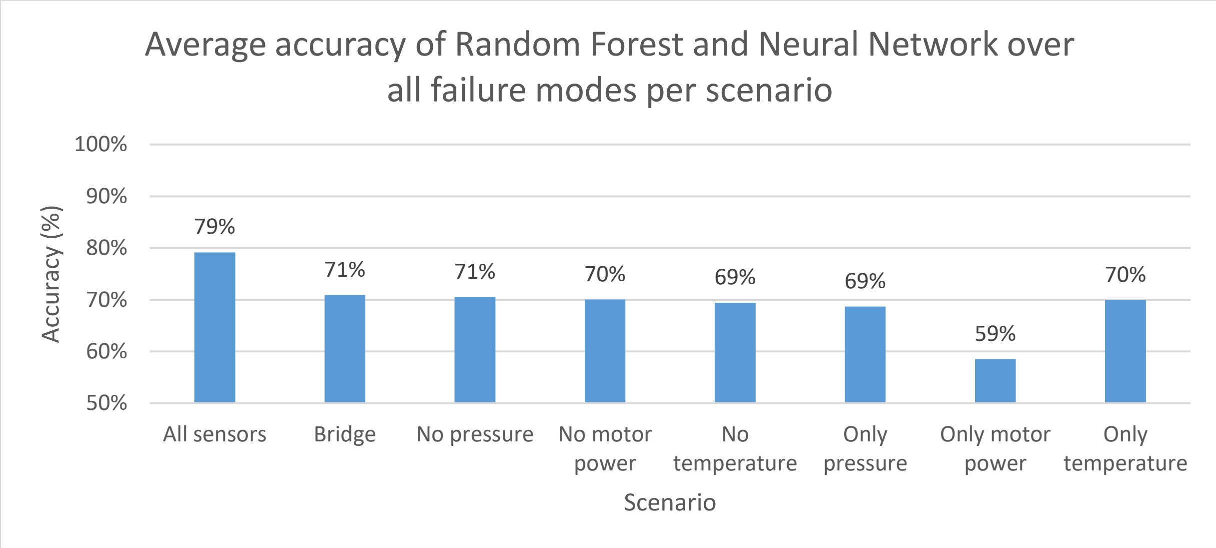


Figure 43 Average accuracy of Random Forest and Neural Network over all failure modes per scenario

For the second part of this analysis, the sensors are assessed individually rather than in groups. This is done based on the random forest model, as this model is explainable based on the input data and still performs good, whilst the Neural Network is not explainable, also known as black box, and other models perform worse.
The For the second part of this analysis, the sensors are assessed individually rather than in groups. This is done based on the random forest model, as this model is explainable based on the input data and still performs well, whereas the neural network is not explainable, also known as a black box, and other models perform worse. The sensors are ranked based on their Gini importance. In random forests, the Gini importance of an input data feature is calculated by computing the total reduction of the Gini impurity achieved by splitting the data on that feature across all trees in the forest. This means that a feature with high Gini importance has strong predictive power and is frequently used in the decision-making process of the model.

The importance of the sensors considered in all sensor scenarios is visible in Figure 44, where the sensors are color coded for the sensor groups. As can be seen, cooler failure is largely dependent on the cooling efficiency sensor. The valve failure has no clear important sensor, although cooling efficiency is important and most of the sensors in the sensor group temperature are important. For predicting pump leakage, volume flow sensors are important, and for accumulator failure, the importance of temperature sensors can be seen, although the importance of individual sensors is somewhat lower.

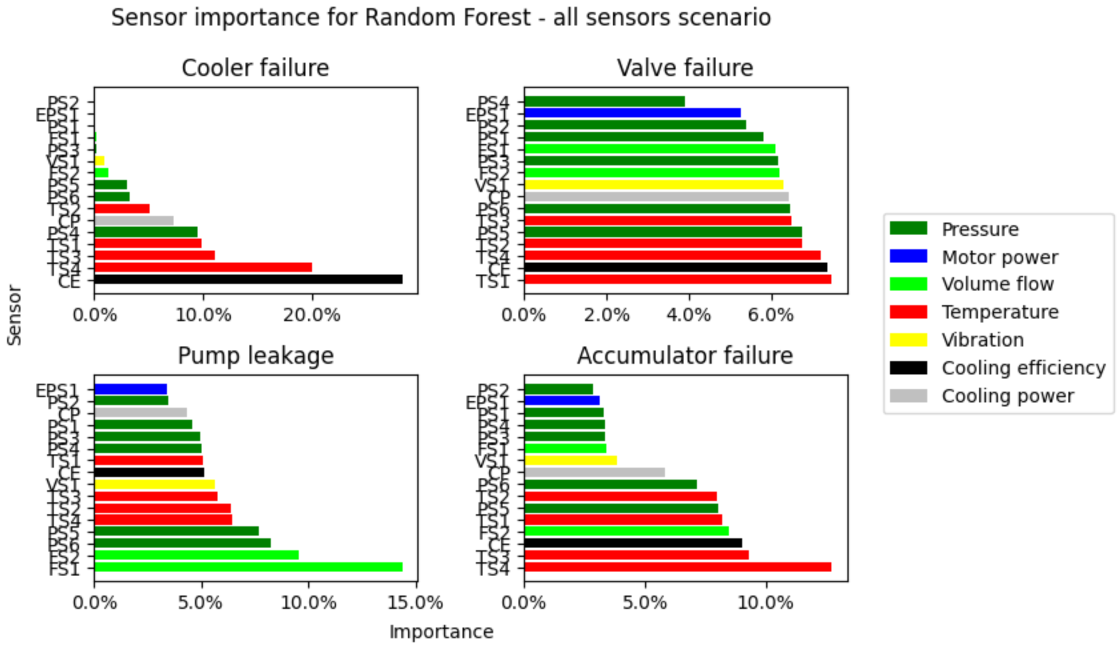


Figure 44 Importance of all individual sensors from the all sensors scenario in the random forest predictive model

The importance is also calculated when fewer sensors are available, as is the case in the bridge scenario. The importance of this is visible in Figure 45. From here again, cooler failure is based mainly on one sensor, and with the absence of cooling efficiency, this is temperature sensor 4. This sensor is also important for determining valve failure, which is still dependent on a lot of sensors. The pump leakage failure method shows a pattern where no sensor group is important, but the combination of sensors is. Accumulator failure is largely dependent on temperature sensor 4 and the other temperature sensors.

Overall, temperature sensors 4 and 3 and pressure sensors 5 and 6 are of great importance for all failure methods. Notably, the motor power sensor is of low importance for all failure modes, as are the other pressure sensors.

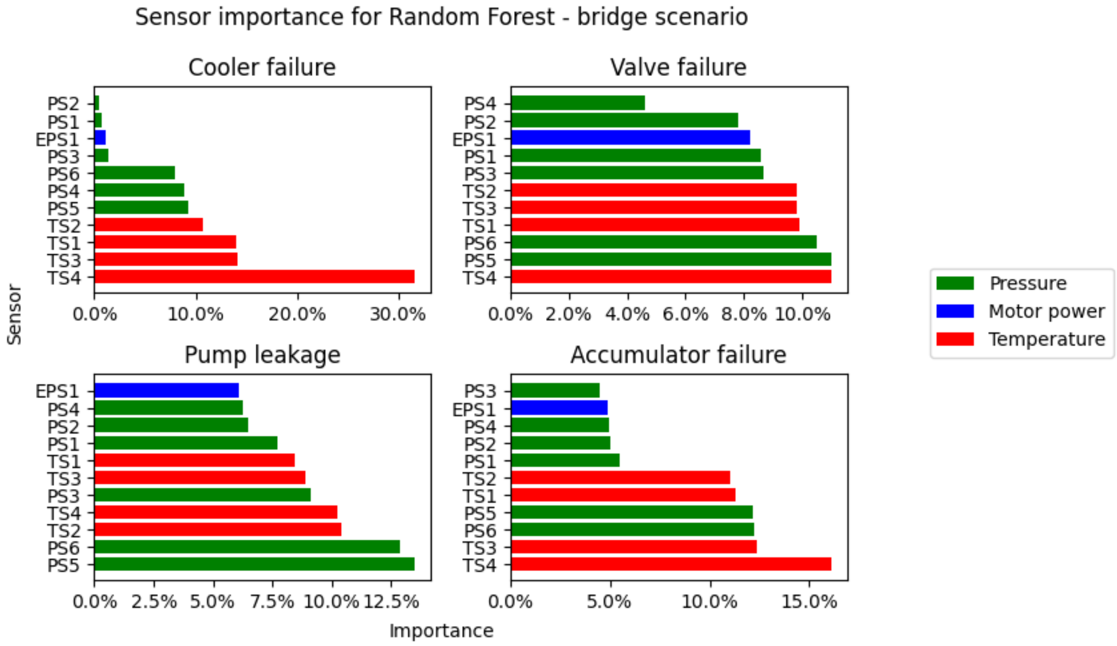


Figure 45 Importance of all individual sensors from the bridge scenario in the random forest predictive model

From the analyses above, it becomes apparent that some sensors are important for a good analysis, while other sensors are irrelevant and could have been left out. Leaving these sensors out can be a cost savings, but it is also important to compare the costs of having an additional sensor versus the expected benefits of the increased accuracy of predictions.

5.3.4 Connecting performance analysis and condition monitoring

In the last part of this condition monitoring analysis, we analyze the added value of predictive techniques over the currently used method of PM. To measure the improvement, we make use of a Monte Carlo simulation as it is a relative simple method to measure improvement.

The simulation consists of the steps described below, in addition there is also a figure available in Appendix L. In the simulation, we use 10,000 iterations, which means we calculate the averages of the performance indicators over the 10,000 iterations. In each iteration, we generate a sequence of 1,000 openings on the bridge. Each opening consists of five variables. The first four variables correspond to the conditions of the failure modes. For example, the first variable shows the condition of cooler failure in the current opening. Each of these condition variables is modeled using the information from the use case. So, with each opening, the condition can change, but it is based on the previous condition and changed randomly according to the failure distribution based on a Poisson distribution. The fifth and last variable corresponds to failure, which indicates whether the opening is a failure. This variable is based on the current condition of the bridge and the probability of random failures, which is set to occur once every 50 openings. So, a failure can occur because one of the failure mechanisms is in a critical state or because of randomness.

Next, a preventive maintenance policy is implemented. It is assumed that once maintenance occurs, every part is in perfect condition. The PM policy is usage-based, so it is dependent on the number of times a bridge is opened. To choose the best configuration, we must decide on the number of usages between maintenance occasions. This results in a tradeoff between availability and costs.

Availability can be determined by dividing the number of failed openings by the number of total openings. Costs can be calculated by multiplying the number of maintenance occasions by the costs per maintenance occasion. This has a linear relationship with the number of maintenance occasions; therefore, the number of maintenance occasions is used as KPI instead.

The comparison of different amounts of usage between maintenance occasions can be seen in Figure 46 and Figure 47. In the use case, we have an average availability of 95%; therefore, the number of usages between maintenance could be 50, as this corresponds with an availability of 94.2%, which is the closest. This also means that there are 20 maintenance occasions.

We can also implement a predictive model. We assume that the model can predict the correct condition with the accuracies from Section 5.3.2 of average over the neural network and random forest. If the model cannot guess the condition, the model does not plan any maintenance for this opening and will try to guess the condition in the next opening. If the predicted condition score falls below a certain threshold, the policy executes a maintenance event. This results in an availability of 97.8% and 19 maintenance occasions. Therefore, we can conclude that a PdM policy indeed shows better performance than the currently used policy as well as many other PM policies. Also, we have shown the relationship between condition monitoring, or maintenance policies in general, and performance analysis. As we do not generate any severity of failures we cannot reason to the risk matrix, however, the availability is higher and therefore the risk matrix is presumed to be more acceptable as well.

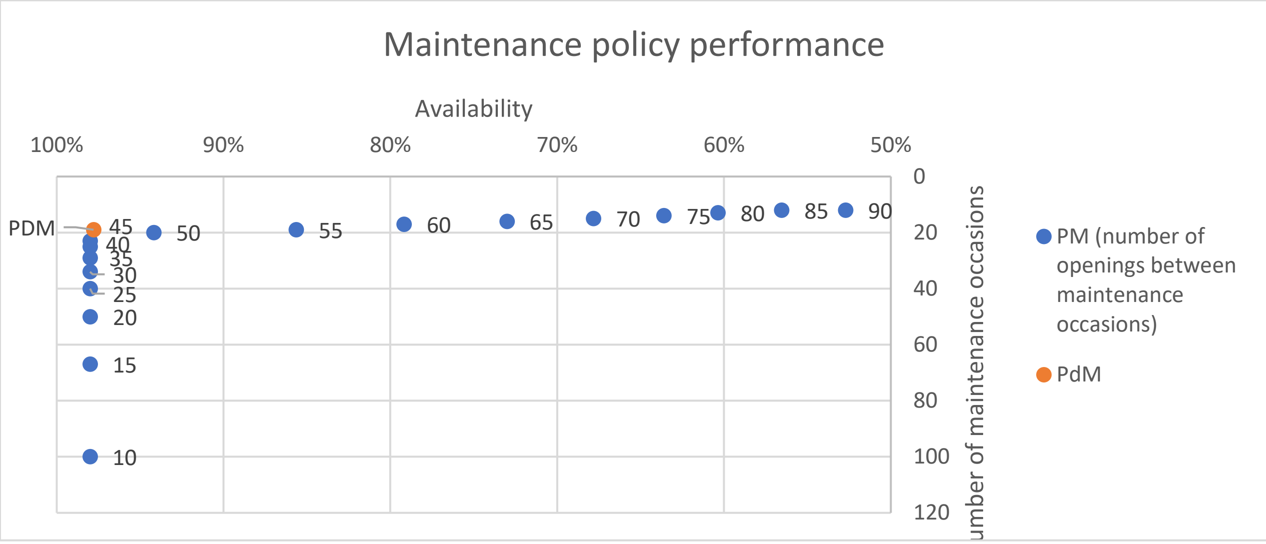


Figure 46 the performance of PM policies and a PdM policy

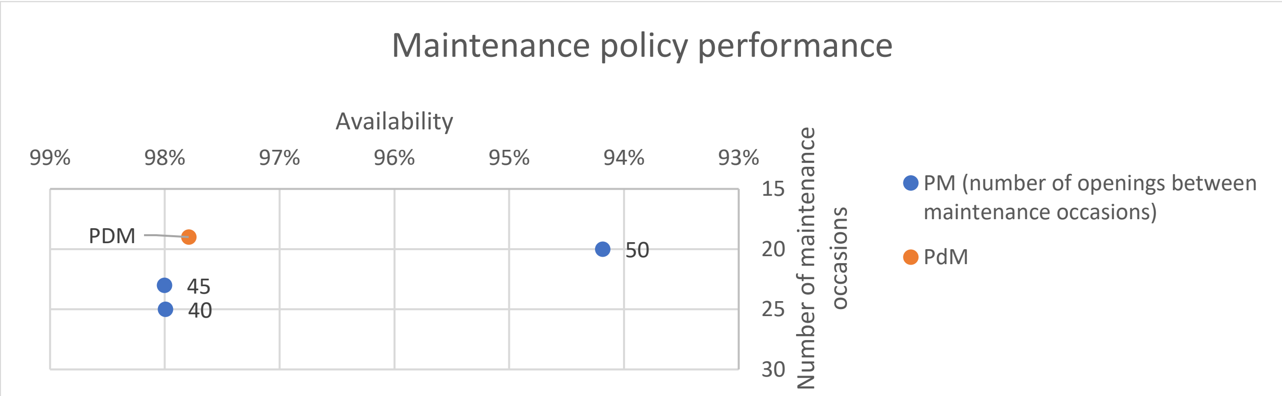


Figure 47 the performance of PM policies and a PdM policy - zoomed in

6 Conclusion and discussion

In the previous chapter, we validated the data model we created in Chapter 4. This data model is needed to connect the steps in the reference model. The model has been expanded with analyses on performance and condition monitoring. Additionally, these methods have been described and added to the Arcadis reference model. Now that we have executed all steps in this research, we present the conclusion and discussion in this chapter.

6.1 Conclusion

This research aimed to extend and improve the reference model of Arcadis by investigating data models, architecture, and data analysis methods following the research question. Through the creation of a data model by evaluating two key groups of static and dynamic data methods, an interface has been created between activities in the reference model. The data model shows the importance of asset classification as the backbone of asset management methods and links these methods together. The model also shows connectivity between FMECA and more dynamic data methods for example the relation between failure modes and condition monitoring or the relation between functions and performance analysis. All three: FMECA, performance analysis and condition monitoring and prediction involve functions and failure methods. Data architecture should facilitate providing this backbone through, for example, an object type library containing common concepts.

The reference model can be improved by including steps for performance analysis, such as availability or costs analysis. This methodology is dependent on performance indicators and can be used to link strategic steps taken by the asset owner of the reference model to the operational steps of the asset manager's role. Performance analysis can help identify relevant follow-up steps and the expected results of projects. Currently the use case is not operating within acceptable performance and therefore the data model helped in showing the need for further analysis.

Another data analytics method that shows improvement in AM is condition monitoring and prediction as part of predictive maintenance. There are various models successful in estimating the current condition of an asset based on sensor data. Most successful is a neural network consisting of long short-term memory network layers, it has the highest accuracy, with an average accuracy of 82.82% over four analyzed failure modes. Another method with high accuracy is the random forest, with an average accuracy of 75.71%. The benefits of random forest over the neural network are the computation time and the explainability of the model. Using a Monte Carlo simulation, we can show that predicting the current condition of the asset helps in improving the availability of the asset, whilst not increasing the costs of maintenance.

To conclude, the Arcadis reference model is expanded with a data model showing improvements in connectivity between activities and departments, in the importance of information architecture based on an OTL, and in possible methodologies to expand the capabilities of Arcadis asset management projects.

6.2 Discussion

This research developed a data model to complement Arcadis' asset management reference model, focusing on the three key aspects of risk, costs, and performance. The reference model is based on literature and includes interfaces, which are essential for standardization reference models. The data model is based on a case study of a movable bridge and uses methods like OTL, GIS, and FMECA. The model is expanded with dynamic data methods, including performance analysis and condition monitoring and prediction. The analysis reveals that the bridge's current performance is unacceptable, and improvements are needed to improve availability and costs. The final data model includes both static and dynamic data methods, enhancing the overall performance of the asset management process.

In line with Angelov et al. (2009), the data model shows various links between methods and belonging data, as also described in the reference model. The data model extends and improves the reference model by explaining the underlying information needed to conduct the methods described in the reference model. The data model also helps in preparing the asset reference model through not yet described methods of performance analysis and condition monitoring and prediction. Both are largely based on data analytics. The data model stimulates the importance of asset classification as the backbone of asset management processes and a common understanding in the organization and communication.

The research contributes to a clearer understanding of the literature on AM practices. It shows the link stipulated by Pudney (2010), that costs, performance, and risks play a crucial role in all the processes, and a balance should be found between these. Practice showed the need for a data model or an interface that is in line with the reference architecture type provided by Angelov et al. (2009). The ISO 55000 standard has been tested on a use case for linking the strategic impact of asset management practices to operational data. As stated by Tinga (2013), predictive maintenance reduces the number of maintenance occasions and costs while increasing performance. Although this comes at the increasing cost of monitoring instruments.

From the analyses, it became clear that it is important to find a balance between the value of extra information and its costs. It could be wise to invest in PdM policies, although this comes at an additional cost of monitoring. The same holds for performance analysis, where the risk matrix can be used as a guideline for which costs should be acceptable. The risk matrix also shows the link between strategic goals and operational impact on these goals, linking business strategy to operational performance. Using the strategy can shape the operational policies and improvements that should be made.

There are some limitations to this research. First is the limited number of use cases, as only one use case is examined. If the number is increased, the model can be further validated and expanded. This somewhat limited generalizability, although the model has been formulated in a generalizable way, allows implementation for other infrastructure assets. Another limitation can be found in condition monitoring and prediction, where instead of real data, synthetic data was used, which was generated under perfect conditions. The same holds for the Monte Carlo simulation showing the comparison between PdM and PM. It is to be expected that data from real assets contains more noise, and therefore requires more extensive data preparation. Additionally, in the Monte Carlo simulation some assumptions are made based on the use case, for example the failure rate. This can be different for other assets and even real life. Lastly, as AM projects tend to take a long time, improvements because of the data model are difficult to validate. Validation is done based on a single use case rather than exact variables like time saved, which can show the improvement more exactly , although this would take more time to measure. Although these limitations exist, this research still succeeded in answering the research question by providing a first iteration of a data model on which future expansions can be made.

Further research should take the role of data quality into account, also following The Institute of Asset Management (2015a), as data quality can have a significant impact on the reliability of the dynamic data analyses. A data quality framework could be used for this is the Total Data Quality Management framework developed by Wang (1998). Further research should also investigate testing the data model on other types of assets to validate the model. Another possibility for future research is including the data needed for the other steps of the reference model that have been excluded in this research, such as portfolio optimization. Further research should also consider evaluating the existing parts of the data models for performance analysis on other performance indicators, such as the other aspects of RAMS. Also, condition monitoring can be expanded to include visual inspections as well, rather than only sensor data. Lastly, GIS can be very complicated, and in this research only the basic aspects are included. To further extend the data model, the entire GIS method should be researched.

