Optimising the data acquisition and data delivery of the WenS process at ProRail.

J.B. (Jan) Feddema
Industrial Engineering Management

Examination committee

First supervisor: Dr. A. I. (Adina) Aldea
Second supervisor: Dr. I. (Ipek) Seyran Topan
External supervisor: B. (Bas) van Wijhe

UNIVERSITY OF TWENTE.

ProRail

Preface

Dear reader,

I hereby present you my bachelor thesis for the Industrial Engineering and Management programme. The research was conducted at ProRail. The goal of the research was to optimise the WenS process, a process that helps ProRail to monitor and maintain its railway network in the Netherlands.

Before proceeding further, a few people must be acknowledged for their contribution to this bachelor thesis.

First of all, I would like to thank my external supervisor at ProRail, Bas van Wijhe, for the opportunity he gave me and all the time he invested in me. Although we could not come to the office due to COVID-19, I have learned a lot from my time at ProRail and enjoyed getting to know the company. I also want to thank Juliette van Driel, manager at ProRail, for providing me with this opportunity and sharing her thoughts with me.

Special thanks to both of my supervisors from the University of Twente, dr. Adina Aldea and dr. Ipek Seyran Topan. Thank you, Adina, for all the guidance, advice, and kindness. Without your help, this thesis would not have been the same. Thank you Ipek for helping me to bring this thesis to a good end.

I hope you enjoy reading this thesis.

Jan Feddema

Enschede, The Netherlands, 13 October 2021

Management summary

ProRail, the railway infrastructure manager of the Netherlands uses the WenS process to increase the railway network availability. WenS gives an overview of the usage of tracks and switches within the railway network in the Netherlands. Materieel Impact, a department within ProRail and product owner of WenS, believes that the acquisition and delivery of source data for the WenS process could be performing better. This is taken as a starting point for further elaboration to get to the core problem. From identification of the core problem, the research's main question was established: "*How can the acquisition and delivery of source data for WenS be improved?"*

To answer this main research question, theory on business process modelling was used. The goal was to determine which business process modelling technique would be best suitable for describing the process of acquiring and delivering all the different data for WenS. To do this, the most popular techniques were listed. Then, criteria for choosing a business process modelling technique were defined. Based on these criteria, Business Process Model and Notation (BPMN) was chosen.

To analyse the current situation for acquiring and delivering each of the data sources, semi-structured interviews were conducted with at least one team member of each data source. Capability Maturity Model Integration (CMMI) and theory on performance measurements served as a framework for these interviews. For each data source, the maturity level was determined by the interviewees. The information quality, process quality and user satisfaction (performance measurements) for each data source were also determined. In combination with these measurements, answers from interviewees were coded, both open and axial, to expose the existing challenges within the acquisition and delivery of each data source and set requirements for the proposed solution. The data sources for WenS that are discussed in this research are Switch, Track, TNR, Quo Vadis, TROTS and Configuration data. CGI, the company that combines these data sources, is also part of the research.

The existing challenges were tackled by proposing solutions for each data source. The following recommendations were given for each data source:

Switch

- Use a script to automatically acquire and deliver the data each month.
- Create the source data from scratch. Use the output of WenS from last month or another source system (needs to be determined) to determine the actual switches. Locate the source of the characteristics for switches.

Track

- Complete the characteristics for each track as soon as possible. Locate the source for all the characteristics. This will increase the quality of the output of WenS.

TNR

- Make sure to schedule a timeslot every month to acquire and deliver TNR data. The best time to do this would be as close as possible to the deadline of delivering all data.

Quo Vadis

- Change the delivery from a daily delivery to a weekly/monthly delivery.

TROTS

- Look into the options for removing an unnecessary activity. This would make the chain shorter, and therefore reducing the risks.
- Analyse whether to use Linux or Windows in the process of acquiring and deliver TROTS archives.

CGI, the company that combines all data sources, was suggested to offer data analysis to their customers. General improvements for the WenS process were focusing on communication, coordination and security.

Contents

Reading guide

Chapter 1. Introduction
The first chapter includes a problem description. In the problem description, the norm and reality are used to formulate the research question. Then, a problem-solving approach is formulated to solve the main research question. In the problem-solving approach, the main research question is divided into five sub-questions.

Chapter 2. Theoretical framework
The second chapter provides the theory which is needed to solve the core problem. The chapter answers the knowledge questions formulated in chapter one. Theory on business process modelling, Capability Maturity Model Integration (CMMI) and performance measurements are given. This is done by a systematic literature review.
.

Chapter 3. Current situation
This chapter describes the current situation for each data source. This is done by using Business Process Model and Notation. Also, a small overview of the identified problems per data source is given.

Chapter 4. Interview results
This fourth chapter discusses all the interviews which have been conducted. The information quality, process quality, user satisfaction and maturity level for each data source is discussed. Answers from interviewees are also coded and interpreted.

Chapter 5. Conceptual design
The fifth chapter proposes a solution per data source. This is done by using the interview results from chapter four. Also, a general solution for the complete WenS process is formulated.

Chapter 6. Conclusion, recommendations, and discussion
In the sixth chapter, the answer to the main research question is given together with the recommendations. The evaluation of the research is also provided. Finally, the limitations of the research and recommendations for further research are provided.

Additional information can be found in the appendix.

1. Introduction

ProRail is an independent Dutch government task organisation, which is responsible for the entire 7000 kilometres of railway network and almost 6000 switches in the Netherlands since its foundation in 2005. The organisation's responsibility includes construction, maintenance, safety and management of the entire Dutch railway. Even though ProRail is responsible for the maintenance, the activities are not done by the organisation itself but are subcontracted to maintenance contractors. However, ProRail does assign the available capacity of the railway network and is also responsible for the railway traffic control.

ProRail wants to be a data-driven organisation where decisions are made based on data. ProRail is responsible for monitoring the number of times a switch or track should be maintained to prevent disruption in the Dutch railway network. A maintenance schedule can be determined by analysing the data about the number of times a switch or track has been driven.

1.1 Current situation

1.1.1 Organization structure

The department Asset Management (AM) is responsible for the prevention of disruptions within the Dutch railway network. This is done by gathering and analysing information about the status of the railway infrastructure. The information indicates whether maintenance or replacement should occur.