7 References

allesinkaagenbraassem. (2021, March 23). Leimuiderbrug deels afgesloten vanwege werkzaamheden. Alles in Kaag En Braassem. https://www.allesinkaagenbraassem.nl/nieuws/leimuiderbrug-deels-afgesloten-vanwege-werkzaamheden.html

Alvarado, S. H., Cortiñas, A., Luaces, M. R., Pedreira, O., & Places, A. S. (2022). Multilevel modeling of geographic information systems based on international standards. *Software and Systems Modeling*, *21*(2), 623–666. https://doi.org/10.1007/s10270021009011

Angelov, S., Grefen, P., & Greefhorst, D. (2009). A classification of software reference architectures: Analyzing their success and effectiveness. *Joint Working IEEE/IFIP Conference on Software Architecture, 2009 & European Conference on Software Architecture*, 141–150. https://doi.org/10.1109/WICSA.2009.5290800

Arunraj, N. S., & Maiti, J. (2007). Riskbased maintenance—Techniques and applications. *Papers Presented at the 2005 Symposium of the Mary Kay O'Connor Process Safety Center*, *142*(3), 653–661. https://doi.org/10.1016/j.jhazmat.2006.06.069

Ataspinar. (2018, April 4). *Machine Learning with Signal Processing Techniques*. ML Fundamentals. https://ataspinar.com/2018/04/04/machine-learning-with-signal-processing-techniques/

Aven, T. (2012). The risk concept—historical and recent development trends. *Reliability Engineering & System Safety*, *99*, 33–44. https://doi.org/10.1016/j.ress.2011.11.006

Beetz, J., & Borrmann, A. (2018). Benefits and limitations of linked data approaches for road modeling and data exchange. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *10864 LNCS*, 245–261. Scopus. https://doi.org/10.1007/9783319916385_13

Braaksma, A. J. J. (2012). *Asset information for FMEA-based maintenance* [Thesis fully internal (DIV), University of Groningen]. https://pure.rug.nl/ws/portalfiles/portal/10443594/AJJ_Braaksma_PhD_Thesis_181012_1.pdf

Davis, R. (2012). *An Introduction to Asset Management*. blah d blah design ltd.

Debell, M., Wilson, C., Jackman, S., & Figueroa, L. (2021). Optimal Response Formats for Online Surveys: Branch, Grid, or Single Item? *J Surv Stat Methodol*, *9*(1), 1–24. https://doi.org/10.1093/jssam/smz039

DIKW Intelligence. (2023). *Het DIKW model*. DIKW Intelligence. https://www.dikw.com/kennis/blogs/het-dikw-model

Einstein, J. (2018). *Natural Language Processing*.

Engineering News-Record. (2020). *ENR 2020 Top 150 Global Design Firms*. Www.enr.com. https://www.enr.com/toplists/2020-Top-150-Global-Design-Firms-Preview

Euzenat, J. & Shvaiko, P. (2013). *Ontology Matching*. Berlin, Heidelberg Springer.

Frické, M. (2009). The knowledge pyramid: A critique of the DIKW hierarchy. Journal of Information Science, 35(2), 131–142. Scopus. https://doi.org/10.1177/0165551508094050

Galster, M. (2015). Software Reference Architectures: Related Architectural Concepts and Challenges. *CobRA '15: Proceedings of the 1st International Workshop on Exploring Component-Based Techniques for Constructing Reference Architectures*, 5–8.

Harmelink, R.L.A. (2022). *Constructing the Service Control Tower* [EngD Thesis, University of Twente]. https://ris.utwente.nl/ws/portalfiles/portal/275069501/Constructing_the_Service_Control_Tower_Harmelink_PDEng_Thesis.pdf

Helwig, N., Pignanelli, E., & Andreas, S. (2018). *Condition monitoring of hydraulic systems*. UCI Machine Learning Repository.

Herder, P. M., & Wijnia, Y. (2012). A Systems View on Infrastructure Asset Management. In Van, P. Herder, & Y. Wijnia (Eds.), *Asset Management: The State of the Art in Europe from a Life Cycle Perspective* (pp. 31–46). Springer Netherlands. https://doi.org/10.1007/9789400727243_3

International Organization for Standardization. (2012). *Geographic information — Ontology (ISO 19150:2012)*. https://www.iso.org/standard/57465.html

International Organization for Standardization. (2014a). *Asset management - Management systems — Requirements (ISO 55001:2014)*. https://www.iso.org/standard/55089.html

International Organization for Standardization. (2014b). *Asset management - Overview, principles and terminology (ISO 55000:2014)*. https://www.iso.org/standard/55088.html

International Organization for Standardization. (2018a). *Asset management - Management systems — Guidelines for the application of ISO 55001 (ISO 55002:2018)*. https://www.iso.org/standard/70402.html

International Organization for Standardization. (2018b). *Condition monitoring and diagnostics of machines — General guidelines (ISO 17359:2018)*. https://www.iso.org/standard/71194.html

International Organization for Standardization. (2018c). *Risk management — Guidelines (ISO 31000:2018)*. https://www.iso.org/standard/65694.html

International Organization for Standardization. (2019). *Risk management — Risk assessment techniques (ISO 31010:2019)*. https://www.iso.org/standard/72140.html

ITF. (2021). Data-driven Transport Infrastructure Maintenance. International Transport Forum Policy Papers, No. 95. OECD Publishing, Paris.

Jensen, C. S., Pedersen, T. B., & Thomsen, C. (2022). Fundamental Concepts. In *Multidimensional Databases and Data Warehousing*. Springer Nature.

Khan, F. I., & Abbasi, S. A. (1998). Techniques and methodologies for risk analysis in chemical process industries. *Journal of Loss Prevention in the Process Industries*, *11*(4), 261–277. https://doi.org/10.1016/s0950-4230(97)00051-x

Kokla, M., & Guilbert, E. (2020). A review of geospatial semantic information modeling and elicitation approaches. *ISPRS International Journal of GeoInformation*, *9*(3). Scopus. https://doi.org/10.3390/ijgi9030146

Kovalev, S. M., Tarassov, V. B., Dolgiy, A. I., Dolgiy, I. D., Koroleva, M. N., & Khatlamadzhiyan, A. E. (2018). Towards intelligent measurement in railcar online monitoring: From measurement ontologies to hybrid information granulation system. *Advances in Intelligent Systems and Computing*, *679*, 169–181. Scopus. https://doi.org/10.1007/9783319683218_18

Luiten, B., Böhms, M., Alsem, D., & O'Keeffe, A. (2019). Asset information management using linked data for the lifecycle of roads. *LifeCycle Analysis and Assessment in Civil Engineering: Towards an Integrated VisionProceedings of the 6th International Symposium on LifeCycle Civil Engineering, IALCCE 2018*, 1529–1536. Scopus. https://www.scopus.com/inward/record.uri?eid=2s2.085063964668&partnerID=40&md5=ff099500041546252eb5a22b299956ba

Muller, G. (2008). *A reference architecture primer*.

NASA. (2008). *RELIABILITY-CENTERED MAINTENANCE GUIDE*.

NEN. (2019). *Conditiemeting gebouwde omgeving (NEN 2767:2019)*. https://www.nen.nl/nen-2767-1-c1-2019-nl-256366

Pudney, S. (2010). *Asset Renewal Decision Modelling with Application to the Water Utility Industry* [PhD Thesis]. https://eprints.qut.edu.au/40933/1/Steven_Pudney_Thesis.pdf

Rijkswaterstaat. (2010). *Leidraad RAMS - Sturen op prestaties van systemen*.

Ruitenburg, R. (2017). *Manoeuvring physical assets into the future – planning for predictable and preparing for unpredictable change in Asset Life Cycle Management* [PhD thesis]. https://doi.org/10.3990/1.9789036543958

Ruitenburg, R. J., Braaksma, A. J. J., & van Dongen, L. A. M. (2017). Asset life cycle plans: twelve steps to assist strategic decision-making in asset life cycle management. In *Optimum decision making in asset management* (pp. 259–287). IGI Global. https://doi.org/10.4018/978-1-5225-0651-5.ch012

SAP. (2023). *What is data modeling? | Definition, importance, & types*. SAP. https://www.sap.com/products/technology-platform/datasphere/what-is-data-modeling.html

Sherwin, D. J. (2000). A review of overall models for maintenance management. *Journal of Quality in Maintenance Engineering*, *6*, 138–164.

Snijder, H. H., & Hesselink, B. H. (2017). Repair, Strengthening and upgrading of steel bridges in The Netherlands. *IABSE Conference, Vancouver 2017: Engineering the FutureReport*, 1177–1184. Scopus. https://www.scopus.com/inward/record.uri?eid=2s2.085050036922&partnerID=40&md5=741cd59703f5b98bb40d106276357eaf

The Institute of Asset Management. (2002). *International Infrastructure Managament Manual*. UK Institute of Asset Management.

The Institute of Asset Management. (2015a). *An Anatomy of Asset Management*. https://theiam.org/media/1486/iam_anatomy_ver3_web-3.pdf

The Institute of Asset Management. (2015b). *Asset Information, Strategy, Standards and Data Management*.

Tinga, T. (2013). *Principles of Loads and Failure Mechanisms - Applications in Maintenance, Reliability and Design* (1st ed.). London Springer.

Voerknecht, H. (2017). *De wegen slibben dicht. Hoe komt CROW tot die conclusie?* CROW. https://www.crow.nl/kennis/bibliotheek-verkeer-en-vervoer/kennisdocumenten/de-wegen-slibben-dicht,-hoe-komt-crow-aan-die-info#:~:text=In%20het%20algemeen%20wordt%20als,miljard%20per%20jaar%20in%202021.

Wang, R. Y. (1998). A product perspective on total data quality management. Communications of the ACM, 41(2), 58–65. https://doi.org/10.1145/269012.269022

Wahid, A., Breslin, J. G., & Intizar, M. A. (2022). Prediction of Machine Failure in Industry 4.0: A Hybrid CNN-LSTM Framework. Applied Sciences, 12(9), 4221. https://doi.org/10.3390/app12094221

Weber, E., Papadopoulos, D. P., Lapedriza, A., Ofli, F., Imran, M., & Torralba, A. (2022). Incidents1M: a Large-Scale Dataset of Images With Natural Disasters, Damage, and Incidents. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1–14. https://doi.org/10.1109/tpami.2022.3191996

White, E. N. (1975). Terotechnology (physical asset management). *Mining Technology*, *57*(5), 293–297.

Wijnia, Y., & de Croon, J. (2015). The Asset Management Process Reference Model for Infrastructures. In J. AmadiEchendu, C. Hoohlo, & J. Mathew (Eds.), *9th WCEAM Research Papers* (pp. 447–457). Springer International Publishing.

Wu, P., Wang, J., Shemery, A., & Hampson, K. (2021). Digital Asset Information Management for Transport Infrastructure: Framework and Implementation. In *Advances in Science, Technology and Innovation* (pp. 413–418). Scopus. https://doi.org/10.1007/9783030484651_69

Appendix A. List of definitions with notes

This appendix contains the full list of definitions from Section 1.4 and includes the notes as described by the International Organization for Standardization (2014b).

- **Asset**

item, thing, or entity that has potential or actual value to an organization.
Note 1 to entry: Value can be tangible or intangible, financial or non-financial, and includes consideration of risks and liabilities. It can be positive or negative at various stages of the asset life.
Note 2 to entry: Physical assets usually refer to equipment, inventory and properties owned by the organization. Physical assets are the opposite of intangible assets, which are non-physical assets such as leases, brands, digital assets, use rights, licenses, intellectual property rights, reputation or agreements.
Note 3 to entry: A grouping of assets referred to as an asset system could also be considered as an asset

- **Risk**

effect of uncertainty on objectives
Note 1 to entry: An effect is a deviation from the expected — positive and/or negative.
Note2 to entry: Objectives can relate to different disciplines (such as financial, health and safety, and environmental goals) and can apply at different levels (such as strategic, organization-wide, project, product and process).
Note 3 to entry: Risk is often characterized by reference to potential "events" (as defined in ISO Guide 73:2009, 3.5.1.3) and "consequences" (as defined in ISO Guide 73:2009, 3.6.1.3), or a combination of these.
Note 4 to entry: Risk is often expressed in terms of a combination of the consequences of an event (including changes in circumstances) and the associated "likelihood" (ISO Guide 73:2009, 3.6.1.1) of occurrence.
Note 5 to entry: Uncertainty is the state, even partial, of deficiency of information related to, understanding or knowledge of, an event, its consequence, or likelihood.

- **Asset management**

coordinated activity of an organization to realize value from assets
Note 1 to entry: Realization of value will normally involve a balancing of costs, risks, opportunities and performance benefits.
Note 2 to entry: Activity can also refer to the application of the elements of the asset management system.
Note 3 to entry: to entry: The term "activity" has a broad meaning and can include, for example, the approach, the planning, the plans and their implementation.

- **Nonconformity**

non-fulfilment of a need or expectation that is stated, generally implied or obligatory

- **Corrective action (in this report also referred to as corrective maintenance)**

action to eliminate the cause of a nonconformity and to prevent recurrence
Note 1 to entry: In the case of other undesirable outcomes, action is necessary to minimize or eliminate the causes and to reduce the impact or prevent recurrence. Such actions fall outside the concept of corrective action, in the sense of this definition.

- **Preventive action (in this report also referred to as preventive maintenance)**

action to eliminate the cause of a potential nonconformity or other undesirable potential situation
Note 1 to entry: This definition is specific to asset management activities only.

Note 2 to entry: There can be more than one cause for a potential nonconformity.
Note 3 to entry: Preventive action is taken to prevent occurrence and to preserve an asset's function, whereas corrective action is taken to prevent recurrence.
Note 4 to entry: Preventive action is normally carried out while the asset is functionally available and operable or prior to the initiation of functional failure.
Note 5 to entry: Preventive action includes the replenishment of consumables where the consumption is a functional requirement.

- **Predictive action (in this report also referred to as predictive maintenance)**

action to monitor the condition of an asset and predict the need for preventive action or corrective action
Note 1 to entry: Predictive action is also commonly referred to as either "condition monitoring" or "performance monitoring".

Appendix B. Setup of surveys

This appendix contains the full survey setup. There are two versions of the survey, each containing different methods based on the domain the experts are operating in. Surveys are written in Dutch, as the respondents all speak Dutch.

Enquête: data gebruik binnen asset management door asset data engineers

Bedankt voor het openen van dit formulier!

In dit formulier stel ik verschillende vragen over methodieken en informatiegebruik binnen het asset management proces.
De resulterende informatie wordt gebruikt voor mijn afstudeeronderzoek.
In dit onderzoek ga ik informatiegebruik en overgang van informatie binnen de verschillende asset management processen (performance analysis, risk management, solution planning etc.) verder onderzoeken en ontwikkelen. Als gevolg kunnen we mogelijk de processen beter op elkaar aan laten sluiten.

Het zou enorm helpen om deze enquête in te vullen, het invullen zal ongeveer 15 minuten kosten.

Indien u een vraag niet wilt beantwoorden, kunt u het antwoord leeg laten.
Informatie gegeven in de enquête zal geanonimiseerd gebruikt worden en er wordt rekening gehouden met confidentialiteit.
Uw naam en adviesgroep vraag ik voor peroonlijke referentie.
De enquête is gelinkt aan mijn Arcadis account, na mijn onderzoek zal ik dus geen toegang meer tot de vragen en antwoorden hebben.

Met vriendelijke groet,
Arjan van Laar

Informatie over uw functie

1. Wat is uw naam?
2. Onder welke adviesgroep valt u?
   a. Asset Management Wegen
   b. Asset Management Rail
   c. Asset Management Infra Objecten & Installaties
   d. Asset Data Management
   e. Information Management
3. Geef een korte omschrijving van uw functie.
4. Beschrijf puntsgewijs een typisch project.
5. Hierbij het procesmodel van Arcadis, in welke van de Asset Manager processtappen plaatst u uw typische werkzaamheden?
   Meerdere keuzes zijn mogelijk
   a. Performance analysis
   b. Risk Management
   c. Solution Planning
   d. Portfolio optimization
   e. Portfolio delivery management
   f. Information management
   g. Performance management

GIS: Data Acquisition and Preparation

In deze en de volgende pagina's vraag ik per methode of u deze methode gebruikt, en indien u dit regelmatig gebruikt, wat voor tooling u gebruikt.

6. Gebruikt u de methode GIS: Data Acquisition and Preparation vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
7. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
8. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
9. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

GIS: Analysis and Trend Recognition

10. Gebruikt u de methode GIS: Analysis and Trend Recognition vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
11. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
12. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
13. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

GIS: Visualisation

14. Gebruikt u de methode GIS: Visualisation vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
15. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
16. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
17. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Data Integration and Structuring (API)

18. Gebruikt u de methode Data Integration and Structuring (API) vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
19. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).

20. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
21. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Time Series Analysis

22. Gebruikt u de methode Time Series Analysis vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
23. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
24. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
25. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Natural Language Processing

26. Gebruikt u de methode Natural Language Processing vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
27. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
28. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
29. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Computer Vision and Image Classification

30. Gebruikt u de methode Computer Vision and Image Classification vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
31. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
32. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
33. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Data Processing and Visualization (without GIS)

34. Gebruikt u de methode Data Processing and Visualization (without GIS) vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)

35. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
36. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
37. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Overige methode(n)

38. Indien u nog een andere methode vaak gebruikt zou u die hier kunnen noemen?
39. Indien u een methode *heeft genoemd*, zou u dan ook de onderstaande vragen willen beantwoorden. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
40. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
41. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Afsluiting

Bedankt voor het invullen van deze enquête.
U kunt nu op submit klikken om de antwoorden te versturen.

Enquête: data gebruik binnen asset management door maintenance engineers

Bedankt voor het openen van dit formulier!

In dit formulier stel ik verschillende vragen over methodieken en informatiegebruik binnen het asset management proces.
De resulterende informatie wordt gebruikt voor mijn afstudeeronderzoek.
In dit onderzoek ga ik informatiegebruik en overgang van informatie binnen de verschillende asset management processen (performance analysis, risk management, solution planning etc.) verder onderzoeken en ontwikkelen. Als gevolg kunnen we mogelijk de processen beter op elkaar aan laten sluiten.

Het zou enorm helpen om deze enquête in te vullen, het invullen zal ongeveer 15 minuten kosten.

Indien u een vraag niet wilt beantwoorden, kunt u het antwoord leeg laten.
Informatie gegeven in de enquête zal geanonimiseerd gebruikt worden en er wordt rekening gehouden met confidentialiteit.
Uw naam en adviesgroep vraag ik voor peroonlijke referentie.
De enquête is gelinkt aan mijn Arcadis account, na mijn onderzoek zal ik dus geen toegang meer tot de vragen en antwoorden hebben.

Met vriendelijke groet,
Arjan van Laar

Informatie over uw functie

1. Wat is uw naam?
2. Onder welke adviesgroep valt u?
    a. Asset Management Wegen
    b. Asset Management Rail
    c. Asset Management Infra Objecten & Installates
    d. Asset Data Management
    e. Information Management
3. Geef een korte omschrijving van uw functie.
4. Beschrijf puntsgewijs een typisch project.
5. Hierbij het procesmodel van Arcadis, in welke van de Asset Manager processtappen plaatst u uw typische werkzaamheden?
    Meerdere keuzes zijn mogelijk
    a. Performance analysis
    b. Risk Management
    c. Solution Planning
    d. Portfolio optimization
    e. Portfolio delivery management
    f. Information management
    g. Performance management

Root Cause Analysis

In deze en de volgende pagina's vraag ik per methode of u deze methode gebruikt, en indien u dit regelmatig gebruikt, wat voor tooling u gebruikt.

6. Gebruikt u de methode Root Cause Analysis vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
7. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
8. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
9. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Failure Mode, Effect and (Criticality) Analysis (FME(C)A)

10. Gebruikt u de methode FME(C)A vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
11. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
12. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
13. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Reliability Centered Maintenance (RCM)

14. Gebruikt u de methode RCM vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
15. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
16. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
17. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Fault Tree Analysis (FTA)

18. Gebruikt u de methode FTA vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
19. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
20. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
21. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Event Tree Analysis (ETA)

22. Gebruikt u de methode ETA vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
23. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
24. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
25. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Pareto Charts

26. Gebruikt u de methode Pareto Charts vaak in uw functie?
    f. Nooit (0/10 projecten)
    g. Zelden (1-3/10 projecten)
    h. Regelmatig (4-6/10 projecten)
    i. Vaak (7-8/10 projecten)
    j. (Vrijwel) Altijd (9-10/10 projecten)
27. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
28. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
29. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Bow Tie

30. Gebruikt u de methode Bow Tie vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
31. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
32. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
33. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Risk Register

34. Gebruikt u de methode Risk Register vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
35. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
36. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
37. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Risk Tree

38. Gebruikt u de methode Risk Tree vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
39. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
40. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
41. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

RAMS(SHEEP)

42. Gebruikt u de methode RAMS(SHEEP) vaak in uw functie?
    a. Nooit (0/10 projecten)
    b. Zelden (1-3/10 projecten)
    c. Regelmatig (4-6/10 projecten)
    d. Vaak (7-8/10 projecten)
    e. (Vrijwel) Altijd (9-10/10 projecten)
43. Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
44. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
45. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Overige methode(n)

46. Indien u nog een andere methode vaak gebruikt zou u die hier kunnen noemen?
47. Indien u een methode *heeft genoemd*, zou u dan ook de onderstaande vragen willen beantwoorden.
    Beschrijf de tooling die u gebruikt voor deze analyse (bijvoorbeeld Excel).
48. Beschrijf kort welke informatie u als input nodig heeft om de methode uit te voeren. Beschrijf uit welke tooling/database u deze informatie verzamelt.
49. Beschrijf kort welke informatie u als output van de methode krijgt. Beschrijf in wat voor tooling/vorm/database u deze informatie levert.

Afsluiting

Bedankt voor het invullen van deze enquête.
U kunt nu op submit klikken om de antwoorden te versturen.

Appendix C. Results of surveys

In this appendix, the results of the surveys are discussed in further detail. Each method asked in the surveys is compared to the process steps. Each method is discussed below. After the methods are discussed, an analysis per process step is done.

RCA

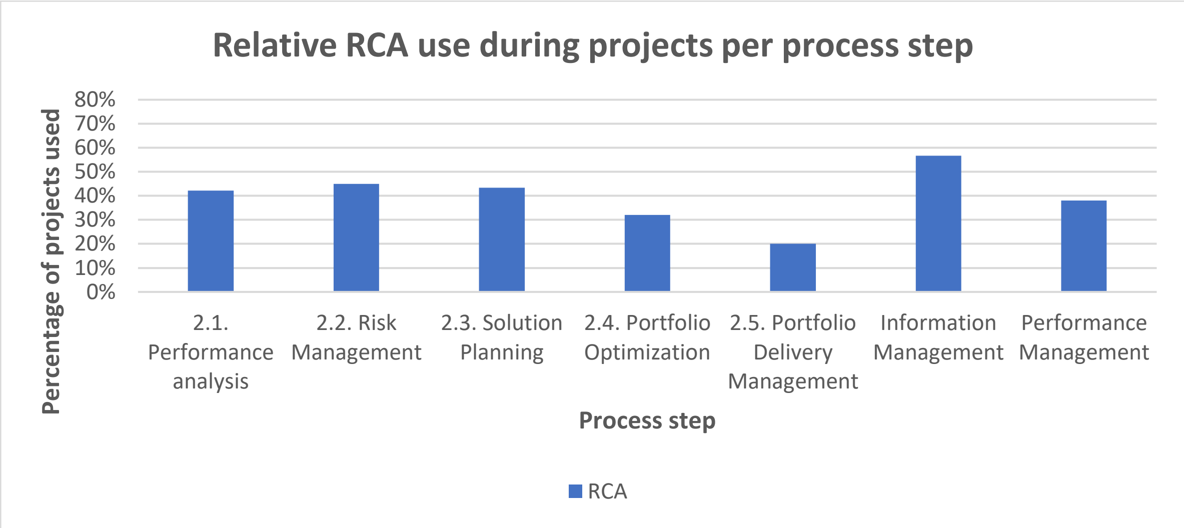


Figure 48 Relative use of RCA in the process steps

As can be seen in Figure 48, RCA is used most often by respondents involved with information management. As discussed in Section 2.3.3, RCA is currently not a very structured approach; therefore, it is logical that people who try to give logic to information use this method. The method is based on qualitative as well as quantitative data. It is often the basis for other methods.

FMECA

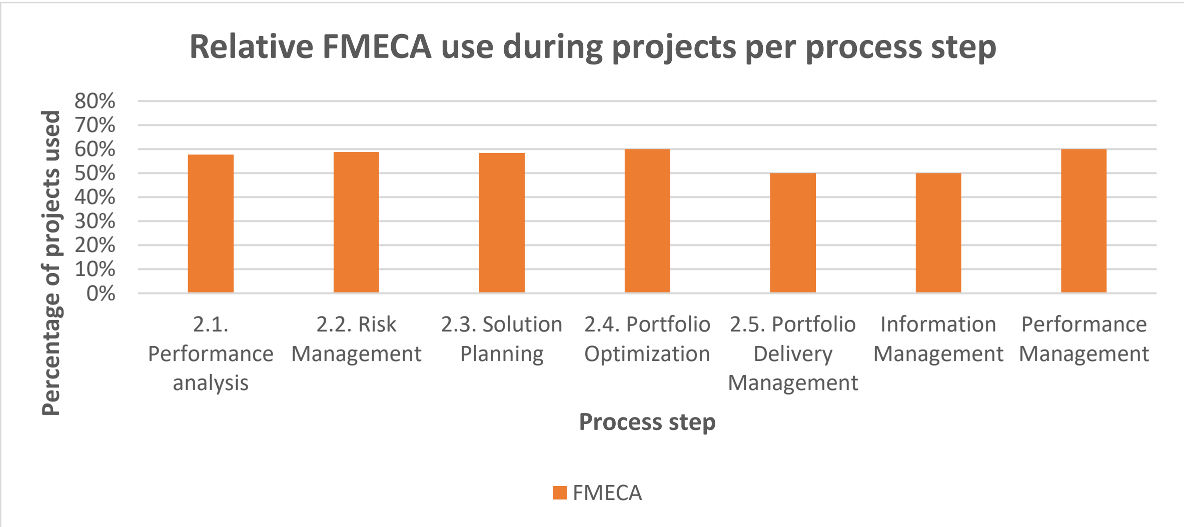


Figure 49 Relative use of FMECA in the process steps

In Figure 49, the use of FMECA in projects is visible per process step. As can be seen, FMECA is an often-used method in all the project steps, especially in performance management and portfolio optimization. FMECA delivers the failure methods that are used for portfolio optimization and gives the key indicators for performance management; therefore, it is logical to use this method in these process steps.

RCM

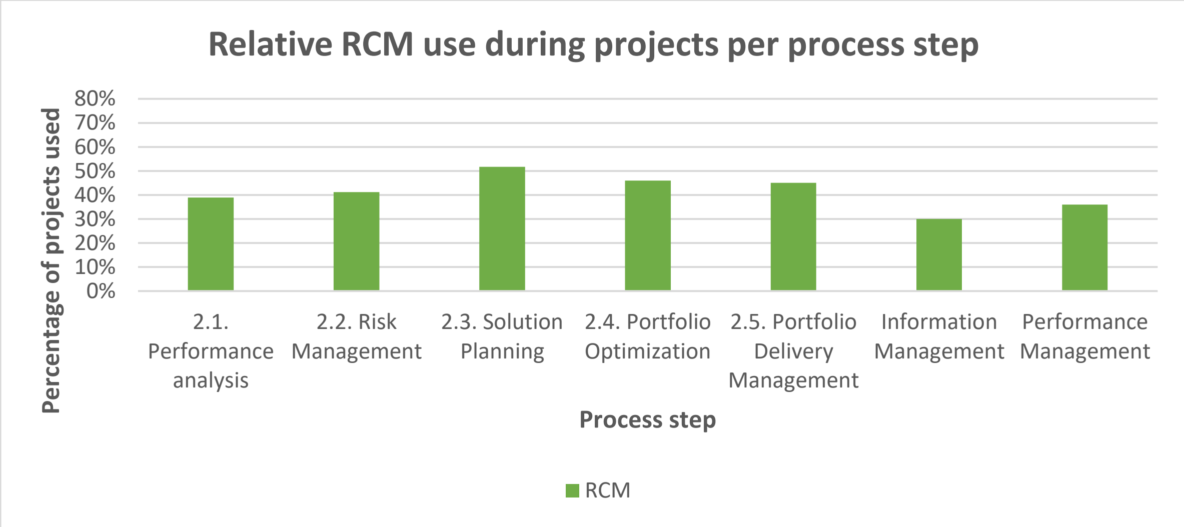


Figure 50 Relative use of RCM in the process steps

Figure 50 shows the use of the Reliability Centered Maintenance methodology in the process steps of the Arcadis reference model. As can be seen RCM, is also an often-used methodology, mainly in solution planning. RCM is an extension of FMECA, which also considers choosing the best solution; therefore, this is logical. Although more of a resemblance to the pattern of FMECA would be expected, a large percentage is in portfolio optimization and performance management.

FTA

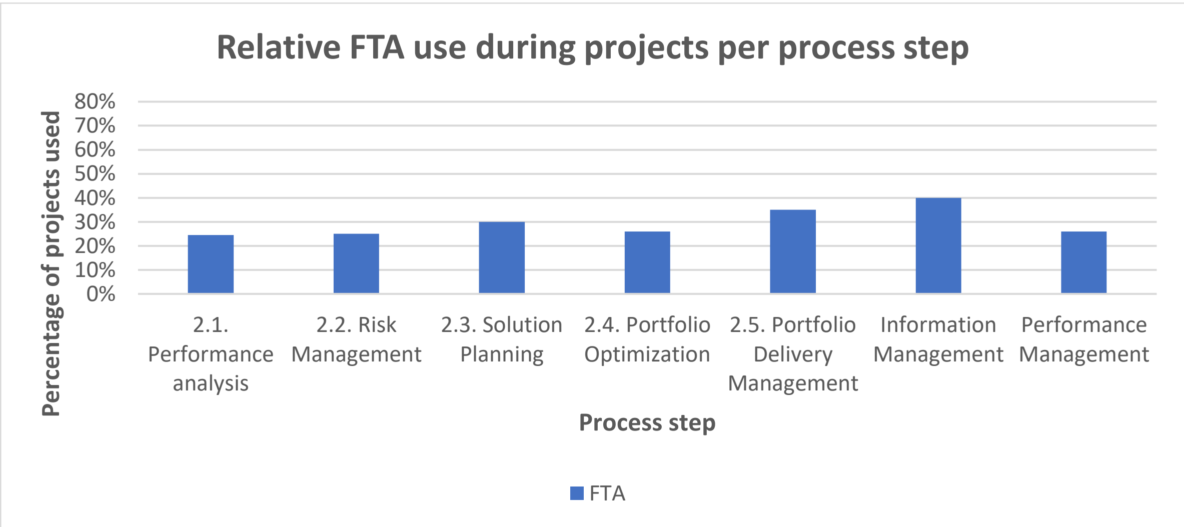


Figure 51 Relative use of FTA in the process steps

The use of Fault Tree Analysis in the process steps is visible in Figure 51. FTA is used less than the methods mentioned above. The largest use of FTA is by the respondents involved with information management.

ETA

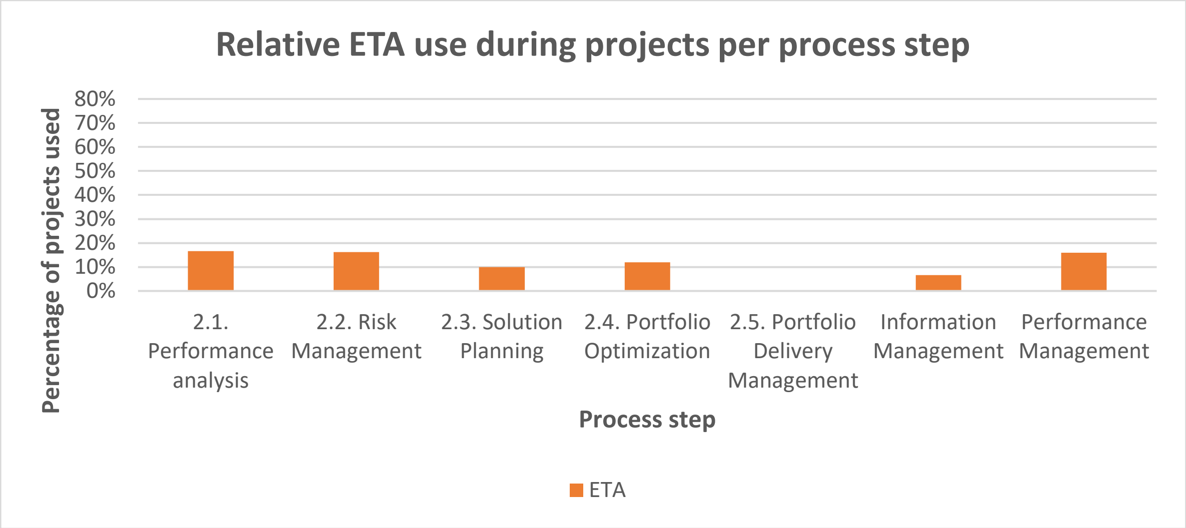


Figure 52 Relative use of ETA in the process steps

The use of Event Tree Analysis in process steps is shown in Figure 52. Here it can be seen that ETA is not commonly used in the process steps. It is most commonly used by respondents involved with performance analysis, risk management, and performance management.

PC

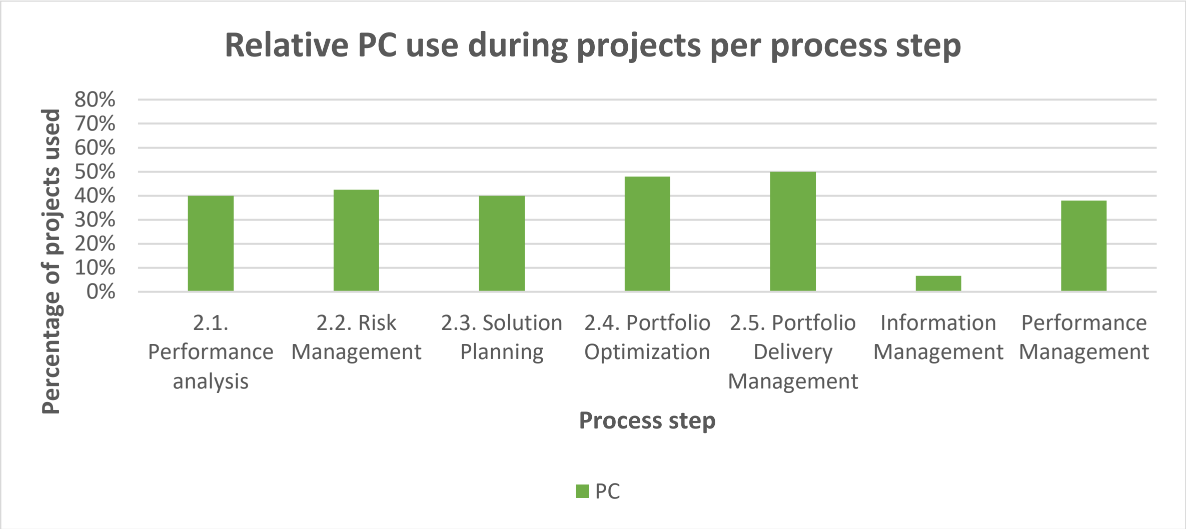


Figure 53 Relative use of PC in the process steps

In Figure 53, the use of the Pareto Chart is shown per process step. PC is commonly used, mainly by the respondents concerned with portfolio delivery management and portfolio optimization. It is logical that this method occurs here more often as PC is suited to assist in making decisions on portfolios with limited budget, which needs to be done in portfolio optimization.

BTA

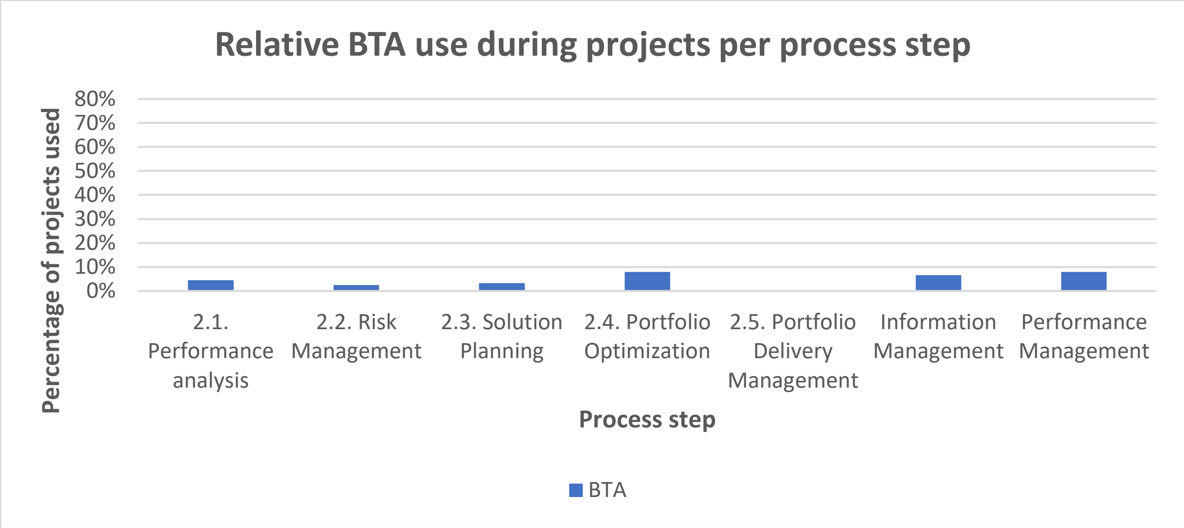


Figure 54 Relative use of BTA in the process steps

The relative use of Bow Tie analysis per process step is shown in Figure 54. The method is not used much, together with the Risk Tree methodology, it is the least used methodology. Its main use is in portfolio optimization.

RR

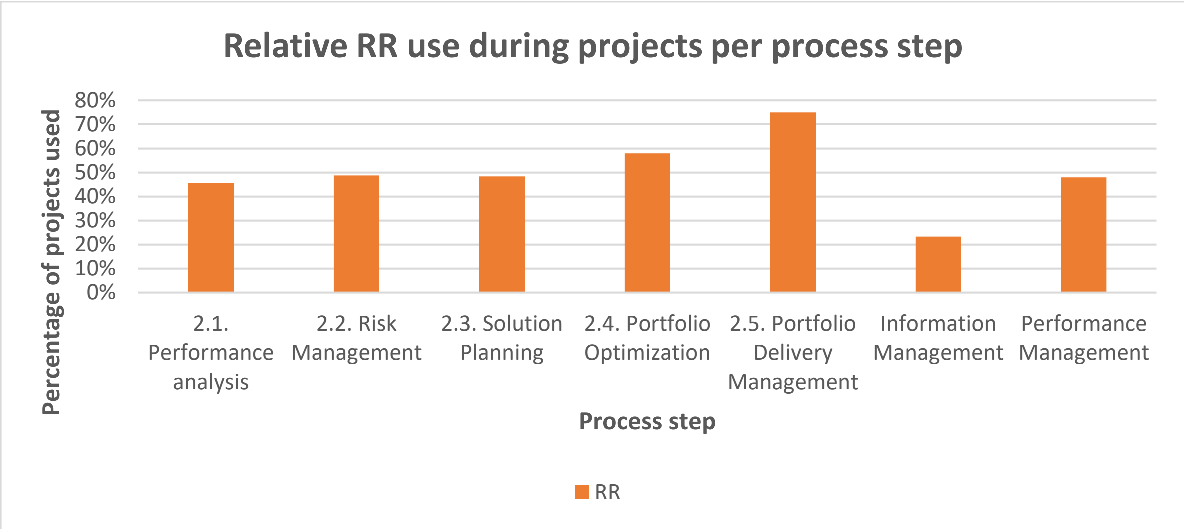


Figure 55 Relative use of RR in the process steps

The Risk Register is used relatively often in portfolio delivery management, as can be seen in Figure 55. Respondents concerned with delivering the portfolio to the customer use a risk register to make the customer aware of the risks in their assets and possibly some linked mitigation measures concerned with reducing or completely mitigating risks.

RT

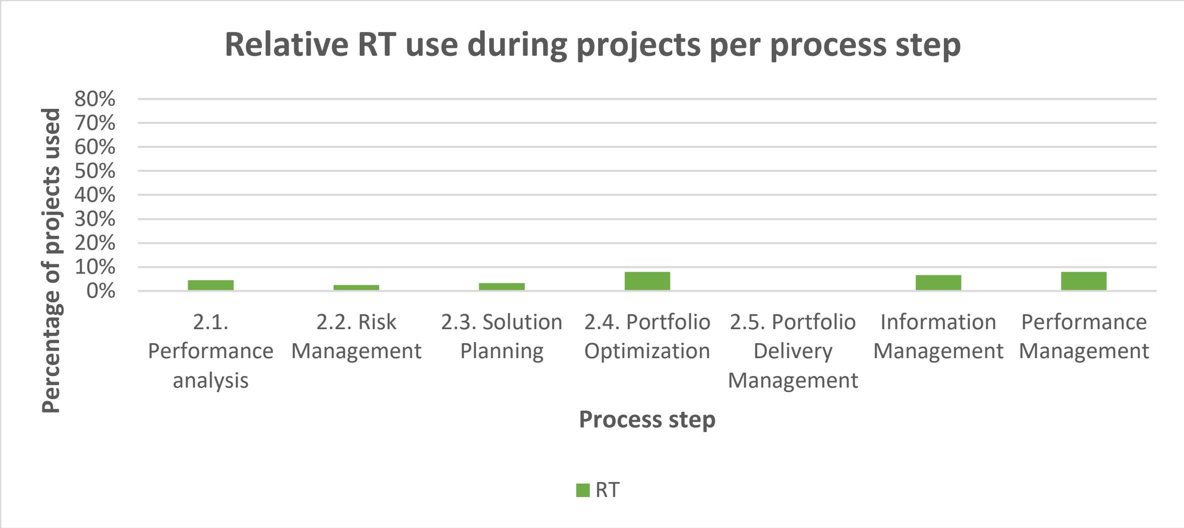


Figure 56 Relative use of RT in the process steps

As already mentioned, the Risk Tree methodology is one of the least used by the respondents. The relative use per process step is shown in Figure 56, where the main use is in portfolio optimization and performance management.