The department AM – Informatie consists of three sub-departments: Vernieuwing (Renewals) Projecten & Implementatie (Projects and Implementation) and Operatie (Operation). Within the subdepartment Operatie, there is a distinction between:

1. Datagestuurd Asset management (Data-driven Asset Management): responsible for analysing data and advising on the use of data.

2. Configuratiedata (Configuration Data): responsible for the management of all objects that form the railway network.

3. Sturingsdata (Monitoring): responsible for the condition of the railway and the corresponding objects.

We can find a visualization of the organization in Appendix C. **Materieel Impact (Material Impact)**, the problem owner of this research, falls within the Sturingsdata department. The cluster Materieel Impact works together with carriers and contractors to help increase the availability of the railway network in the Netherlands. Higher availability is needed due to the increasing number of daily passengers. Data is delivered and analyses are done in cooperation with the carriers and contractors for more efficient and effective track maintenance. Materieel Impact is also responsible for measuring the weight of trains, checking the quality of the tracks and the material driving on the tracks. This is done to optimize the maintenance process, prevent disruptions and determine the user fee.

1.1.2 WenS process

To increase railway network availability, Materieel Impact provides “switch use” and “track load” data to stakeholders monthly, in Excel spreadsheets. Appendix A and B provide more information on these spreadsheets. These excel sheets are created by the IS called WenS. Data is collected, stored and processed to provide these excel sheets. The excel sheets give an overview of the actual load and usage of the railway network. To create the excel sheets, the following five sources are combined:

TROTS: *Train Observation Tracking System.*

This system determines the position and identity of trains and passes it on to the train control posts of ProRail. The department Rail Traffic Control needs the position of trains for controlling rail traffic.

Quo Vadis: *Measuring driving trains*

Quo Vadis is a system for measuring the weight, axes and dynamic force on driving trains. The system measures the deflection of the rail with optical sensors. Given this deflection, the static and dynamic force that a wheel of the train exerts on the rail is deduced. The 45 Quo Vadis systems are strategically placed across the Dutch railway network to measure as many trains as possible. Quo Vadis is an important data source for the dataset SwitchUse and RailLoad as the weight of the trains influences the load class of the tracks and switches.

SAP: *Systems, Applications and Products.*

SAP contains static information about the railway network in the Netherlands. The information contains several characteristics about the switches and tracks. Information about railway tracks includes length, local speed (both passenger and freight), construction date and leading-in switches. Information about switches includes location, type, angle, speed and construction date. SAP is described as switches and tracks in figure 1.

MCS: *Measuring points Configuration Service*

Measuring points Configuration Service delivers system configuration of measuring points that determine when and where train activities on the railway infrastructure should be measured.

TNR: *Train Number Range*

Through this application, all train numbers are documented. This source data is needed to connect a train number to a train type (passenger train or freight train) and a carrier.

Figure one visualises the data sources within the whole WenS process. Each of the departments delivers their source data monthly to Conseillers en Gestion et Informatique (CGI), a service provider managing the software and therefore the processing of WenS. CGI combines all the data sources into excel spreadsheets (Appendix A). These excel spreadsheets are sent to Materieel Impact for validation. If the spreadsheets are correct, Materieel Impact delivers the spreadsheets to their customers. If not, Materieel Impact analyses and fixes the problem in consultation with CGI. According to Materieel Impact, this happens rarely.

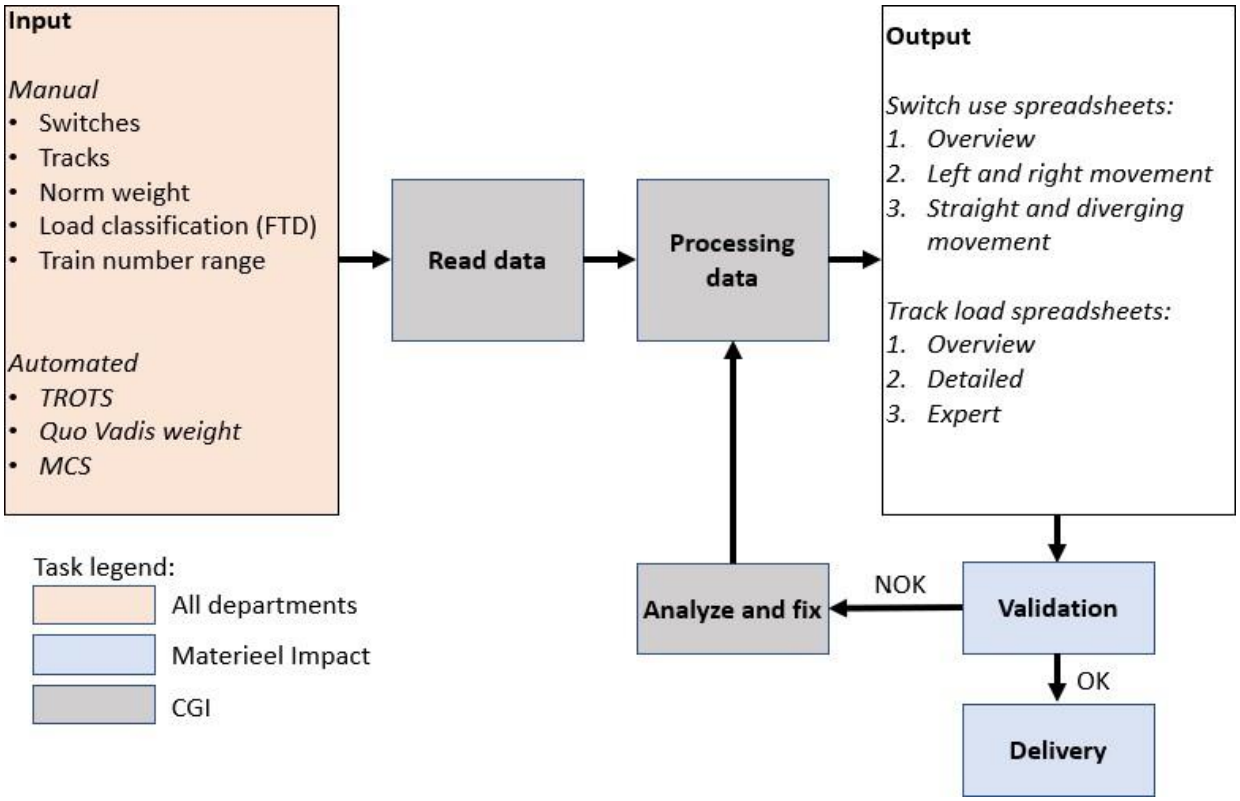


Figure 1: The WenS process

1.2 Problem description

1.2.1 Problem cluster

Materieel Impact believes that the acquisition and delivery of source data the WenS process could be performing better. Therefore, Materieel Impact perceives the need for more insight into the supply of source data for WenS. This can be substantiated by anecdotal evidence. Multiple information specialists and a process manager within Materieel Impact have indicated the need for more insight into the data sourcing for WenS. To determine the main causes for this perceived need, we analyse and visualise the WenS process (Figure 1) and use a problem cluster to identify a core problem (Figure 2).