RAMS

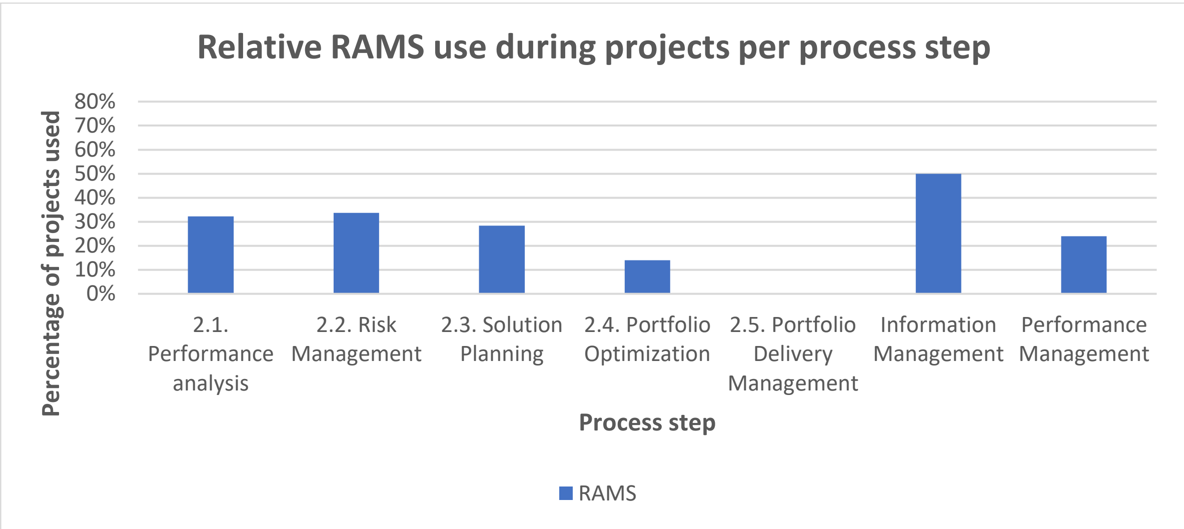


Figure 57 Relative use of RAMS in the process steps

The last discussed methodology is the RAMS(SHEEP) methodology. This is not necessarily a methodology but more of a set of principles that can be used to assess an asset. The relative use of these principles is visible in Figure 57. The main use is in information management.

GIS DAP

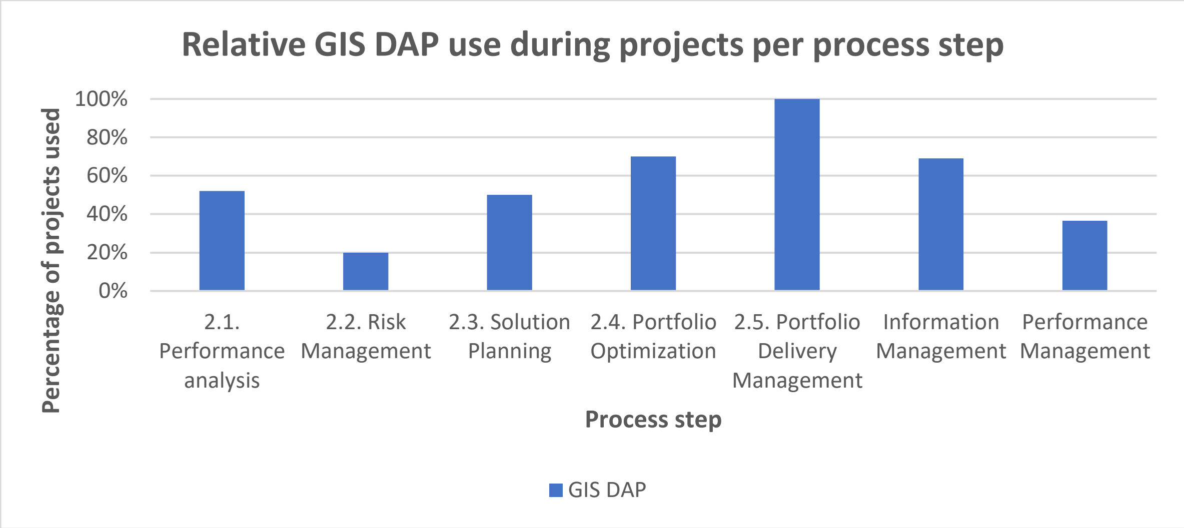


Figure 58 Relative use of GIS DAP in the process steps

The method of Geographic Information Systems Data Acquisition and Preparation is often used as can be seen in Figure 58, which shows the relative use per project step. The method is most often used by the respondent, who is working with portfolio delivery management. This is followed by portfolio optimization and information management.

GIS ATR

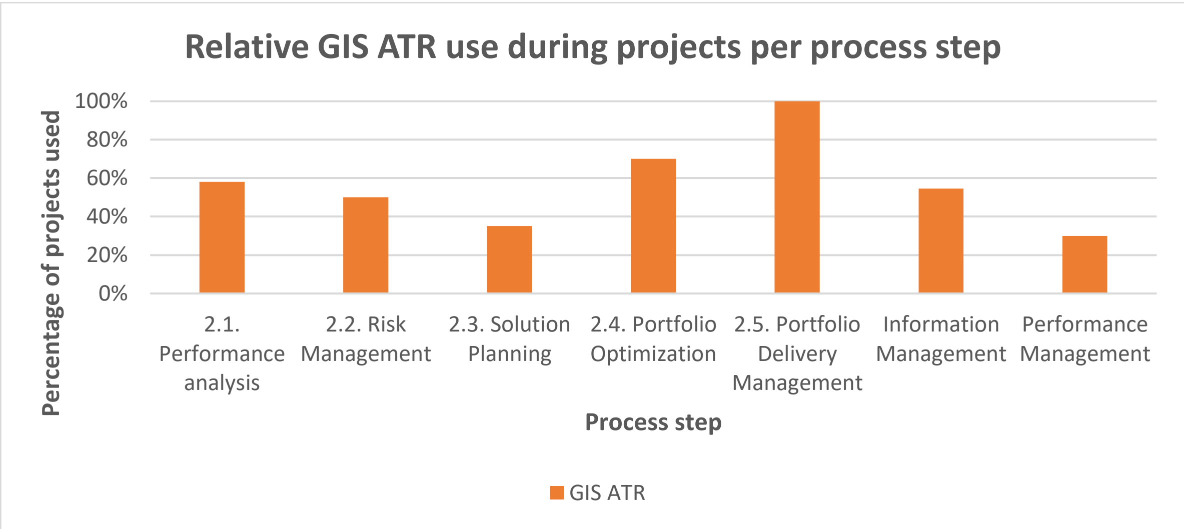


Figure 59 Relative use of GIS ATR in the process steps

Another often-used methodology is Geographic Information Systems Analysis and Trend Recognition. Its relative use is visible in Figure 59. Again, portfolio delivery management is the relatively most often mentioned process step. This is again followed by the portfolio optimization process step.

GIS V

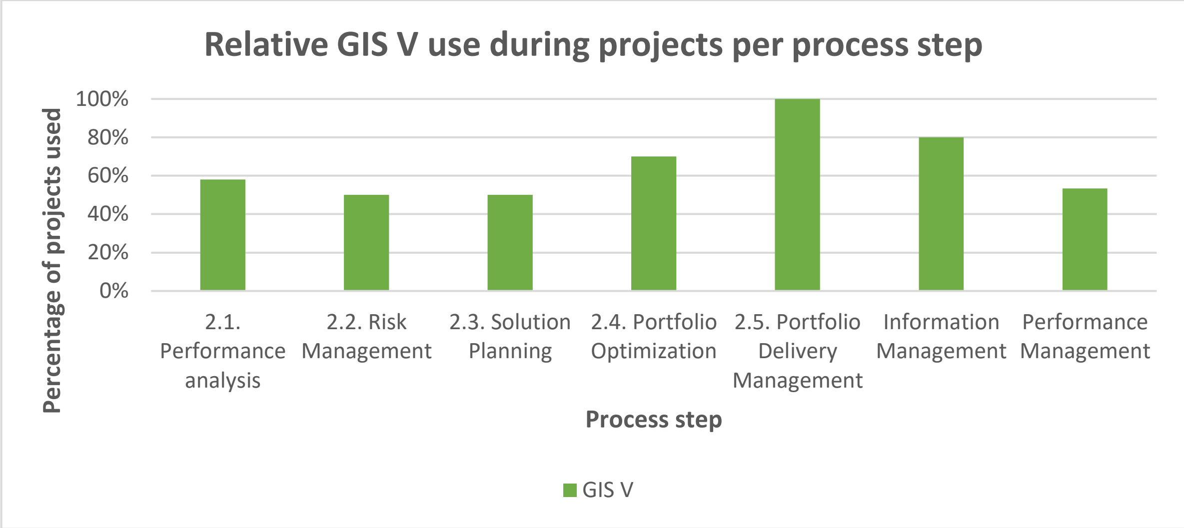


Figure 60 Relative use of GIS V in the process steps

The most acknowledged methodology is Geographic Information Systems Visualization. The relative use is visible in Figure 60. Again, the respondents that acknowledged portfolio delivery management are working often with GIS V. This is now followed by the process step of information management.

DIS

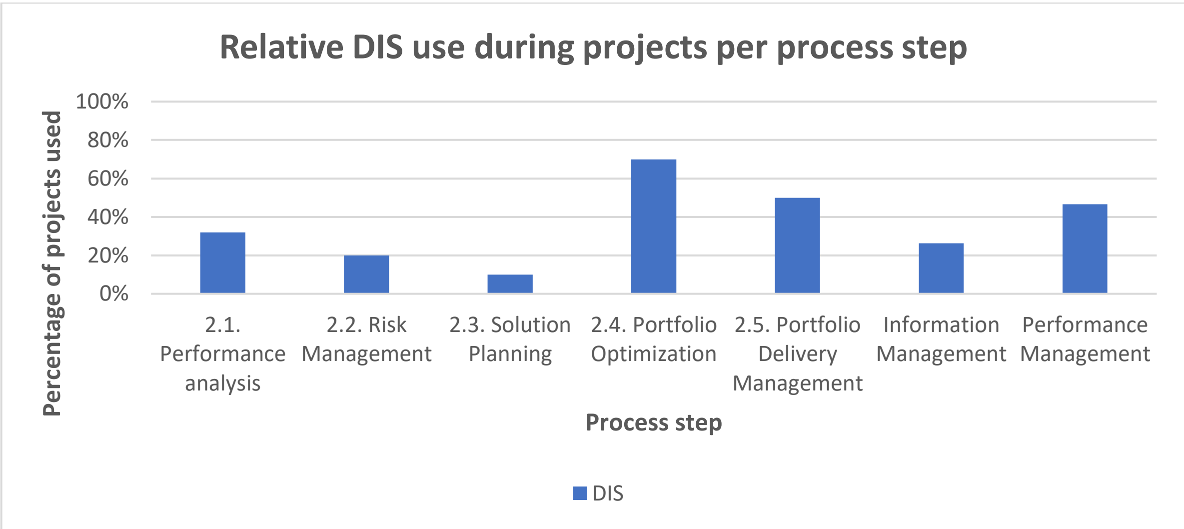


Figure 61 Relative use of DIS in the process steps

In Figure 61, the relative use of Data Integration and Structuring is shown. This method is most often by respondents who acknowledge portfolio optimization. Additionally, respondents working in portfolio delivery management and performance management also regularly use the method.

TSA

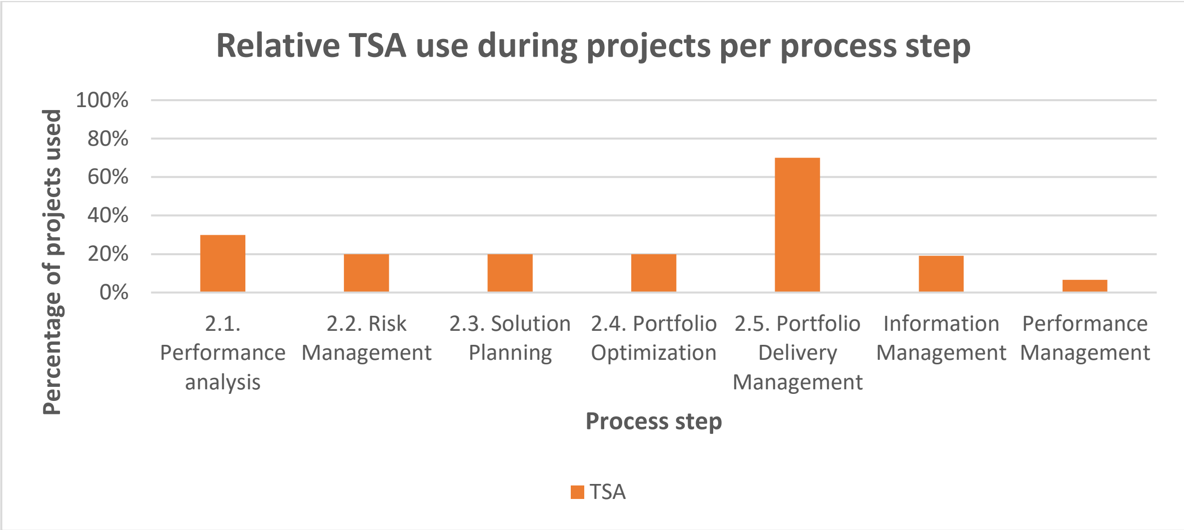


Figure 62 Relative use of TSA in the process steps

Time Series Analysis is another method used sometimes by employees at Arcadis. The relative use per process step is visible in Figure 62. Again, portfolio delivery management is relatively the most mentioned process step.

NLP

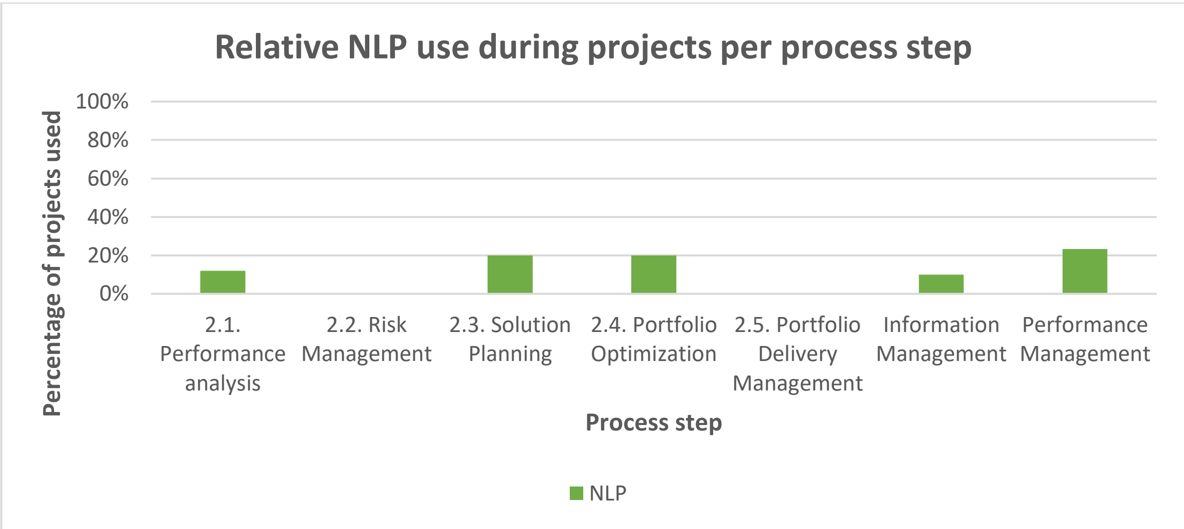


Figure 63 Relative use of NLP in the process steps

Natural Language Processing is a methodology where computer algorithms can transform text into useful information and data. The relative use of the method is visible in Figure 63. It can be concluded that NLP is the least often mentioned method by the respondents. It was mentioned most often by respondents involved with information management.

CVIC

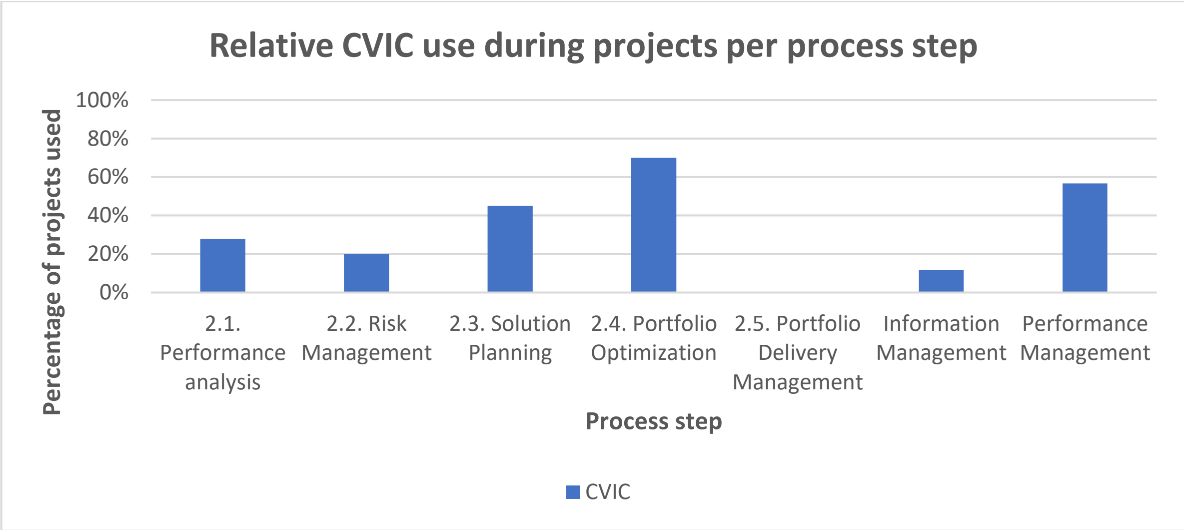


Figure 64 Relative use of CVIC in the process steps

Computer Vision and Image Classification is a method where computer algorithms identify objects in images and videos and analyze these objects. The relative use of the method in the process steps is available in Figure 64. Here it can be seen that respondents working in portfolio optimization are using the method most often, followed by performance management.

DPV

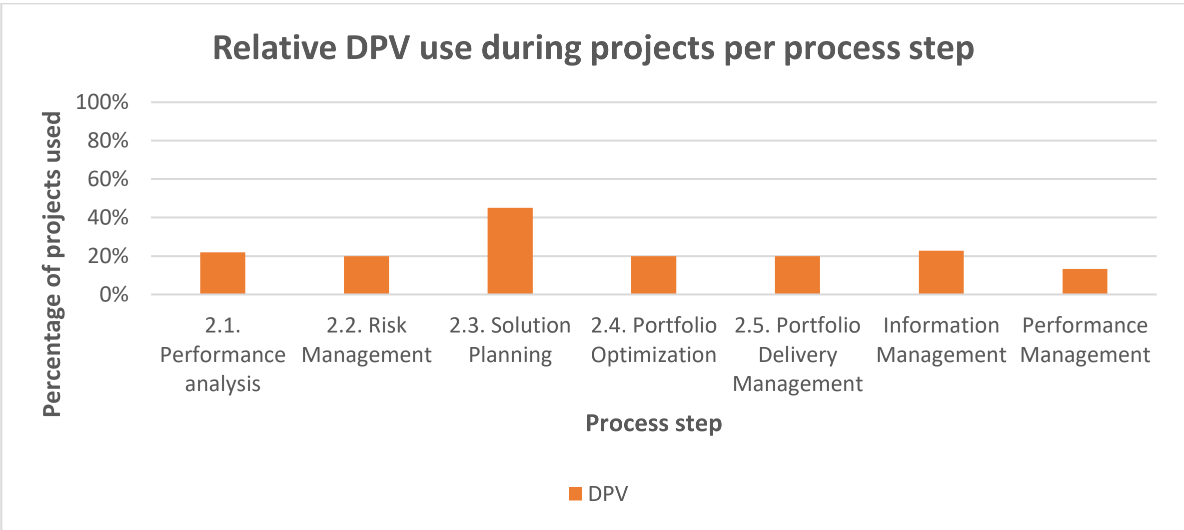


Figure 65 Relative use of DPV in the process steps

In Figure 65, the relative use per process step of the method Data Preparation and Visualization is shown. In this method, structured data is used to gather conclusions and visual representations of the data. It is used most by the respondent involved with solution planning.

2.1. Performance analysis

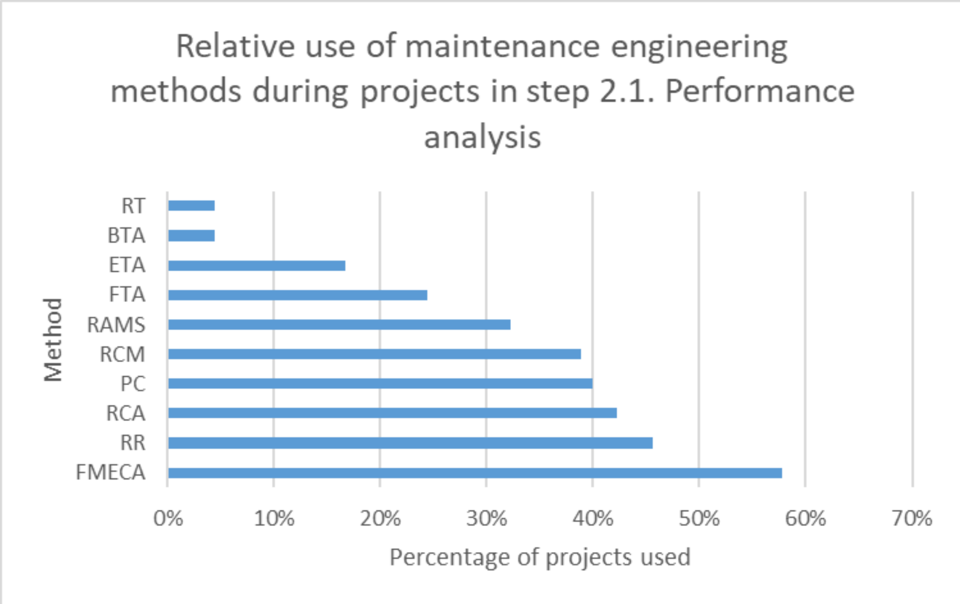


Figure 66 Relative use of the maintenance engineering methods in the process step 2.1. performance analysis

In Figure 66, the relative use per maintenance engineering method in the step performance analysis is shown. As can be seen for this step, FMECA is the most used method.

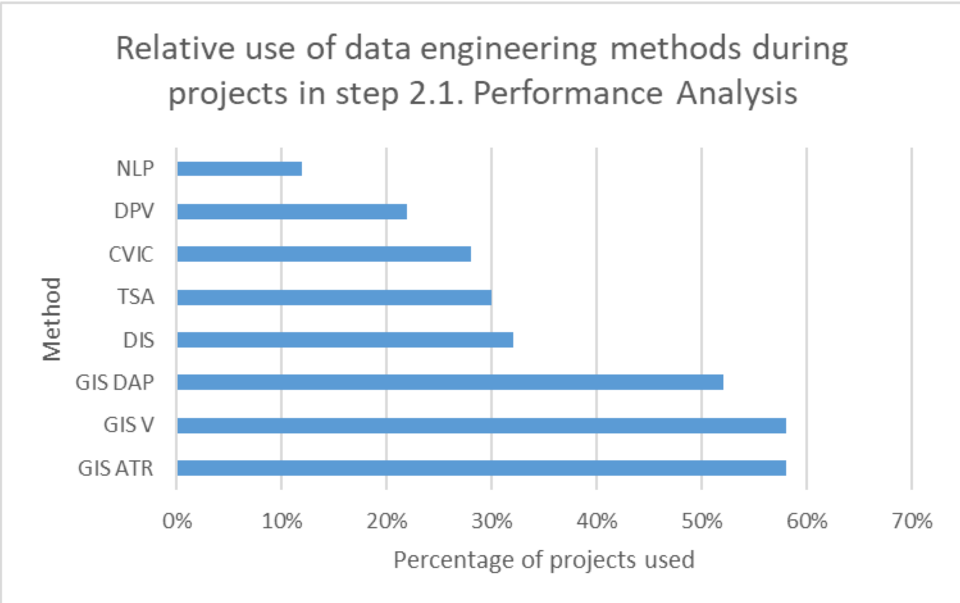


Figure 67 Relative use of the data engineering methods in the process step 2.1. performance analysis

In Figure 67, the relative use of the data engineering methods in the step performance analysis is shown. As can be seen for this step, GIS analysis and trend recognition is the most used method, together with GIS visualization.

2.2. Risk Management

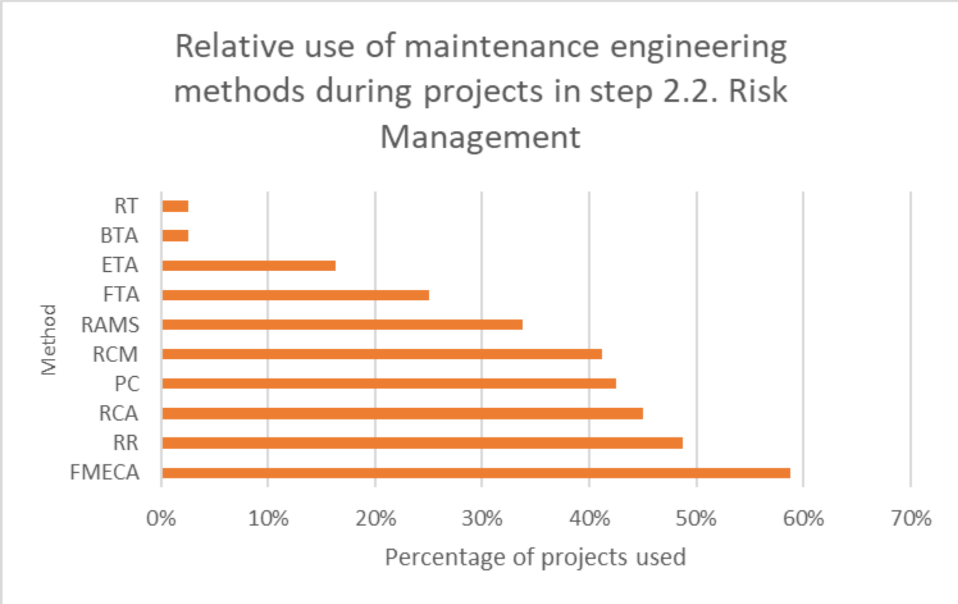


Figure 68 Relative use of the maintenance engineering methods in the process step 2.2. risk management

In Figure 68, the relative use of the maintenance engineering methods in the step risk management is shown. As can be seen for this step, FMECA again is the most used method.

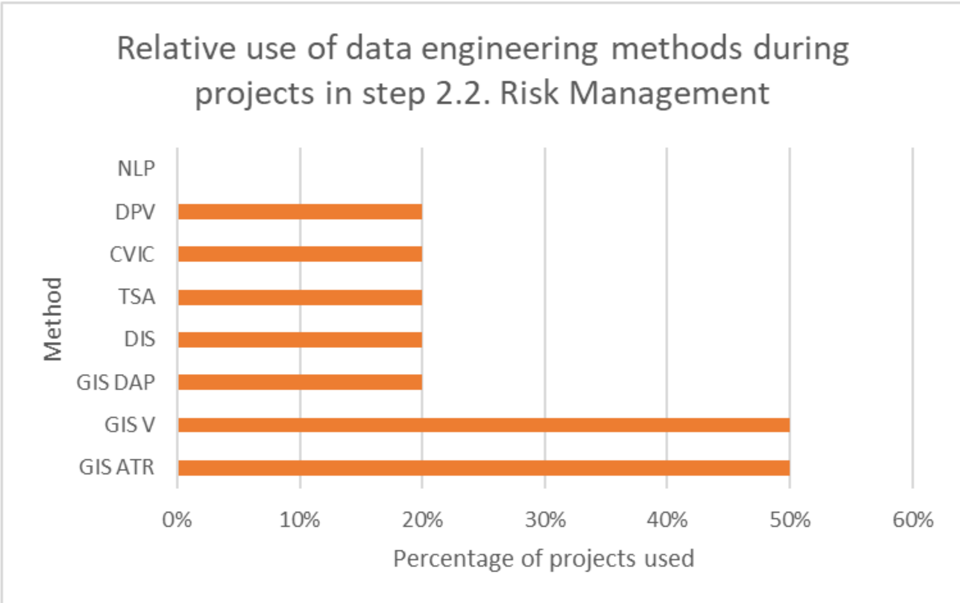


Figure 69 Relative use of the data engineering methods in the process step 2.2. risk management

In Figure 69, the relative use of the data engineering methods in the step risk management is shown. As can be seen for this step, GIS analysis and trend recognition and GIS visualization are again the most used methods.

2.3. Solution Planning

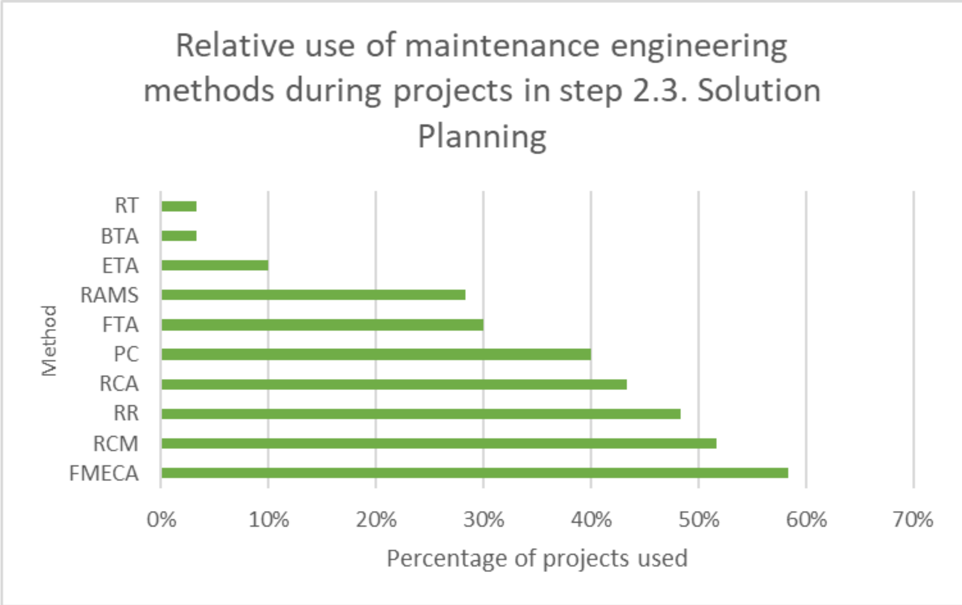


Figure 70 Relative use of the maintenance engineering methods in the process step 2.3. solution planning

In Figure 70, the relative use of the maintenance engineering methods in the step solution planning is shown. As can be seen for this step, FMECA again is the most used method.

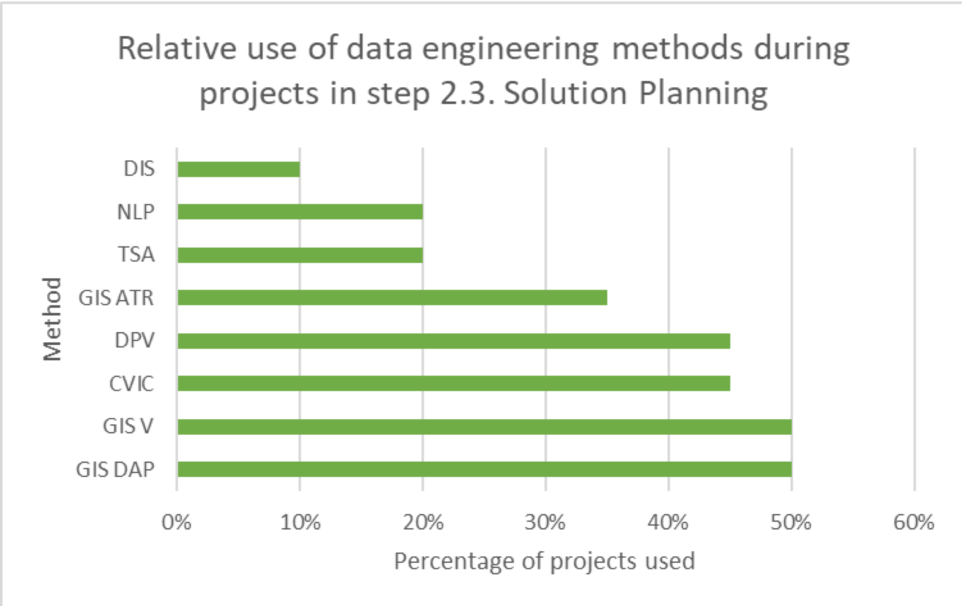


Figure 71 Relative use of the data engineering methods in the process step 2.3. solution planning

In Figure 71, the relative use of the data engineering methods in the step solution planning is shown. As can be seen for this step, GIS data acquisition and preparation and GIS visualization are the most used methods.

2.4. Portfolio Optimization

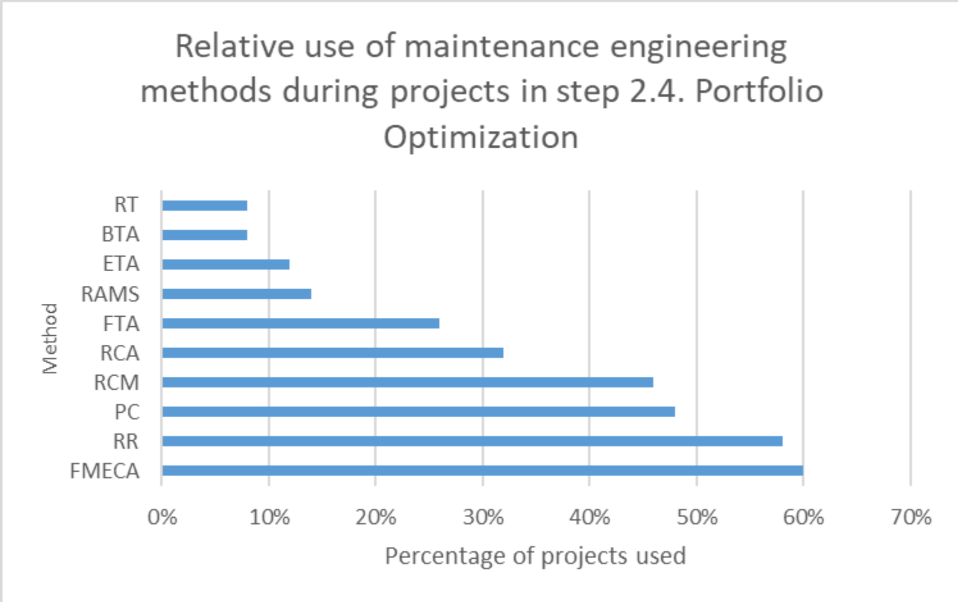


Figure 72 Relative use of the maintenance engineering methods in the process step 2.4. portfolio optimization

In Figure 72, the relative use for the maintenance engineering method, in the step portfolio optimization is shown. As can be seen for this step, FMECA again is the most used method.

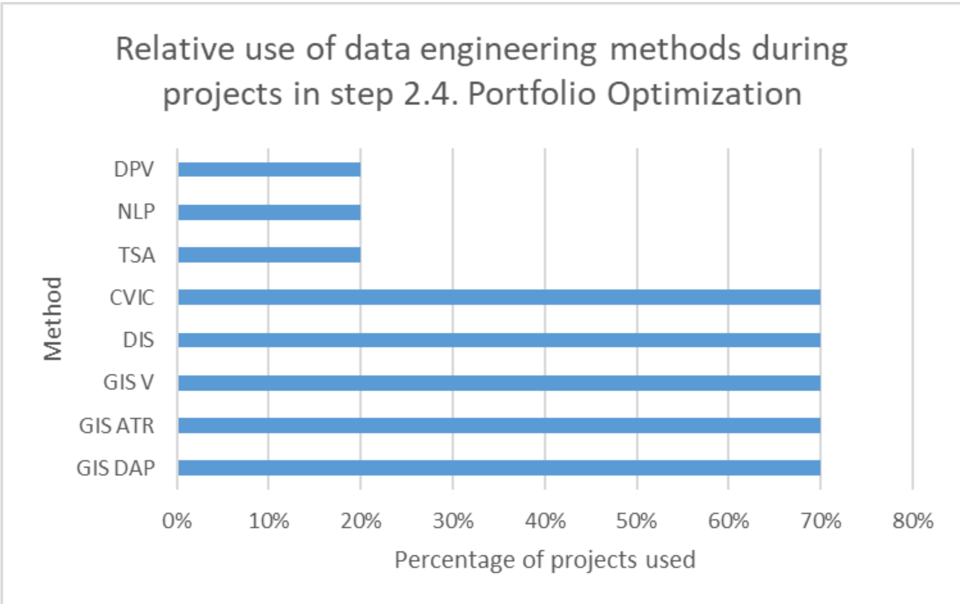


Figure 73 Relative use of the data engineering methods in the process step 2.4. portfolio optimization

In Figure 73, the relative use of the data engineering methods in the step portfolio optimization is shown. As can be seen for this step, five different methods are the most used.

2.5. Portfolio Delivery Management

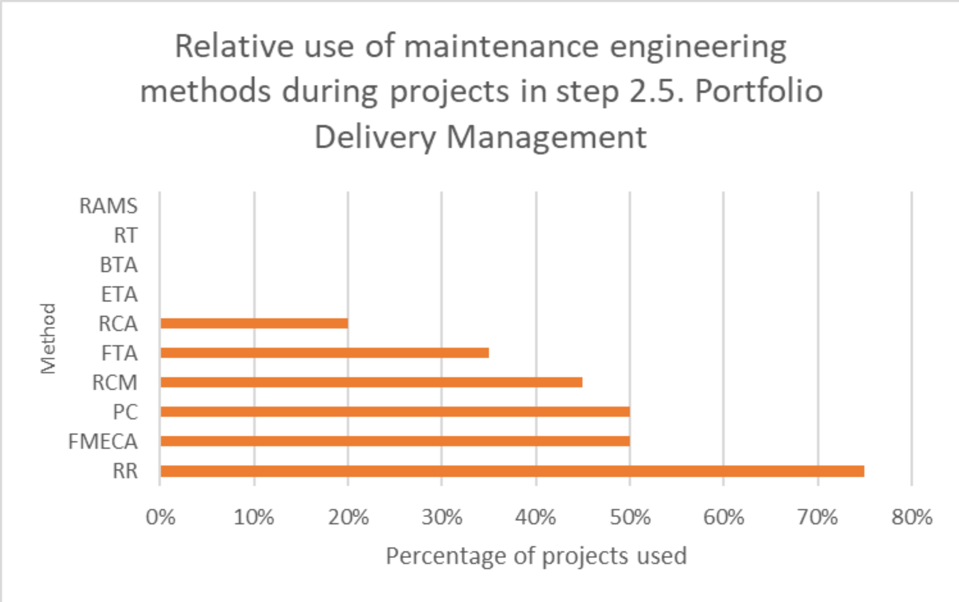


Figure 74 Relative use of the maintenance engineering methods in the process step 2.5. portfolio delivery management

In Figure 74, the relative use of the maintenance engineering methods in the step portfolio delivery management is shown. As can be seen for this step, Risk Register is the most used method.

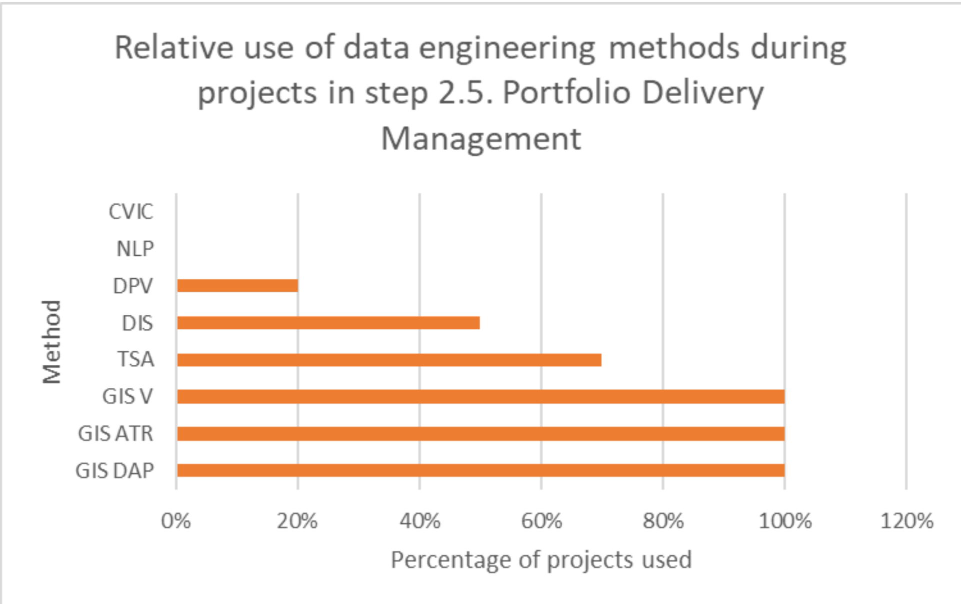


Figure 75 Relative use of the data engineering methods in the process step 2.5. portfolio delivery management

In Figure 75, the relative use for the data engineering methods in the step portfolio delivery management is shown. As can be seen for this step, the three GIS methods are the most used methods.