The problem cluster has been designed by the method provided in the book Solving Managerial Problems Systematically by Heerkens & van Winden (2017). According to the book, a problem cluster is a visualization of several problems and their mutual relationships. A problem cluster serves as a means of structuring the problem context and is used to identify the core problem. The arrow respectively represents cause → effect. Problems that do not have a cause are possible core problems. From the problem cluster, it becomes clear that several possible core problems are causing an inefficient process. One of these possible problems is the core problem. The core problem should be able to be solved in 10 weeks by using IEM theory learned in the last years.

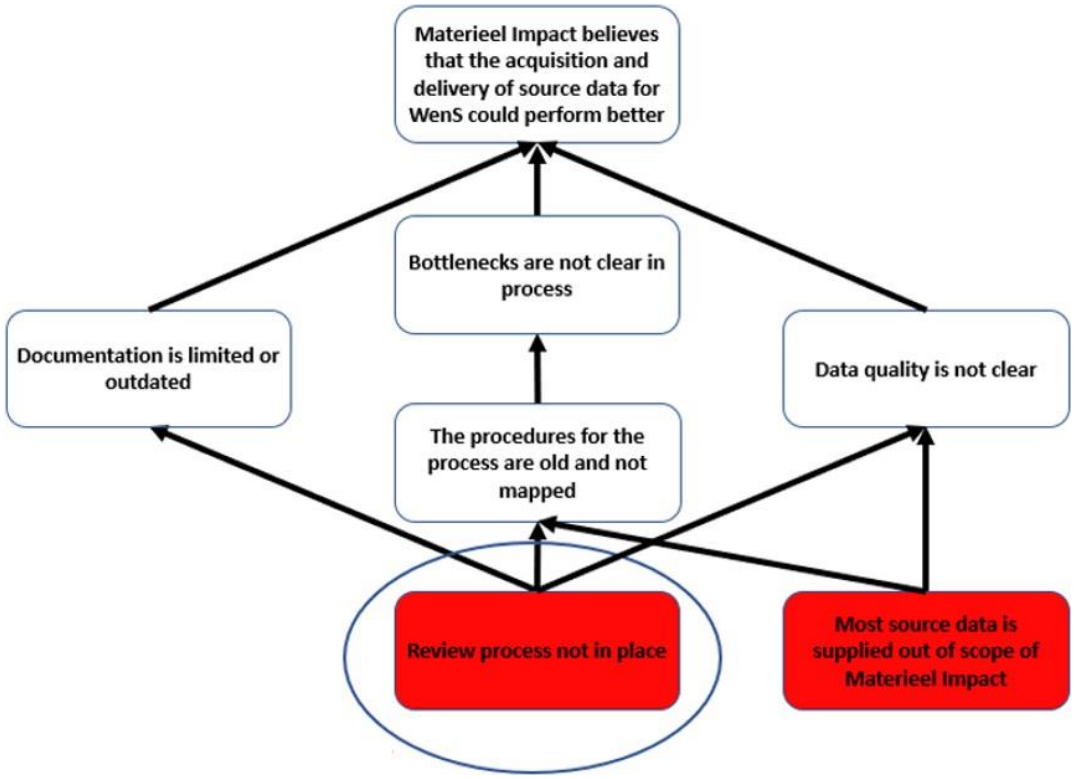


Figure 2: Problem cluster

The current supply of data has several challenges. Materieel Impact believes that the acquisition and delivery of source data could perform better. Interviews have made clear that three causes make Materieel Impact believe the process could perform better. Documentation about the process is limited or outdated and both bottlenecks and data quality in the process are not clear. In the end, two core problems are identified. Most source data is supplied out of the scope of Materieel Impact and a review process has not been conducted for a long time. In other words: Due to these core problems, Materieel Impact is unable to determine the performance of the acquisition and delivery of source data for WenS. As Materieel Impact is the problem owner of this research, it requires knowledge about the processes of data acquisition and delivery for the Wens process to be able to improve it.

The red problems are potential core problems. One of them is the problem: “Review process not in place”. This problem lies at the core of all other problems. Solving it should therefore lead to improving the performance of WenS. For instance, CC BI delivers its source data one to three times a day. Materieel Impact believes that is not the most efficient way. Therefore, Materieel Impact needs to know how the data sourcing for WenS is functioning and performing. There are also some inefficiencies that Materieel Impact perceive in delivering their data source. Materieel Impact uses the WenS spreadsheets from the previous month to determine newly built switches. The switch name and corresponding information are manually added to the switch data source. This is done by the information specialists within Materieel Impact. As the new switches are added manually to the switch data source, this process is error-prone and time-consuming and could therefore be one of the bottlenecks.

The second and last core problem, “most source data is supplied out of scope of Materieel Impact”, has not been chosen. A global cost-benefit analysis tells that there are great costs for changing this situation. Data is managed within ProRail by different departments. To change this, the complete structure within ProRail should be changed.

After analysing the problem cluster, we conclude that the focus of this research is on reviewing and improving the acquisition and delivery of data sources for WenS. We choose this core problem as we keep in mind the conditions of this research; it should be solvable in ten weeks and consist of IEM theory.

1.2.2 Action problem

The action problem is a discrepancy between the norm and the reality (Heerkens & van Winden, 2017). The action problem can be derived from the core problem. In more detail, the action problem can be formulated as follows: “Materieel Impact has not conducted a review process on the acquisition and delivery of data for the WenS process for a long time; a review process should be conducted to be able to design a To-Be business process model including possible improvements for the acquisition and delivery of data for WenS.

1.2.2.1 Reality

The reality represents the current situation. The reality has been determined by interviews with team members of Materieel Impact. Materieel Impact has indicated that they do not know how the acquisition and delivery of data sources for WenS is performing. There are indicators for a bad performing data acquisition and delivery, but underlying causes are not clear.

1.2.2.2 Norm

The norm represents the desired situation. Team members of Materieel Impact have indicated a need for more insight into the acquisition and delivery of data sources. According to these team members, the norm is to know how the complete data acquisition and delivery for WenS is performing and where bottlenecks are located.

1.2.3 Research question

To solve the action problem, a research question should be constructed. This research project revolves around solving the research question. Therefore, the following research question is composed: *'How can the acquisition and delivery of source data for WenS be improved?"*

To answer the main research question, sub-questions are needed. More elaboration on the sub-questions can be found in the problem approach. In the problem approach, each sub-question is assigned to a phase in the chosen method.

1.2.3.1 Research goal

The research will be dedicated to solving this main question. To be more specific, the research will contain an analysis of the current processes for acquiring and delivering the source data. The analysis will possibly lead to areas of interest for improvement. After the analysis, advice on the areas with the most improvement potential will be provided to Materieel Impact. The advice contains a To-Be business process model which depicts how the data acquisition and delivery could be improved. Materieel Impact must decide whether they want to implement the recommendations. Implementation is not included in this research.

1.2.4 Definition of key constructs

WenS process = An information system that combines several data sources for a monthly delivery of information to stakeholders about the degree of use and load of switches and tracks managed by ProRail.

Data sourcing / Data collection = Data collection (also data sourcing) is the systematic approach to gathering and measuring information from a variety of sources to get a complete and accurate picture of an area of interest. Data collection enables a person or organization to answer relevant questions, evaluate outcomes and make predictions about future probabilities and trends (McLaughlin, 2020).

Efficiency = Efficiency signifies a peak level of performance that uses the least number of inputs to achieve the highest amount of output. Efficiency requires reducing the number of unnecessary resources used to produce a given output including personal time and energy. It is a measurable concept that can be determined using the ratio of useful output to total input. It minimizes the waste of resources such as physical materials, energy, and time while accomplishing the desired output (Banton, 2020).

Input-Process-Output (IPO) model = A general approach in system analysis and software engineering to describing the structure of an information processing program.

System quality = System quality is a measure of the information processing itself that includes software and data components, and a measure of the technical soundness of the system (Benmoussa et al., 2018). DeLone and McLean (2003) measured the quality of the system in terms of ease of use, functionality, reliability, flexibility, data quality, portability, integration and importance.

Information quality = Information quality (IQ) is measured in terms of accuracy, timelines, completeness, relevance, and consistency (Benmoussa et al., 2018).

User satisfaction: User satisfaction is the sum of feelings and attitudes regarding distinct factors, which affect user's satisfaction positively or negatively. (Baily and Pearson, 1983).

1.3 Problem Solving Approach

1.3.1 MPSM method for the first phase

A well-known research strategy at the University of Twente is "Solving Managerial Problems Systematically" (MPSM) by Heerkens & van Winden (2017). The first phase of this method is an effective starting method for identifying the right problem. The method consists of the following phases:

1. Defining the problem
2. Formulating the approach
3. Analysing the problem
4. Formulating (alternative) solutions
5. Choosing a solution
6. Implementing the solution
7. Evaluating the solution

The first phase of this method, defining the problem, is an effective starting method for identifying the right problem. However, for further research, other methodologies might be more suitable to tackle the given problem. The MPSM method is not specifically designed for research that involves the transformation of data and analysis of this data. For the acquisition and delivery of data in the WenS process (IS), it would be helpful to approach this research from a data transformation and – analysis perspective.

1.3.2 Design Science Research Methodology

The Design Science Research Methodology (Peffers, K., et al, 2007) is a methodology for conducting design science (DS) in information systems (IS). The objective of this method is to design an artefact to

solve the identified problem. The paper defines artefacts as: "Artifacts are potentially constructs, models, methods, or instantiations or new properties of technical, social, and/or informational resources. Conceptually, a design research artefact can be any designed object in which a research contribution is embedded in the design" (Peffers, K., et al, 2007). A significant difference between the Design Science Research Methodology (DSRM) and MPSM is the artefact. To solve the identified core problem in this research, the DSRM is the best suitable fit. Generally, the MPSM method focuses on solving an action problem (bringing the problem from reality to norm) while the DSRM method focuses on the creation of an artefact.

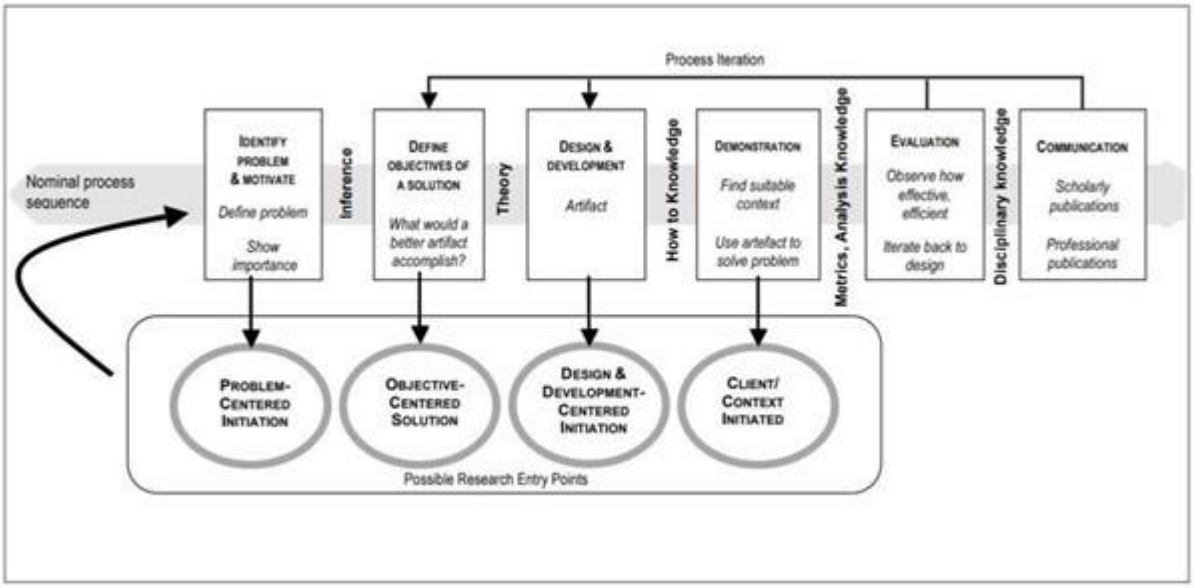


Figure 3: Design Science Research Methodology

The DS process includes six steps: problem identification and motivation, definition of the objectives for a solution, design and development, demonstration, evaluation, and communication. A schematic overview of these steps is depicted in figure 1. For this research, the six steps are elaborated more below:

Phase 1: Identify problem and motivate

Phase 1 has already been executed in section 1.1 using the MPSM method. Continuing from here, DSRM will be used to execute the research.