Information Management

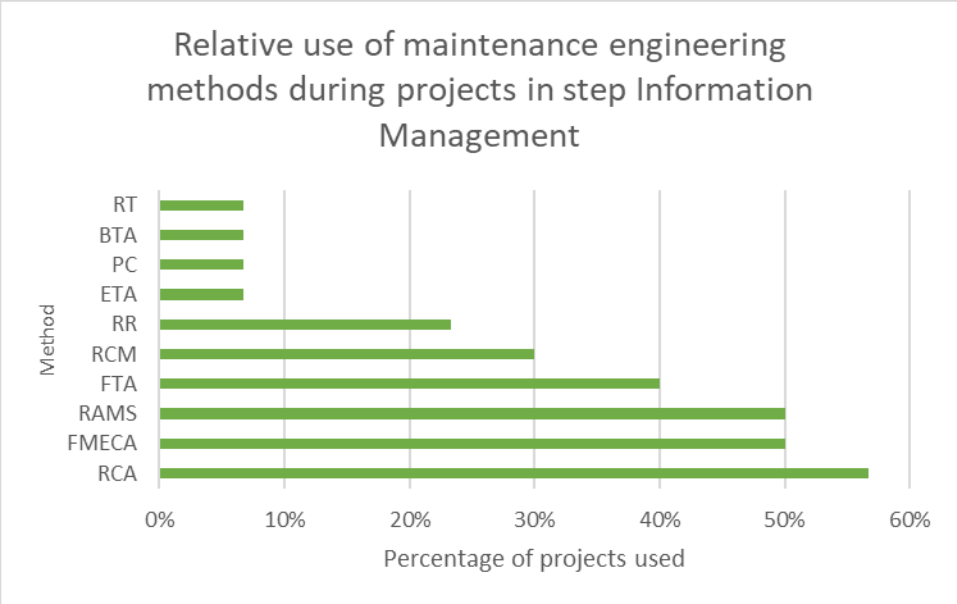


Figure 76 Relative use of the maintenance engineering methods in the process step information management

In Figure 76, the relative use for the maintenance engineering methods in the step information management is shown. As can be seen for this step, Root Cause Analysis is the most used method.

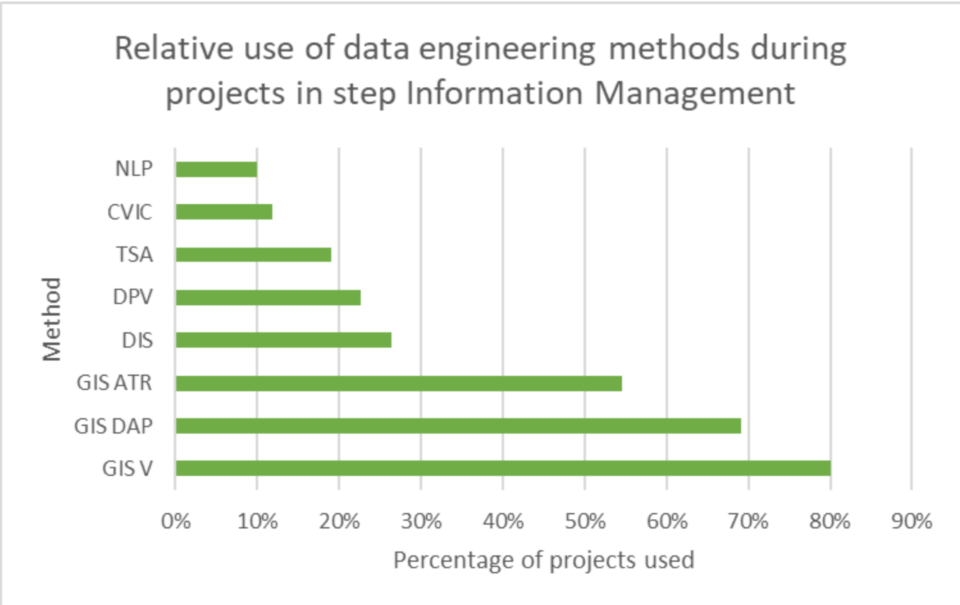


Figure 77 Relative use of the data engineering methods in the process step information management

In Figure 77, the relative use for the data engineering method, in the step information management is shown. As can be seen for this step, the GIS Visualization method is most used.

Performance Management

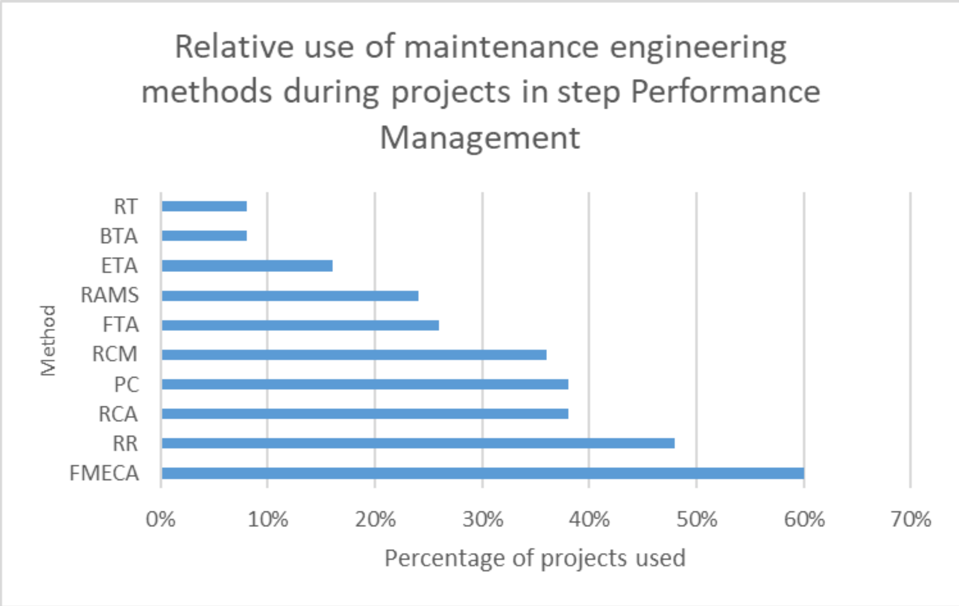


Figure 78 Relative use of the maintenance engineering methods in the process step performance management

In Figure 78, the relative use for the maintenance engineering method in the step performance management is shown. As can be seen for this step, FMECA is the most used method.

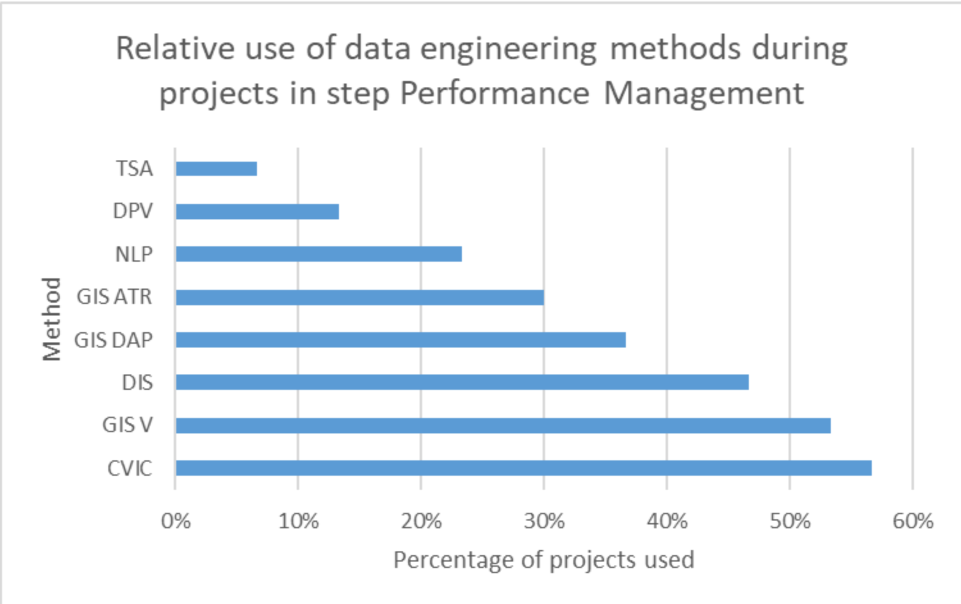


Figure 79 Relative use of the data engineering methods in the process step performance management

In Figure 79, the relative use for the data engineering method in the step performance management is shown. As can be seen for this step, the Computer Vision and Image Classification method is most used.

Appendix D. Arcadis OTL and movable bridge decomposition

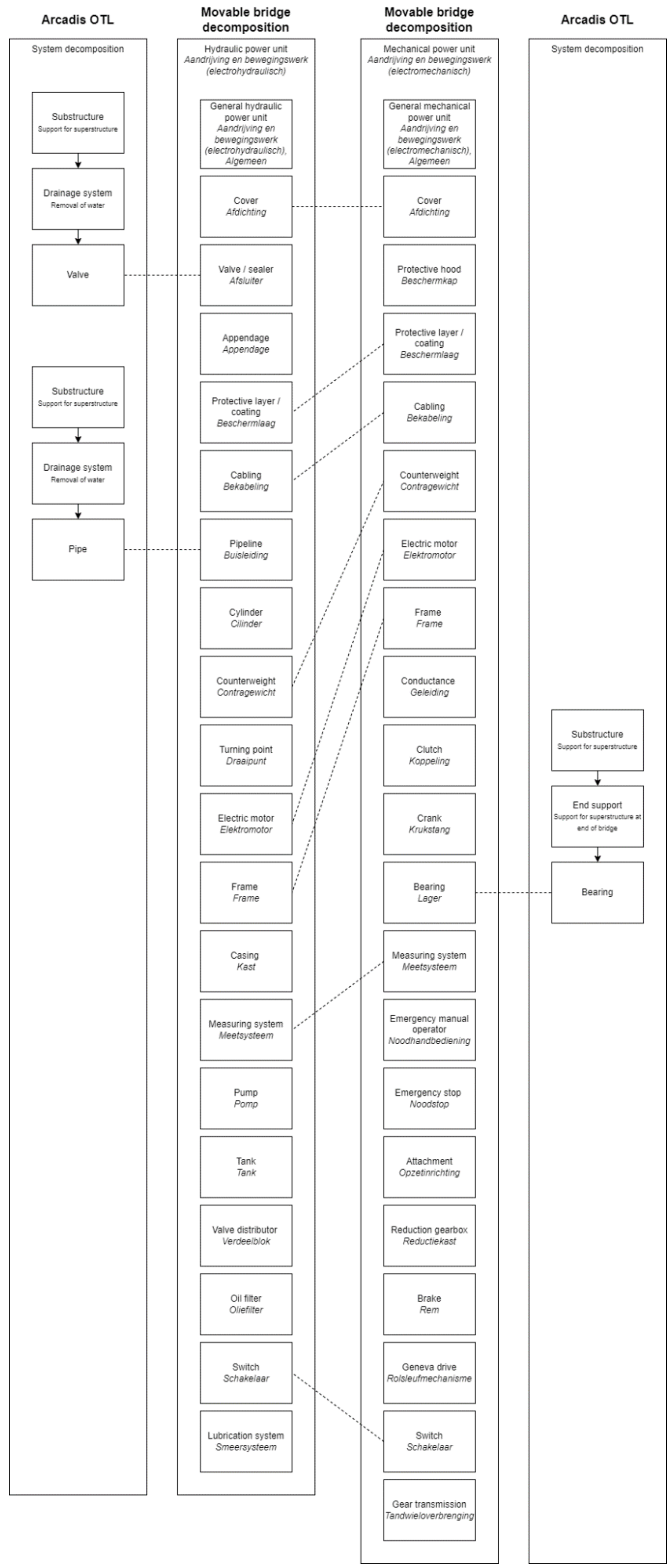


Figure 80 Arcadis OTL and movable bridge comparison containing power units

Appendix E. Attributes for AM from the OTL

In Table 18, the attributes useful for AM are shown. Each attribute has a name, definition, and measurement unit. Some examples of classes for which the attribute is present are given, as are the class levels at which the attribute is present. Level 0 represents the highest level of abstraction: the bridge, level 5 represents the lowest level of abstraction with components like reinforcement couplers and welded joints.

Table 18 Attributes useful for AM from the OTL

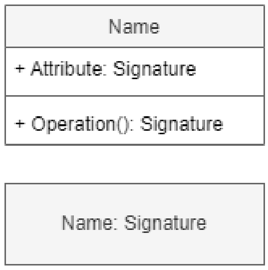
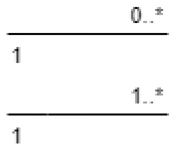
Name	Definition	Measurement unit	Example classes	Class levels	Use in AM
Annual average daily traffic	The total volume passing a point or segment of a highway facility in both directions for one year, divided by the number of days in the year	Time	Bridge	Level 0	Use in performance analysis and usage-based maintenance.
Dead load	Dead loads remain relatively constant over time and comprise, for example, the weight of a building's structural elements, such as beams, walls, roof and flooring components. Dead loads result from the weight of the structure or other fixed elements before any live loads are taken into consideration.	Force	Bridge, Deck	Level 0, Level 2	Use in risk analysis and determining RUL.
Deflection	The displacement of a structural member or system under load.	Distance	Substructure, Superstructure, Deck, Abutment	Level 1 - Level 3	Useful in risk analysis and condition-based maintenance.
Design working life	Assumed period for which a structure or part of it is to be used for its intended purpose with anticipated maintenance but without major repair being necessary.	Time	Bridge, Substructure, Superstructure, Pipe	Level 0 – Level 5	Useful in determining requirements for AM.
Flutter	Self-induced harmonic motion. A self-excited aerodynamic instability that can grow to very large amplitudes of vibrations.	Frequency	Bridge	Level 0	Useful in risk analysis and condition-based maintenance.

Live load	Live loads (also known as applied or imposed loads, or variable actions) may vary over time and often result from the occupancy of a structure. Typical live loads are derived from: people, mobile equipment, vehicular traffic, wind, water, and/or earthquakes.	Force	Bridge, Deck	Level 0, Level 2	Use in risk analysis and condition-based maintenance.
Mass	A measure of the amount of matter in an object, usually measured in kilograms.	Mass	Substructure, Superstructure	Level 1 – Level 5	Use in risk analysis.
Redundancy	A structural condition where there are more elements of support than are necessary for stability.	Dimensionless	Substructure, Superstructure	Level 1	Use in risk analysis.
Redundant member	A member in a bridge that renders it a statically indeterminate structure; the structure would be stable without the redundant member whose primary purpose is to reduce the stresses carried by the determinate structure.	Boolean	Beam, Stiffener	Level 2 – Level 3	Use in risk analysis.
Resonance	The regular vibration of an object as it responds in step (at the same frequency) with an external force.	Frequency	Bridge	Level 0	Use in risk analysis.
Volume	Volume equates to the amount of space that a substance or object occupies	Volume	Substructure, Superstructure	Level 1 – Level 5	Use in risk analysis.

Appendix F. UML explained

In this research, we use class diagrams of UML. Class diagrams exists of three elements; these are explained in Table 19.

Table 19 UML explained

Element	**Name**	**Explanation**	**Example**
Class	Name	Name of the class. If attributes are absent this is followed by the Signature of the class denoted by colons.	
	Attribute	A class can have none to multiple attributes. An attribute can have a signature, denoted with colons.	
	Operation	A class can have none to multiple operations. An operation can use attributes, denoted between the brackets. An operation can have a signature, denoted with colons.	
Relationship	1	This relationship means a class is unique from that side.	
	0..*	This relationship means the class possibly cannot occur but can occur up to infinitely many times.	
	1..*	This relationship means the class must occur at least once and can occur up to infinitely many times.	
	0..n	This relationship means the class possibly cannot occur but can occur up to n times.	
	1..n	This relationship means the class must occur at least once and can occur up to n times.	
Signature	String / URL	A combination of characters, or a URL in the case of URL. Example: ‘Apple’	+ Attribute: Signature + Operation(): Signature
	Integer / Float / Year	A number where integer represents whole numbers, float represents decimal numbers, year represent integers in a year format. Example: 35.67	
	Boolean	A binary value. Example: TRUE	
	Enumeration: Signature	A predefined list where each item in this list follows the structure of another Signature. Example: {‘Apple’, ‘Banana’, ‘Orange’}	
	Series: Signature	Range of multiple variables of another type of signature	

Appendix G. Combined data models

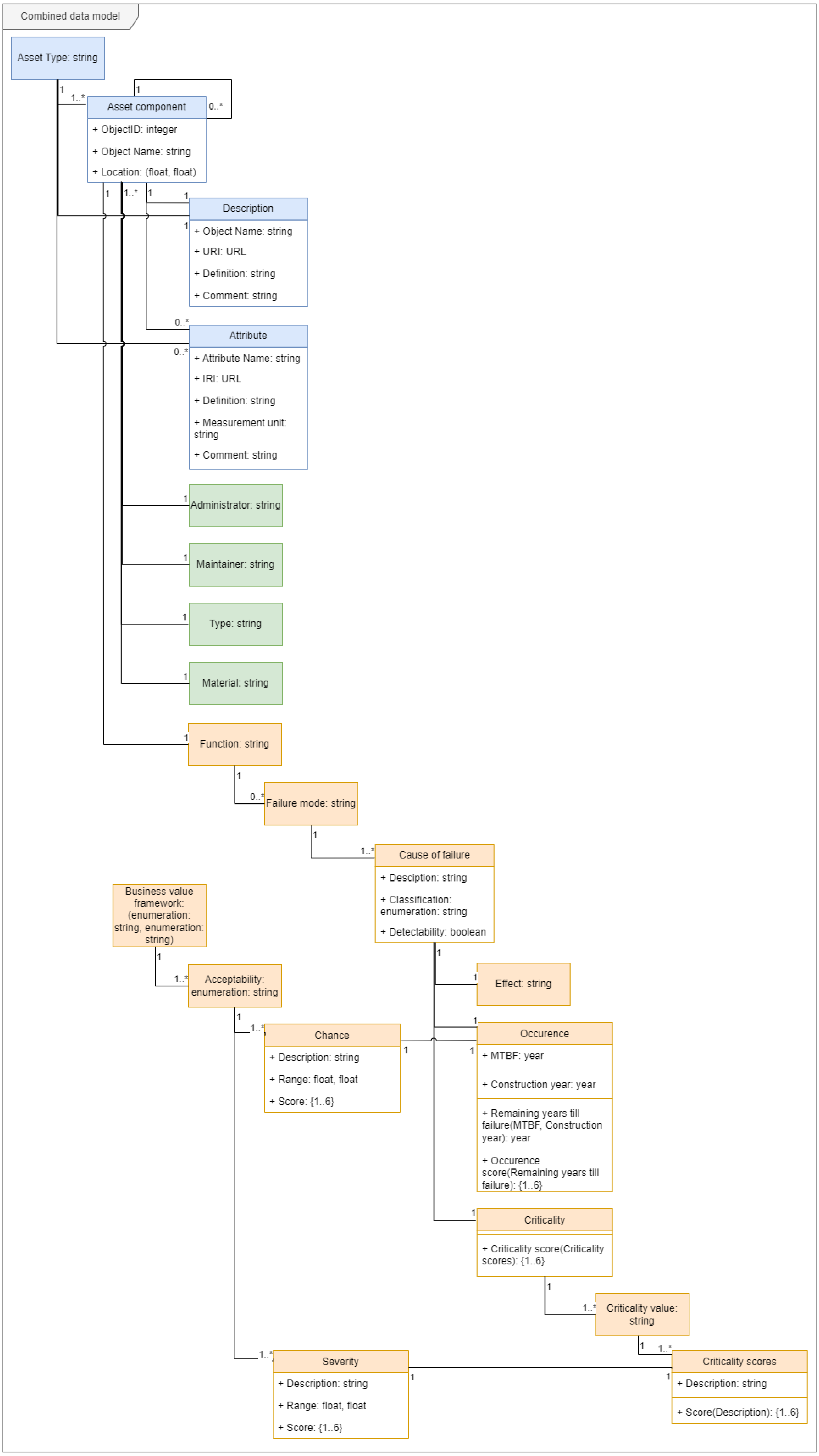


Figure 81 Logical data model of OTL, GIS and FMECA combined

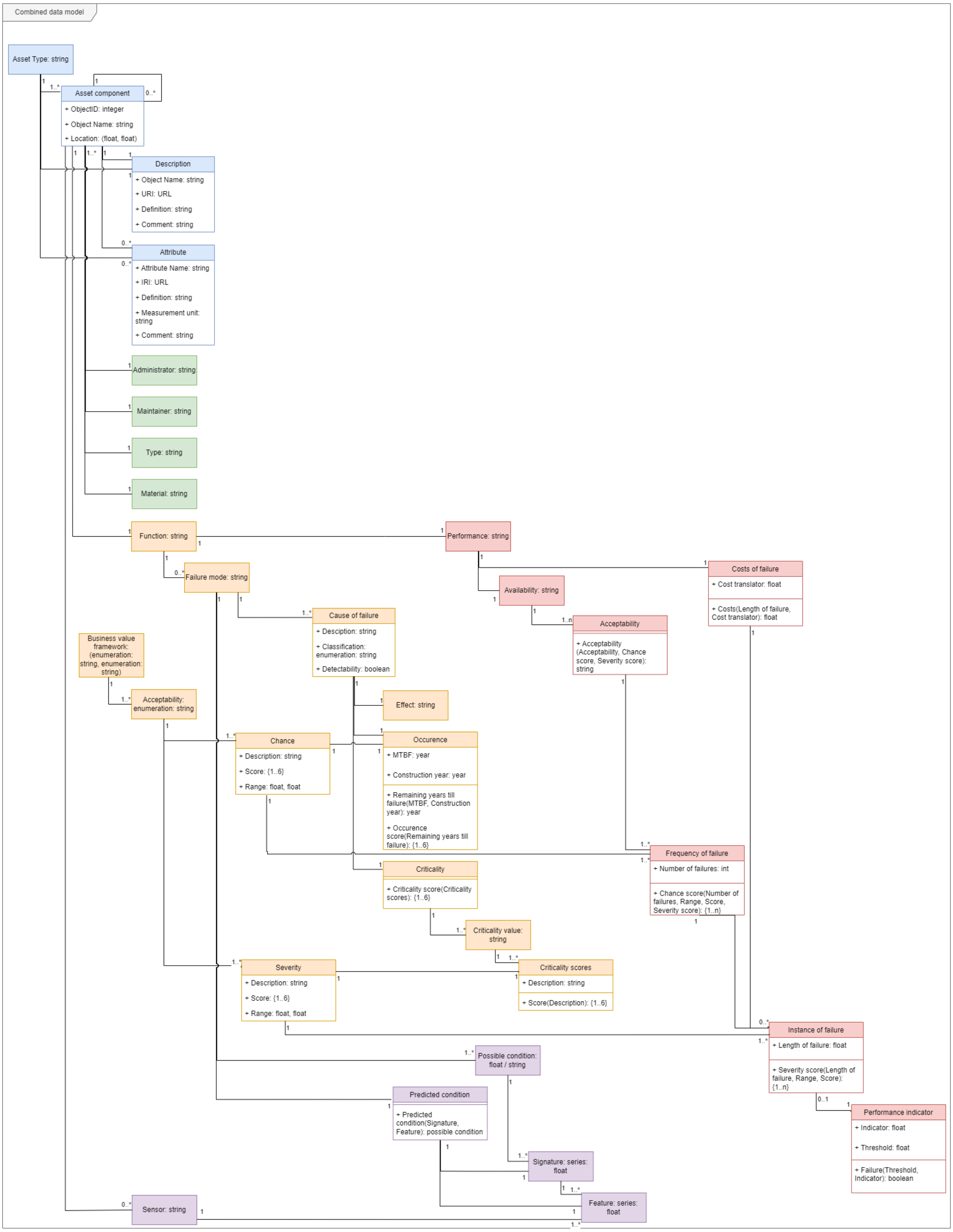


Figure 82 Logical data model of OTL, GIS and FMECA, performance analysis and condition monitoring and prediction combined