Phase 2: Objectives of a solution

The objective is to develop a business process model which can be used as a research mechanism to be able to improve the acquisition and delivery of data sources for WenS. According to "The Design Science Research Methodology (Peffers, K., et al, 2007)", prior literature is an important fundament for developing a process model. The development of a process model should build upon the strengths of prior research. During this phase, research should be done on how to solve the identified problem before being able to conclude the objectives of a solution. As stated before, a review process on the acquisition and delivery of source data has not been conducted for a long time. This reduces the ability to improve the acquisition and delivery of data for WenS. Creating a model is the logical step in solving the problem. A

brief scoping research has been conducted. This research showed that business process modelling helps to find a solution for the identified problem. Therefore, the theoretical perspective is business process modelling. However, there are several methods for modelling information system processes. Finding the right method is an important step for a good problem-solving approach. This leads to the following research question:

1. What kind of business process modelling technique is suitable for the improvement of the acquisition and delivery of source data?

This question has been answered by a systematic literature review. Core concepts to answer this question are business process modelling and information systems. Business Process Modelling Notation will be used within this research. The systematic literature review can be found in chapter 6. After determining that this thesis includes business process modelling, more information should be gathered on how to analyse the business process models. Maturity models are a good fit to analyse the current situation because a maturity model shows how capable the business processes are of achieving continuous improvement. This leads to the following research question:

2. Which type of maturity model can be used for business process modelling?

This question will be answered by a general review of the literature. The chosen maturity model will be used to determine at which level the process is performing. The level of maturity is chosen by analysing the answers given in the interviews.

Phase 3: Design and development

This phase describes the artefact that is created. The delivered artefact will be a business process model for all the source data. A business process model is the graphical representation of a company's business process. For this research, it is the acquisition and delivery of data for WenS. Different graphing methods can be used for the modelling. This phase consists of two different stages: design and development. To be able to design the model, more knowledge on the current situation should be gathered.

3. What does the As-Is model of acquiring and delivering the source data look like?

This question will be answered by organising a meeting with one stakeholder of each of the five data sources. These conversations/interviews will be the main input for designing the model. After these questions have been answered, the model of the current situation should help to find possible improvements in the process. Knowledge question one gave an answer to which business process modelling technique should be used. The model can be designed by BPMB. A program that suits this notation well is Bizagi. From now on, this model will be referred to as the As-Is model.

Phase 4: Demonstration

Using the differently designed business process models and the knowledge gathered about improving information systems, a new, optimised model can be created. The designed model from the past phase, also called the As-Is model, can be used to create a To-Be model. A To-Be model depicts how the future process can be improved. To be able to create the To-Be model, the following questions should be answered:

4. What alterations are needed to improve the acquisition and delivery of the chosen source data for WenS?

The purpose of question three is to determine which sources of data should be improved. This will be determined by assessing the maturity of the business process models of the current situation. If an As-Is model does not represent any main issues, further research on this As-Is model can be ignored. A review will be conducted on which maturity model to use for this thesis.

The fourth question will be answered by previous knowledge gathered in the interviews and the second round of interviews. This will result in an overview for improvements in the acquisition and delivery of the chosen source data.

Phase 5: Evaluation
Observe and measure how well the To-Be model improves the acquisition and delivery of source data for WenS. This can be done by a survey questioning the employees who are involved in the process. Check whether the new process model is a significant improvement for the data acquisition and delivery of data sources for WenS. This leads to the following knowledge question:

5. How to evaluate the To-Be business process model?

The results of the evaluation can be used for further improvement of the process. A small literature study will elaborate more on how to evaluate the To-Be business process model.

Phase 6: Communication
Communicate about the research and the resulting advice. Communicate about its importance, utility and novelty with the stakeholders. The stakeholders consist of ProRail and the University of Twente. After the research is finished, the research will be orally defended for the exam committee and published on the internet. The results will also be presented to all people interested from ProRail.

1.3.3 Deliverables

Based on the problem definition, methodologies and all other knowledge, the following will be delivered in this research:

- Map of the current situation (BPMN) for each of the data sources
- Needs/requirements of team members for improved data acquisition and delivery
- Advice on how the data acquisition and delivery for the WenS process can be improved using a To-Be model
- Evaluation on the opinions of the team members for the To-Be model

1.4 Research design

Previously, the research- and knowledge questions have been determined. The next step is to determine an approach to solve these questions. Determining the approach will be done within this chapter. The questions have been arranged in a table. The table represents an overview of the method that is used to solve each of the questions. This includes the type of research, the research population, the research strategy, the method of data gathering and a method of data processing. All these subjects will be elaborated within this chapter later.

Knowledge problem	Type of research	Research population	Research strategy	Method of data gathering	Method of data processing	Method of analyzing
What kind of business process modelling technique is suitable for the improvement of the acquisition and delivery of source data?	Descriptive	Literature	Broad-qualitative	Literature study which results in systematic literature review	Qualitative	Content analysis
Which type of maturity model can be used for business process modelling?	Descriptive	Literature	Deep qualitative	Literature study w	Theoretical framework for assessing the As-Is models	Systematic literature review
What does the As-Is model of acquiring and delivering the source data look like?	Descriptive	One team member of each of the departments	Deep qualitative	Semi-structured interviews	Visual representation using a business process model	Content analysis
What alterations are needed to improve the acquisition and delivery of the chose source data for WenS?	Descriptive	The different departments at ProRail responsible for the source data	Deep qualitative	Semi-structured interviews	Qualitative Quantitative	Content analysis Quantitative analysis
How to evaluate the To-Be business process model?	Descriptive	Team members working with the process and literature study	Qualitative	Semi-structured interview	Qualitative Quantitative	Content analysis Quantitative analysis

1.4.1 Type of research

First, the different available types of research should be determined. The types of research can be classified as reporting, descriptive, explanatory or predictive. These research types can be distinguished by the following characteristics (Cooper and Schindler, 2014):

- Reporting: A reporting study provides a summation of data, often recasting data to achieve a deeper understanding or to generate statistics for comparison.
- Descriptive: A descriptive study tries to discover answers to the questions who, what, when where and how.
- Explanatory: An explanatory study goes beyond description and attempts to explain the reasons for the phenomenon that the descriptive study only observed.
- Predictive: A predictive study attempts to predict when and in what situation an event will occur.

To determine how and why certain events happen, business research should be done. To be able to understand whether the research type is quantitative or qualitative within this research, a distinction between the two should be made. Qualitative research can be described as "an array of interpretive techniques which seek to describe, decode, translate, and otherwise come to terms with the meaning, not the frequency, of certain more or less naturally occurring phenomena in the social world" (Van Maanen, 1979). This differs from quantitative research, as quantitative methodologies answer questions related to how much, how often, how many, when, and who. This methodology answers a precise measurement. (Cooper and Schindler, 2014) After these definitions have been set, we can conclude that this research will mostly consist of qualitative research. The mapping of the current process does not require to have further precise measurements for quantification and answers the question who, what when, where and how. Therefore, this is a descriptive study.

1.4.2 Research population

The research population for this research consists of all the team members involved in the data acquisition and delivery. To map the current situation, one team member of every department will be interviewed. To determine the bottlenecks and possible solutions, more team members per department will be questioned. The different departments and their responsibilities have been discussed in chapter 1.3. For the first part of this thesis, problem identification, all the departments have been questioned to understand the reasons for a perceived need of Materieel Impact for a more efficient WenS process.

1.4.3 Explaining choice of data gathering method

This research will use two different data gathering methods. The first data gathering method is semi-structured interviews. According to Cooper and Schindler (2014), a semi-structured interview generally starts with a few specific questions before following the respondent's tangents of thought with interviewer probes. Another data gathering method is questionnaires. This is defined by Gehlbach (2014) as the design and development of self-administered surveys. Questionnaires allow researchers to collect large amounts of information within a relatively small amount of time. Below, the reasoning for using or not using these methods have been explained for each phase.

Problem identification
Within this phase, semi-structured interviews have been used to gather information. Documentation was read and a few questions were prepared to appear competent during the interview. The preparation helped to keep the interviews going and knowing what to focus on. These semi-structured interviews contributed to creating a problem cluster. The reason behind using interviews for this phase is that at the start of this phase, there is no knowledge yet. The interviews are a great start to get to know the company, the employees and the problem. A first personal encounter with the employees will benefit for better research in the future when the employees are needed for more interviews or as respondents for questionnaires.

Design and development
The phase design and development creates an artefact; the As-Is model of acquiring and delivering source data for WenS. The business process model is most easily created by semi-structured interviews. Questionnaires will not meet the requirements for the As-Is model. The questions can not be efficiently designed to extract the information needed as follow-up questions should be asked depending on the answers of the interviewees. Using semi-structured interviews, follow-up questions can be asked immediately.

Demonstration
To create the To-Be model, which is the goal of the demonstration phase, semi-structured interviews will be used. The interviews will give an answer to which alterations are needed for the As-Is model. Combining this information contributes to getting to a proposed solution.

Evaluation
Evaluation of the newly designed business process model, the To-Be model, will be executed by a semi-structured interview. The evaluation phase's goal is to determine whether the goals of the research have been met and the newly designed model is useful. Using a semi-structured interview leaves space for the interviewee to share his or her thoughts.

1.4.4 Explaining choice of data analysis method

Content analysis (CA) is a method to draw conclusions using a systematic process from the content of messages. Krippendorff (2004) explains content analysis as "a research technique for making replicable and valid inferences from texts (or other meaningful matter) to the contexts of their use". Content analysis suits this research best, as the data output from interviews will be qualitative. By using content analysis, meaning can be interpreted from the responses given by the team members. This can be used to draw conclusions.

To analyse the semi-structured interviews, the meetings will be recorded. Recording meetings will prevent that information is forgotten. Before recording, written approval for recording is gathered.

2. Theoretical framework

This chapter discusses the theoretical framework used throughout the research. The goal of this chapter is to gather all the needed theoretical knowledge for successful research. The knowledge which is needed has been described in Chapter 1.3.2 by the following two research questions: “What kind of business process modelling technique is suitable for the improvement of the acquisition and delivery of source data?” and “Which type of maturity model can be used for business process modelling?”. This chapter will answer both questions.

2.1 Business process modelling for WenS

The problem identification showed that a review process on the acquisition and delivery of source data has not been conducted for a long time. This reduces the ability to improve the acquisition and delivery of data for WenS. Creating a model would be the logical step in solving the problem. A brief scoping research has been conducted. This research showed that business process modelling helps to find a solution for the identified problem. Therefore, the theoretical perspective is business process modelling. However, there are several methods for modelling information system processes. Finding the right method is an important step for a good problem-solving approach. Therefore, this section discusses the relevant theory for this research found on business process modelling techniques.

2.1.1 Business process modelling techniques

This section gives an overview of popular business process modelling techniques. Business process models specify the activities, with their relationships, that are performed within a single organisation and therefore specify process orchestrations. Process orchestrations provide a detailed view of the activities of processes and their execution constraints (Weske, 2012). Aldin & De Cesare (2009) give an overview of several popular modelling languages that can be used for visualizing process orchestrations. For each of the techniques, an example is included for better understanding.

Flow charts

A flow chart is a graphical representation that shows the flow of control throughout a process by providing a step-by-step illustration of what occurs given a specific situation. Flow charts are used predominantly in software engineering, but their simplicity and ease of use have enabled managers and business owners to adopt this technique for organisational purposes as well. (Aldin & De Cesare, 2009).