Appendix H. Data preparation for performance analysis

In this appendix, data cleaning steps for the opening duration dataset are discussed. The steps taken are shown in Figure 83. The descriptive statistics of the raw data are shown in Table 20. The descriptive statistics of the data after cleaning are shown in Table 21. Cleaning is done in three steps: first, there are duplicates in the dataset. Each passing ship has an entry in the set, even though they could have passed through during the same opening of the bridge, thereby pretending there are more openings than there are. Therefore, these duplicates are removed. Second, the outliers are determined and removed. These are opening durations above 1000 minutes; this value is based on combining high-duration entries with news articles. From these articles, it became apparent that this was planned maintenance or other causes not defined as failures. Afterwards, the 20 entries after a peak are checked to see if these entries also contain a peak; if they do, these entries in between are also removed.

In the dataset are also various entries with low opening durations, even too low to let any boat pass. This can also be seen from the minimum in the tables. However, as the roads are far more important than the traveling of ships, coming from the risk valuation in the risk matrix, opening of the bridge for a short amount of time does not have a big impact.

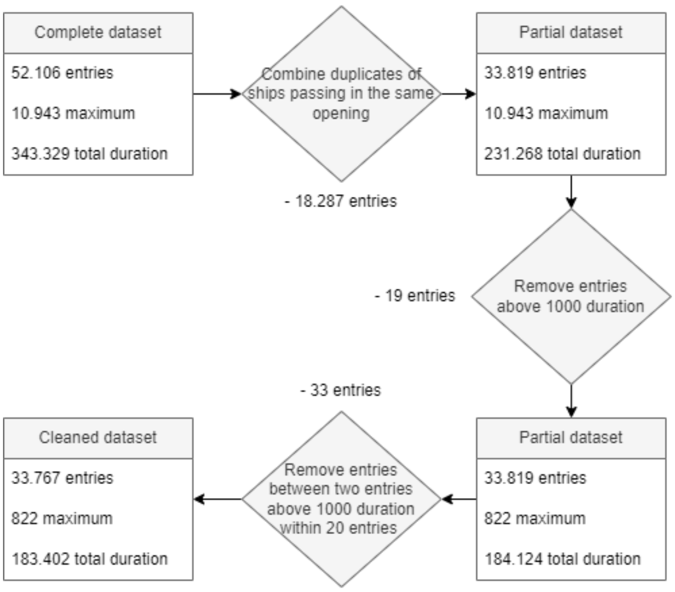


Figure 83 Data cleaning steps

Table 20 Descriptive statistics opening duration before cleaning

Year	Number of passages	Minimum opening duration	Maximum opening duration	Average opening duration	Mode opening duration	Total opening duration
2016	9009	0	10943	8,64	4	77925
2017	8131	0	290	4,62	4	37591
2018	7795	0,55	297	5,35	4	41738,79
2019	8189	1	711	5,45	4	44623
2020	8059	1	19	4,87	4	39238
2021	4156	0,12	7993,95	11,65	4	48412,44
2022	5516	0,12	1199,77	8,04	6,17 6,25	44324
2023	1251	0,12	60,57	7,58	6,42	9476,86

Table 21 Descriptive statistics opening duration after cleaning

Year	Number of passages	Minimum opening duration	Maximum opening duration	Average opening duration	Mode opening duration	Total opening duration
2016	5603	0	822	5,241478	4	29368
2017	5175	0	290	4,40058	4	22773
2018	4646	0,55	297	5,10271	4	23707,19
2019	5242	1	711	4,853682	4	25443
2020	5153	1	19	4,604308	4	23726
2021	2933	0,12	132,5	7,022308	4	20596,43
2022	4003	0,12	610,27	7,60968	6,25	30461,55
2023	1012	0,12	60,57	7,240158	6,42	7327,04

In Figure 84, letter-value plots are sown for each year regarding the opening duration. From the plots 2016 has a wide spread, but relatively low number of outliers. Years with a narrower spread like 2020 do suffer from more outliers, which can be explained by the fact that the threshold for outliers is also narrower.

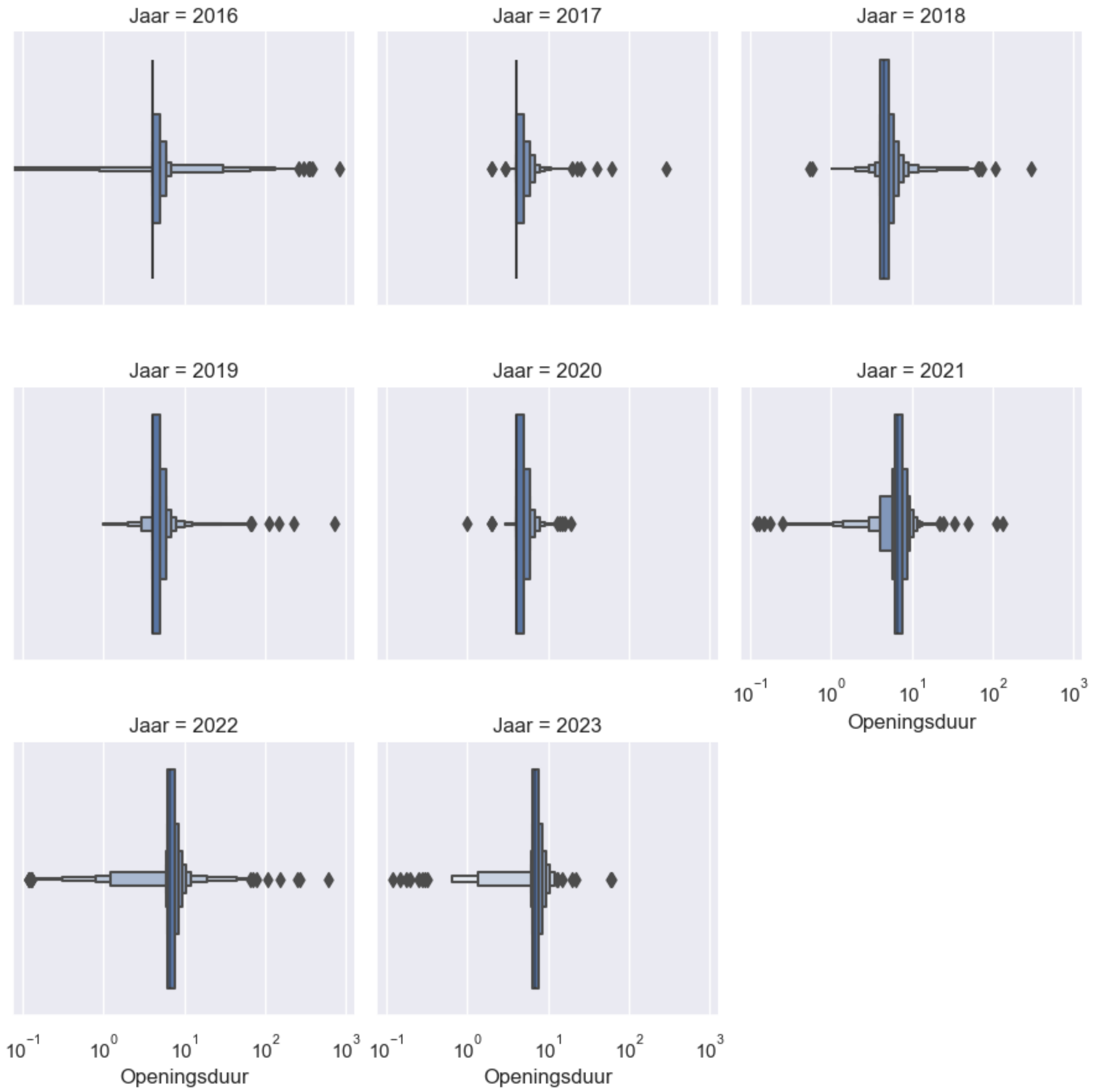


Figure 84 Letter-value plot describing the cleaned data

Appendix I. Relation between sensors and failure modes

In this appendix, the relationship between the sensors and the failure modes of the test rig is shown. For the failure modes, the most extreme conditions are highlighted in red, while normal conditions are marked in blue.

For the cooler failure method, the cooling efficiency and cooling power sensors are dominant. Flow sensor 2 also shows a correlation. All the sensor data over the cycles is visible in Figure 85.

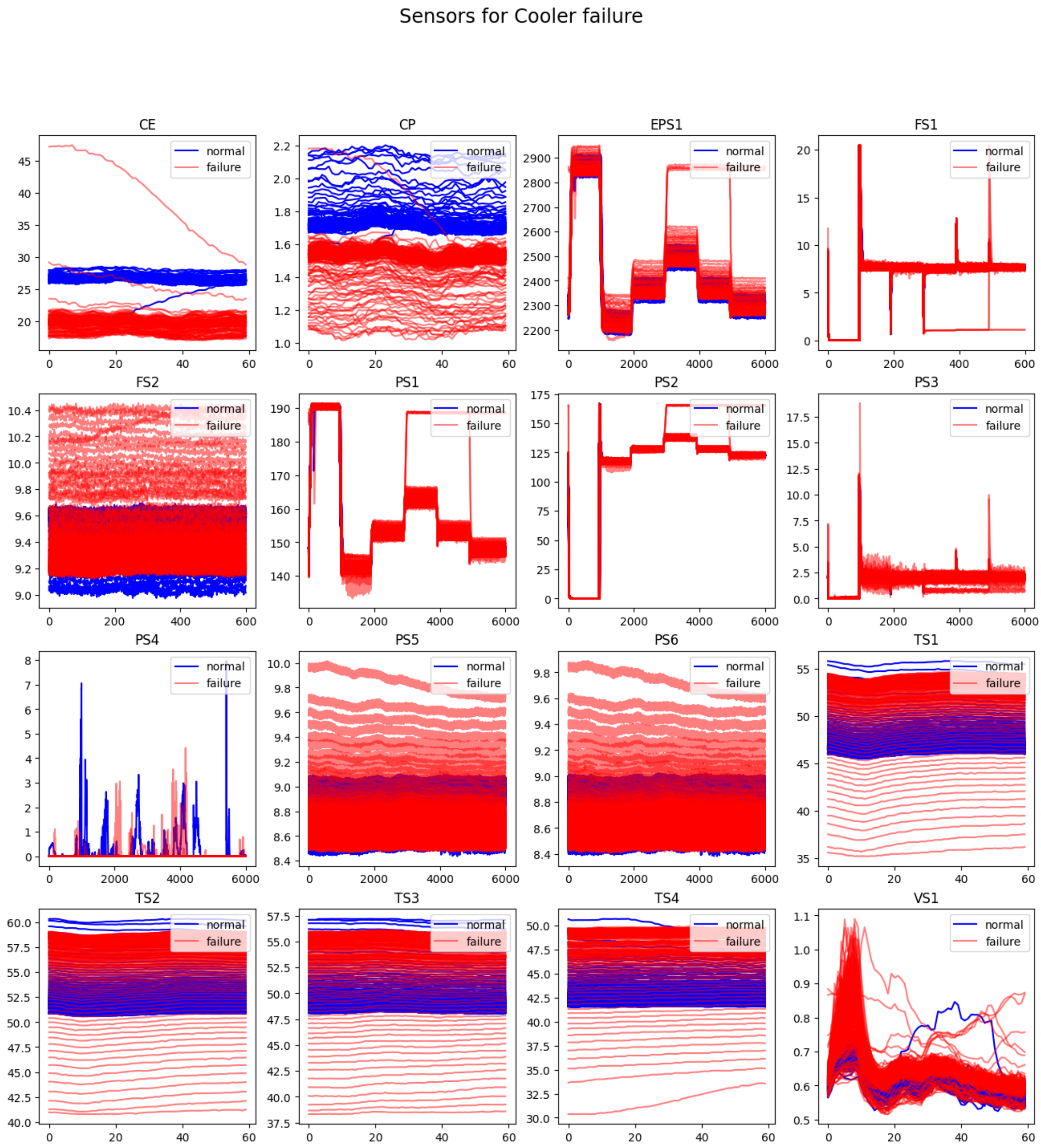


Figure 85 Sensors and cooler failure

For the valve failure method, the most dominant sensors are all the temperature sensors as well as pressure sensors 5 and 6, and flow sensor 2 shows a correlation. All the sensor data over the cycles is visible in Figure 86.

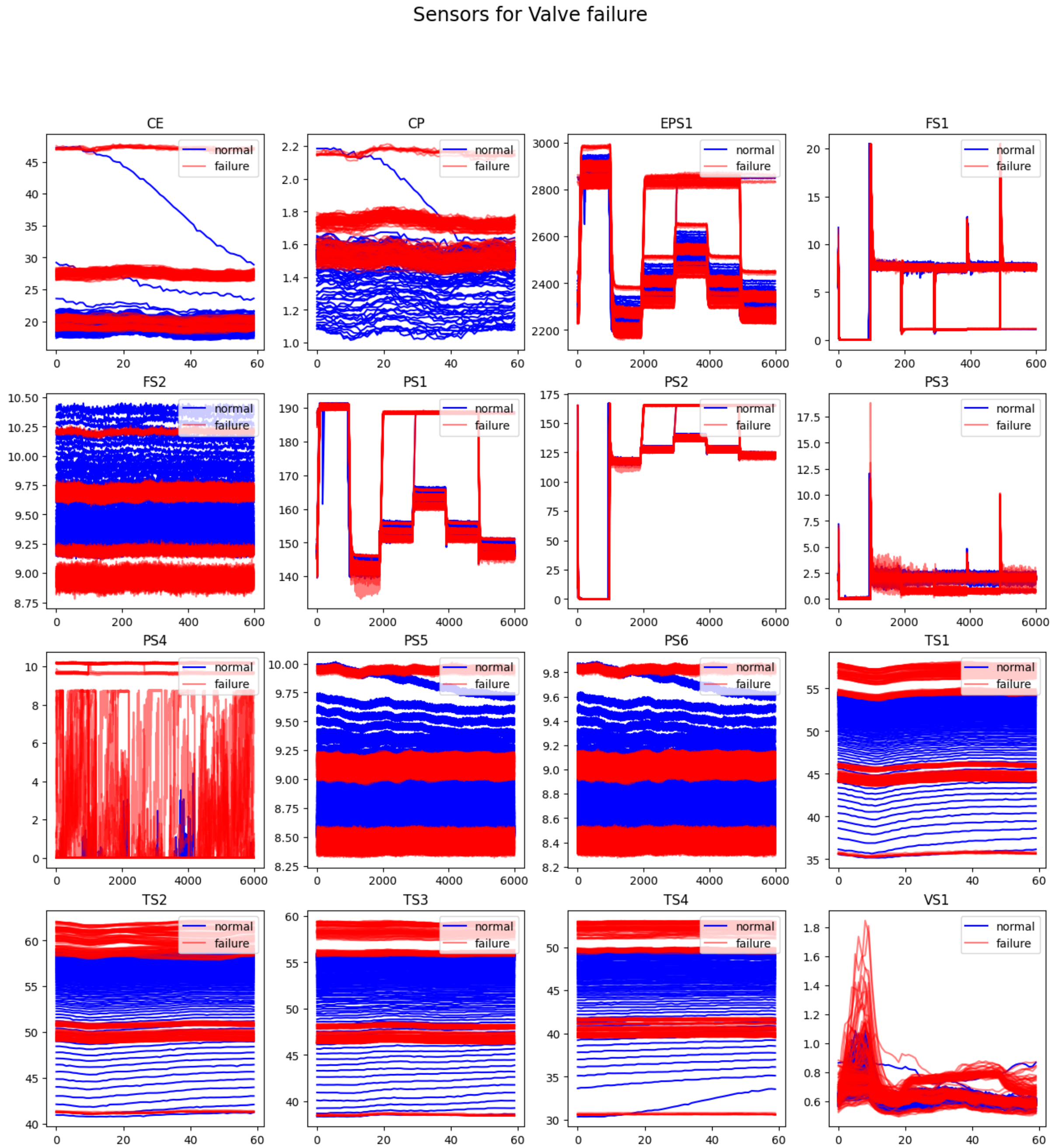


Figure 86 Sensors and valve failure

For the pump leakage failure method, the most dominant sensor is pressure sensor 4 as well as the vibration sensor. Pressure sensors 5 and 6 also show strong correlations, as do all temperature sensors. All the sensor data over the cycles is visible in Figure 87.