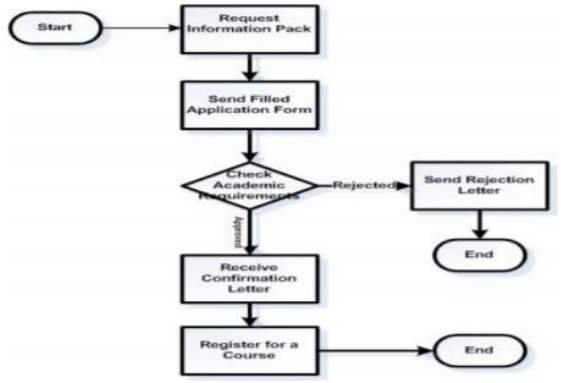


Figure 4: Example of a flow chart (Aldin & De Cesare, 2009)

Petri nets

Petri nets are a well-known technique for specifying business processes formally and abstractly. Petri nets consist of places, transitions and directed arcs connecting places and transition (Weske 2012). According to Aldin & De Cesare (2009), a Petri net is a mathematical/graphical representation that is appropriate for modelling systems with concurrency. Petri nets are used for modelling computer software, hardware, control and business processes.

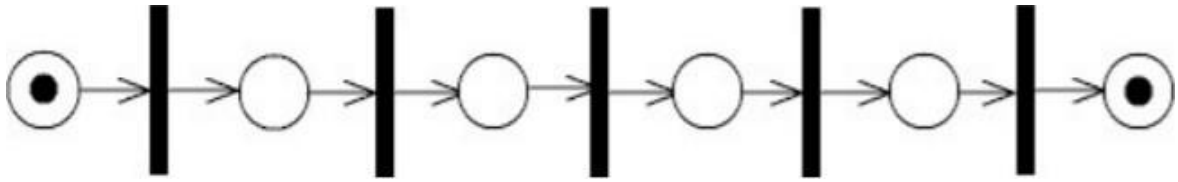


Figure 5: Example of a Petri net (Aldin & De Cesare, 2009)

Data flow diagram (DFD)

A data flow diagram is a graphical representation used to show system functionality. This includes underlying processes and the flow of data. It is mainly used for studying systems analysis and design in software engineering. (Aldin & De Cesare, 2009)

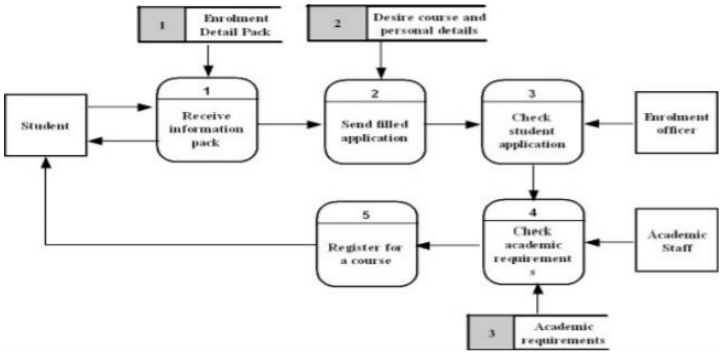


Figure 6: Example of a data flow diagram (Aldin & De Cesare, 2009)

Role activity diagram (RAD)

RADs are a graphical representation of processes in terms of the roles presented within these processes, their component activities and their interactions, together with external events and the logic, which determines what is the sequence of those activities (Aldin & De Cesare, 2009)

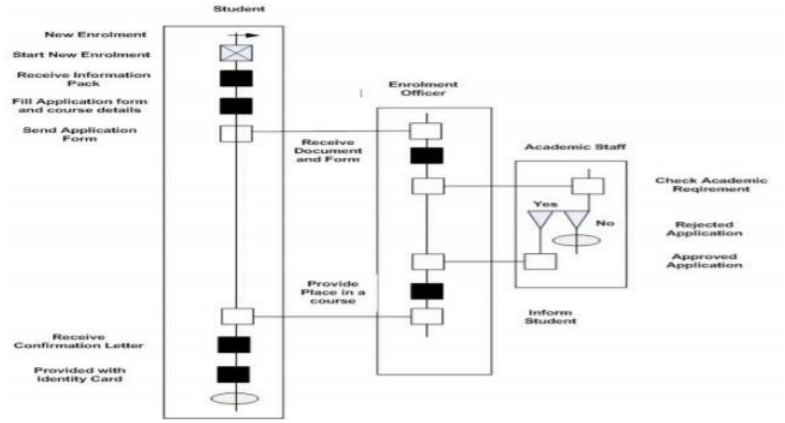


Figure 7: Example of a role activity diagram (Aldin & De Cesare, 2009)

Business Process Modelling Notation (BPMN)

The Business Process Modelling Notation aims at supporting the complete range of abstraction levels, from a business level to a technical implementation level. The primary goal of this notation is a high understanding for all users. To conclude, BPMN creates a standardized bridge for the gap between the business process design and process implementation (Weske, 2012).

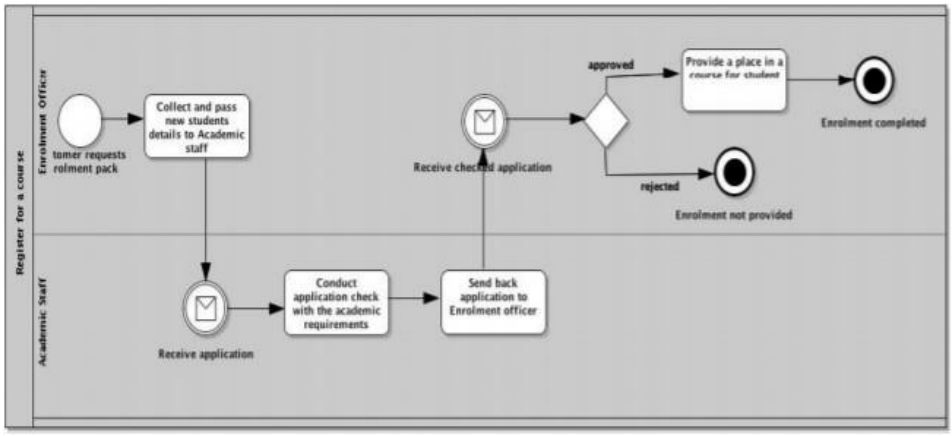


Figure 8: Example of business process modelling notation (Aldin & De Cesare, 2009)

2.1.2 Choosing a business process modelling technique

This section discusses how a suitable business process modelling technique has been chosen. As the most popular business process models have been described, a decision must be made on which technique will be used within this research. To do this, the criteria for choosing a business process model must be defined.