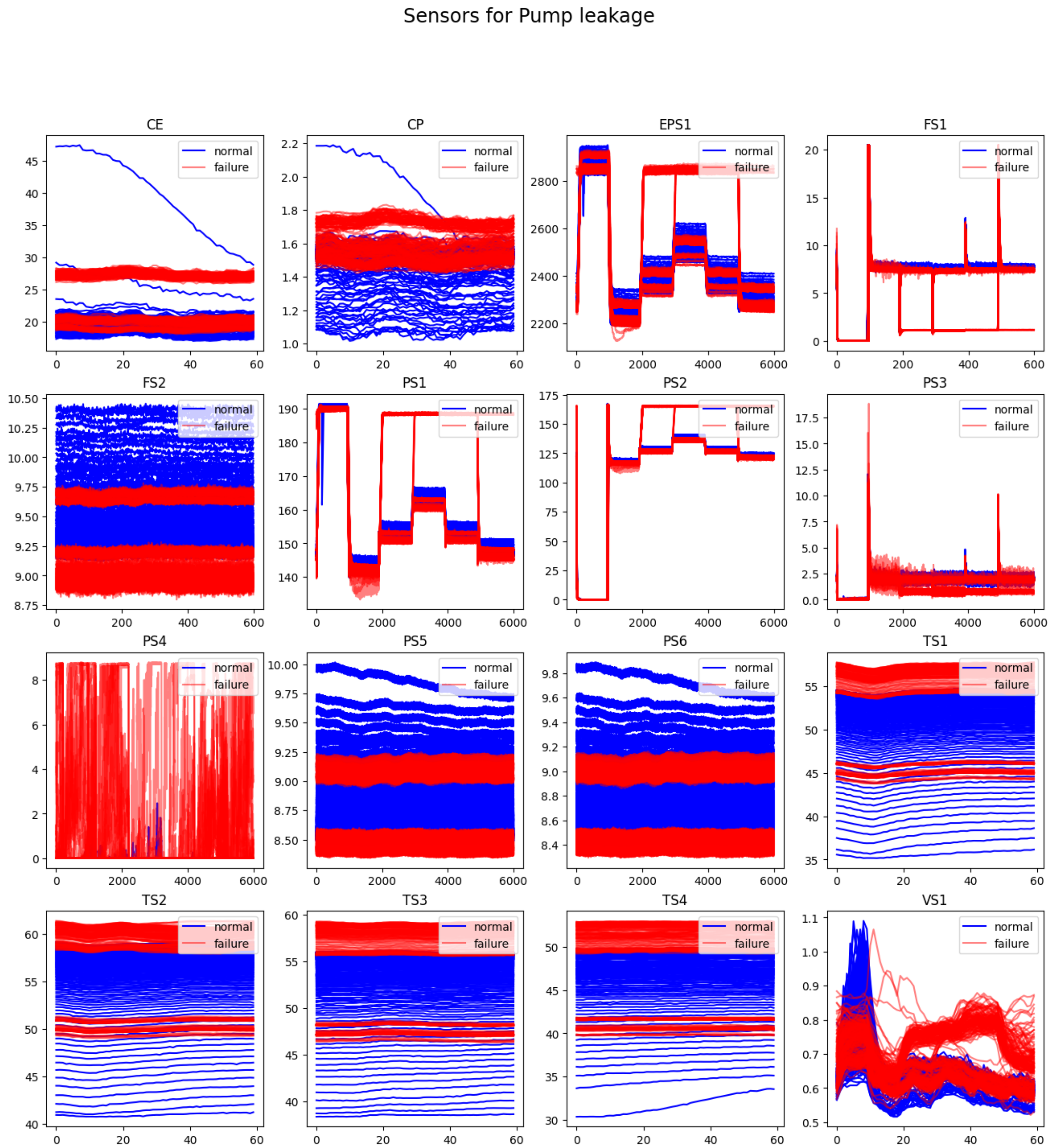


Figure 87 Sensors and pump leakage

The sensor data for the accumulator is visible in Figure 88. From here, it can be seen that pressure sensors 5 and 6 are dominant, as well as the flow sensors. The power sensor and cooling power sensors show a correlation with the failures.

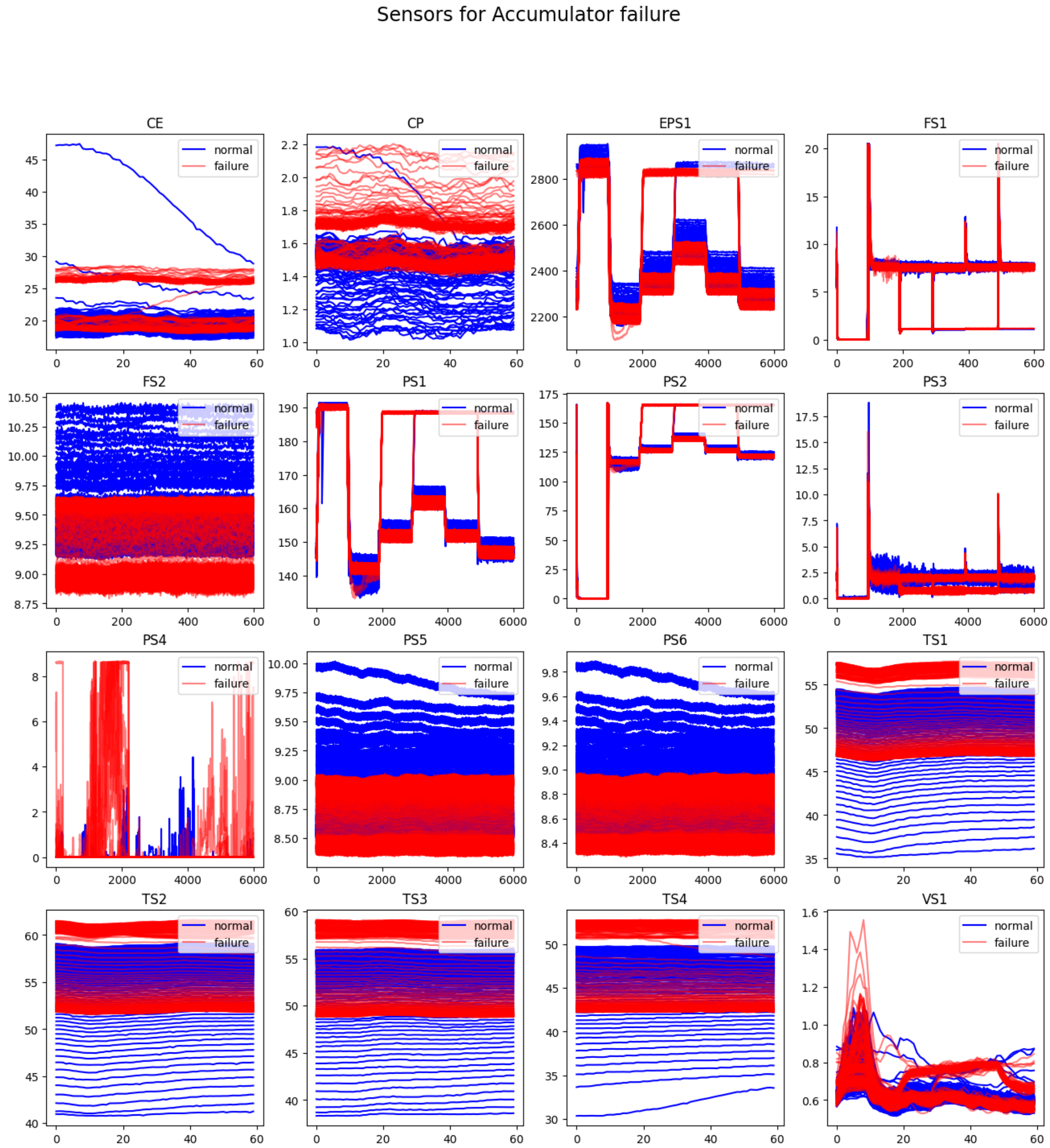


Figure 88 Sensors and accumulator failure

Appendix J. Hyperparameter tuning Neural Network

In this appendix, the hyperparameters for the long short-term memory network are tuned. The parameters in Table 22 are tuned based on the method of sensitivity analysis. To get the best results, 5-fold cross validation is used.

Table 22 Hyperparameters of neural network

Parameter	Possible values	Default value	Best value
Number of nodes in layer	16, 32, 64	32	64
Percentage of dropout	10%, 30%, 50%	30%	50%
Number of epochs	5, 10, 20	10	10 (20)
Optimization algorithm	Stochastic Gradient Descent, ADAM	Stochastic Gradient Descent	Stochastic Gradient Descent
Activation function	Sigmoid, ReLU, Softmax, hyperbolic	Sigmoid	Sigmoid, hyperbolic
Learning rate	0.1, 0.01, 0.001	0.01	0.01
Number of layers	1, 2, 3	2	2

The first is the number of layers. Results are visible in Table 23. As can be seen, the number of nodes per layer has a relatively low impact on all the failure modes. 64 nodes performed best, but 32 is chosen as the default for the sensitivity analysis due to lower computation times and a small difference with 64.

Table 23 Accuracy for the number of nodes in a layer

Accuracy	Cooler failure	Valve failure	Pump leakage	Accumulator failure	Average accuracy
16 nodes	98.68%	83.67%	78.10%	66.17%	81.65%
32 nodes	98.91%	83.67%	77.69%	69.43%	82.43%
64 nodes	98.91%	83.67%	77.87%	69.75%	82.55%

The percentage of dropout is used to prevent the model from overfitting the data to the training data. The results of the percentage on accuracy are visible in Table 24. As can be seen, the dropout percentage also has a low impact on accuracy. 50% dropout is the best, but 30% seems to be performing almost as well and is the default.

Table 24 Accuracy for the dropout percentage

Accuracy (%)	Cooler failure	Valve failure	Pump leakage	Accumulator failure	Average accuracy
10% dropout	99.14%	83.67%	77.78%	66.08%	81.67%
30% dropout	98.68%	83.67%	77.46%	70.39%	82.55%
50% dropout	98.68%	83.67%	77.73%	70.52%	82.65%

The effect of the number of epochs on accuracy is visible in Table 25. The higher the number of epochs, the better the accuracy. However, this means the training takes more time; therefore, 10 epochs are chosen as the default and best value. Higher numbers of epochs are left out due to computation time.

Table 25 Accuracy for the number of epochs

Accuracy (%)	Cooler failure	Valve failure	Pump leakage	Accumulator failure	Average accuracy
5 epochs	97.10%	83.67%	77.69%	64.76%	80.80%
10 epochs	99.00%	83.67%	77.87%	70.70%	82.81%
20 epochs	99.68%	83.67%	79.14%	71.20%	83.42%

Next is the effect of the optimization algorithm on accuracy; this is presented in Table 26. As can be seen Stochastic Gradient Descent outperforms ADAM and is therefore the best algorithm. It is also the default algorithm.

Table 26 Accuracy for the optimization algorithms

Accuracy (%)	Cooler failure	Valve failure	Pump leakage	Accumulato r failure	Average accuracy
Stochastic Gradient Descent	98.87%	83.67%	77.78%	70.02%	82.58%
ADAM	98.73%	83.67%	78.19%	66.08%	81.67%

The next hyperparameter is the activation function. Three functions are tested. The results are shown in Table 27. As can be seen, the activation function has a small impact on the accuracy of the neural network, except for Softmax. Sigmoid and Hyperbolic are the best-performing functions; Sigmoid is also the default function.

Table 27 Accuracy for the activation functions

Accuracy (%)	Cooler failure	Valve failure	Pump leakage	Accumulator failure	Average accuracy
Sigmoid	98.55%	83.67%	77.82%	70.88%	82.73%
ReLu	98.96%	83.67%	78.32%	69.07%	82.51%
Softmax	86.85%	70.75%	55.56%	56.51%	67.41%
Hyperbolic	98.63%	83.67%	78.68%	69.93%	82.73%

Another parameter is the learning rate. The accuracy for three different learning rates is shown in Table 28. Notably, the learning rate of 0.01 outperforms 0.01 on some failure modes; however, 0.01 is the best on average. Therefore, 0.01 is chosen as the best learning rate, it is also the default learning rate.

Table 28 Accuracy for the learning rate of the optimization algorithm

Accuracy (%)	Cooler failure	Valve failure	Pump leakage	Accumulator failure	Average accuracy
0.1	99.55%	83.67%	69.43%	70.29%	80.74%
0.01	98.50%	83.67%	77.69%	69.57%	82.36%
0.001	78.96%	83.67%	77.69%	63.36%	75.92%

The last hyperparameter to choose is the number of layers in the neural network. The results on accuracy are visible in Table 29. As can be seen, the number of layers has a very low impact on accuracy. The best-performing number of layers is 2, which is also the default value.

Table 29 Accuracy for the number of layers in the neural network

Accuracy (%)	**Cooler failure**	**Valve failure**	**Pump leakage**	**Accumulator failure**	**Average accuracy**
1 layer	98.91%	83.67%	77.78%	68.75%	82.28%
2 layers	98.64%	83.67%	77.82%	69.21%	82.34%
3 layers	98.78%	83.67%	77.46%	69.07%	82.24%

The best values for all hyperparameters are shown in the last column of Table 22.

Appendix K. Extended input data analysis of predictive models

In this appendix, the analysis from Section 5.3.3, is extended with three additional scenarios. These scenarios are:

- Only pressure – a scenario where only the pressure sensors are considered.
- Only motor power – a scenario where only the motor power sensor is considered.
- Only temperature – a scenario where only the temperature sensors are considered.

The only pressure scenario is the sixth considered scenario. The accuracy of the predictive models is visible in Table 30. Cooler failure and valve failure are still at the level of prediction of the earlier scenarios. The accumulator failure drops further in accuracy for the random forest.

Table 30 Accuracy of prediction algorithms over the failure modes for the only pressure scenario

Accuracy	**Cooler failure**	**Valve failure**	**Pump leakage**	**Accumulator failure**	**Average accuracy**
Random Forest	92.93%	40.23%	53.70%	33.47%	55.08%
Neural Network	98.32%	83.67%	77.78%	69.07%	82.21%
Average accuracy	95.62%	61.95%	65.74%	51.27%	68.64%

The seventh scenario is the only motor power sensor scenario. This means there is only a single input sensor for the predictions. Accuracy can be found in Table 31. The accuracy drops drastically with cooler failure, while valve failure increases its accuracy.

Table 31 Accuracy of prediction algorithms over the failure modes for the only motor power scenario

Accuracy	**Cooler failure**	**Valve failure**	**Pump leakage**	**Accumulator failure**	**Average accuracy**
Random Forest	33,61%	51,02%	55,37%	36,64%	44,16%
Neural Network	66,80%	83,67%	77,69%	63,36%	72,88%
Average accuracy	50,20%	67,35%	66,53%	50,00%	58,52%

The last scenario considers only the temperature sensors. The performance can be found in Table 32. From here, accuracy drops a bit, but not much compared to the scenarios with two sensor groups or all bridge sensors, showing the importance of the temperature sensors.

Table 32 Accuracy of prediction algorithms over the failure modes for the only temperature scenario

Accuracy	**Cooler failure**	**Valve failure**	**Pump leakage**	**Accumulator failure**	**Average accuracy**
Random Forest	99,55%	43,17%	55,33%	38,37%	59,10%
Neural Network	97,96%	83,67%	77,69%	63,36%	80,67%
Average accuracy	98,75%	63,42%	66,51%	50,86%	69,89%

In Figure 89, the average accuracy for all eight scenarios is visible. As already concluded in Section 5.3.3, the accuracy drops as the amount of input data decreases. The decrease between the two-sensor group and one-sensor group scenarios is relatively small, with the exception of the only motor power scenario. Notable is that the accuracy of Valve failure is relatively constant over all scenarios, as is the prediction for cooler failure, except for the only motor power scenario.

Figure 43 shows a simplified version of Figure 89, only containing the average of all failure modes.

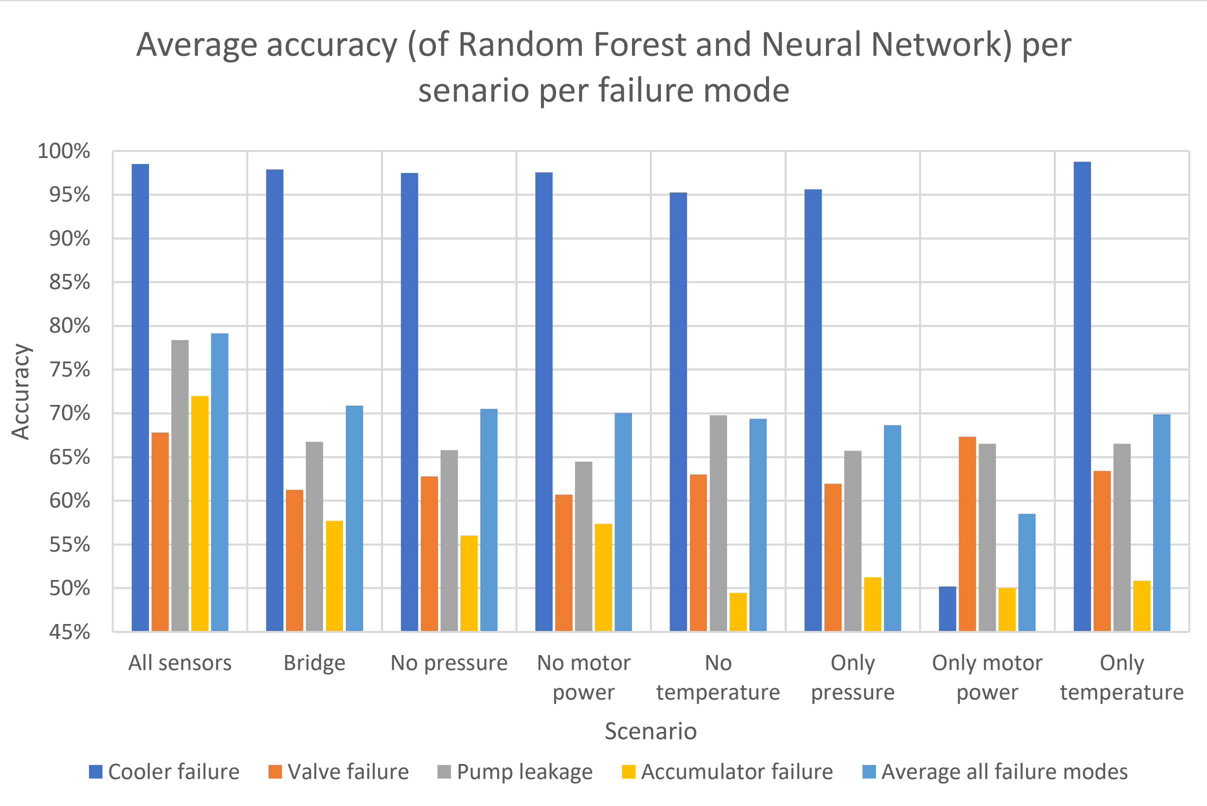


Figure 89 Average accuracy of Neural Network and Random Forest, compared over all scenarios for each failure mode – extended

Appendix L. Monte Carlo simulation

In the figure the steps taken in the Monte Carlo simulation from Section 5.3.4 are shown.

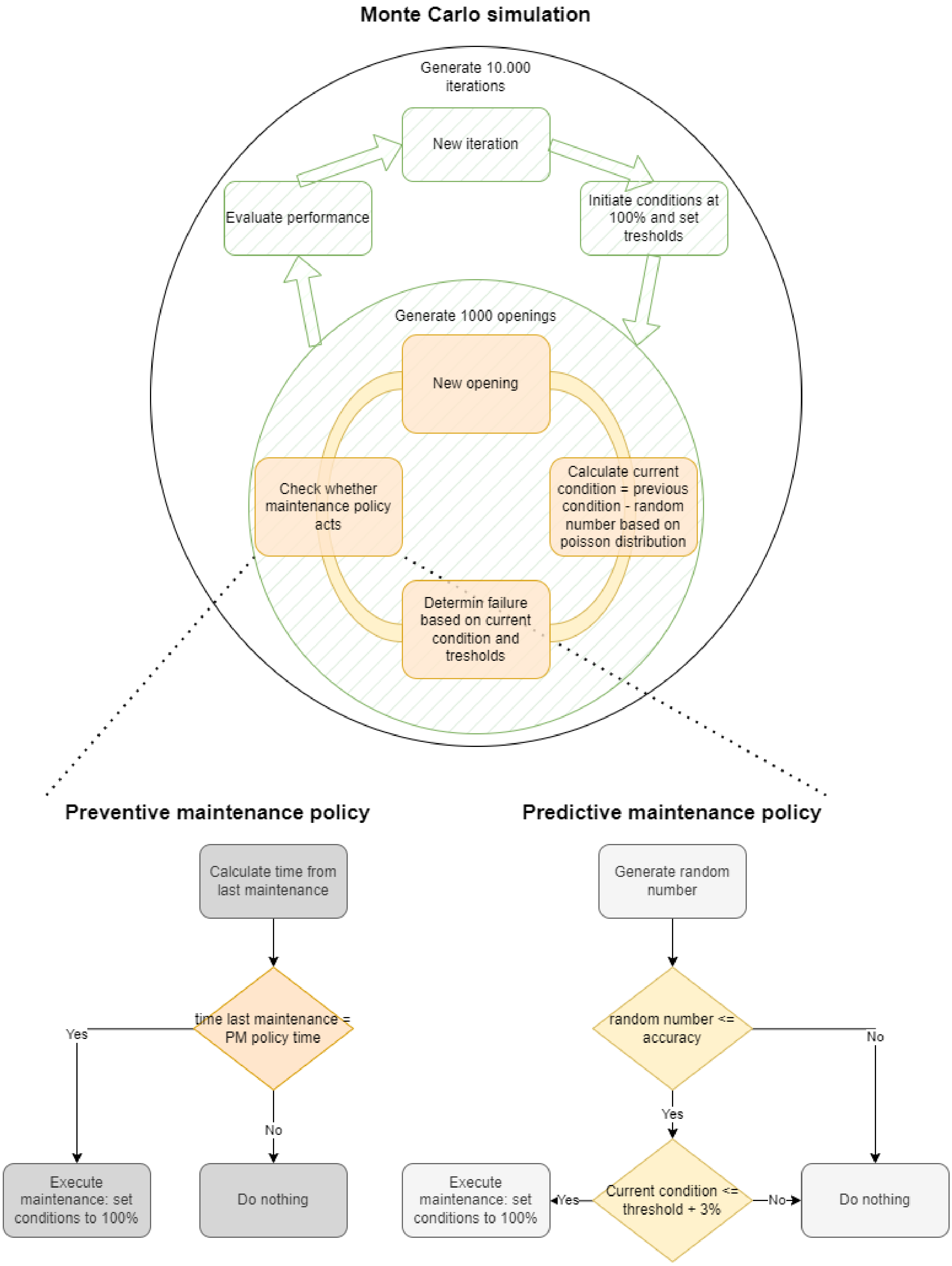


Figure 90 Steps taken for Monte Carlo simulation