Aldin & De Cesare (2009) conducted a comparative analysis of popular business process techniques based on five criteria. These five criteria are scope, flexibility, ease of use, understandability and simulation. Aguilar-Savén (2004) classifies it into two dimensions. These dimensions are the purpose of the model and the model change permissiveness. Luo & Tung (1999) include four characteristics for comparing different methods. These characteristics are formality, scalability, enactability and ease of use. Some of the criteria are named differently while meaning the same. These criteria are put in the same row in Table 1: Criteria for choosing a business process modelling techniqueTable 1. Based on these three papers, a list of five different criteria for choosing a business process modelling technique could be obtained.

Table 1: Criteria for choosing a business process modelling technique

Criterium	Aldin & De Cesare (2009)	Aguilar-Savén (2004)	Luo & Tung (1999)
1.	Scope	Purpose of the model	Scalability
2.	Flexibility	Model change permissiveness	Formality
3.	Ease of use		Ease of use
4.	Understandability		
5.	Simulation		Enactability

There is no need for the business process model to be capable of dynamically simulating a business process as the purpose of the business process model is to serve as a visualisation of the current situation. There is also not a need for stakeholders to be able to apply the technique. The stakeholders are only interested in the current situation. Therefore, the criteria ease of use and simulation are rejected. As a result, the criteria used to choose a model are scope, flexibility and understandability.

The first criterium is scope. The scope can be defined by the extent to which the process modelling elements are represented by constructs of the technique. The elements include process, activity, service and product, role, goal, event and rule. BPMN provides more details than a flowchart when it comes to defining processes. Therefore, a flowchart is excluded too. Data Flow Diagrams are excluded because this research does not specifically focus on data flow. Flexibility is another important criterion. It should be possible to change the model without completely replacing it. After reading documentation and conducting semi-structured interviews, the created model must be validated by a team member. It is possible that the created business process model is not correct and that changes are necessary. Therefore, the business process model should be easily changeable. The last criterium is understandability. Understandability is important because the business stakeholders need to understand the business process models for the implementation of the recommended changes. A Petri net is an abstract technique and therefore, more difficult to understand. As a result, Petri nets are excluded.

After comparing all the techniques based on the selected criteria, Business Process Modelling Notation (BPMN) has been selected. The BPMN technique is flexible, easy to understand, and all elements are included. Although some other techniques fulfil these criteria as well, BPMN has been chosen because of its flexibility. BPMN is a well-structured technique for modelling the different aspects of processes in an organisation. The source data is provided by different kinds of processes with fundamental differences. An example of a difference is that some source data is acquired and delivered automatically while other source data is acquired and delivered manually. Therefore, BPMN is suitable for all these processes. More elaboration on the BPMN technique can be found in chapter 2.1.3.

2.1.3 Business Process Model and Notation

This section elaborates more on the chosen business process modelling technique, Business Process Model and Notation. The technique uses different kinds of elements. For clarity to the stakeholders, all used elements in the models within this research are discussed.

Business Process Model and Notation (BPMN) is a graphical representation for specifying business processes in a business process model. The technique was developed under the coordination of the Object Management Group. BPMN aims at supporting the complete range of abstraction levels, from a business level to a technical implementation level. “The primary goal of BPMN is to provide a notation that is readily understandable by all business users, from the business analysts that create the initial drafts of the processes, to the technical developers responsible for implementing the technology that will perform those processes, and finally, to the businesspeople who will manage and monitor those processes. Thus, BPMN creates a standardized bridge for the gap between the business process design and process implementation.”(Weske, 2012)

BPMN includes both simple modelling elements and extensive modelling elements. The simple elements in BPMN express simple structures in business processes while the extensive elements bring more detail to the business processes. The basis elements are easy to understand, while the extensive elements can be used when process designers are more familiar with the process. The elements in BPMN can be divided into four different categories. Each category includes different sets of elements. These categories can be found in Figure 9. Based on the used elements in this research, more elaboration is given on some sets of elements.

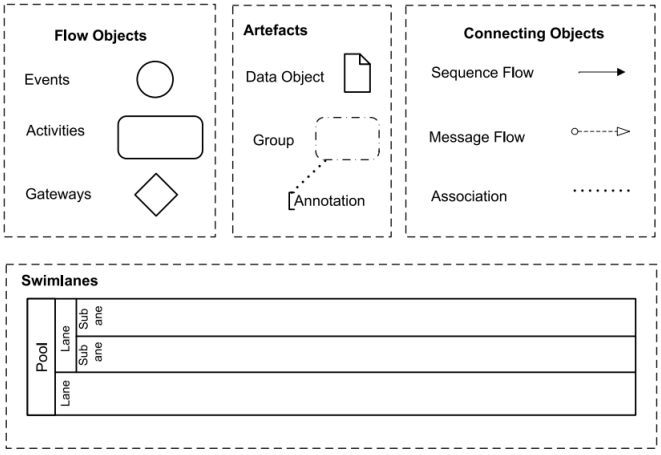


Figure 9: BPMN: categories of elements (Weske, 2012)

Figure 10: Activity types which specify the kind of task that is represented (Weske, 2012)

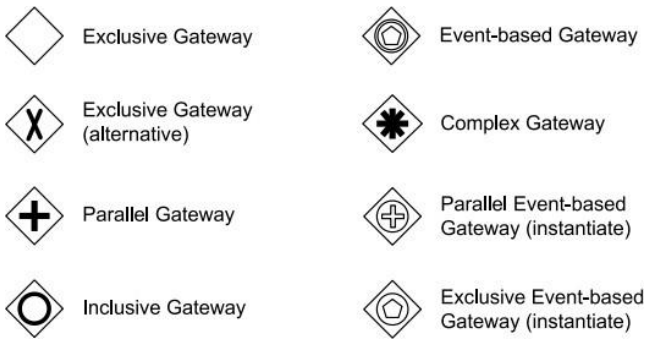


Figure 11: Gateway types in BPMN (Weske, 2012)

2.2 Capability Maturity Model Integration for WenS

This section discusses the chosen maturity model for this research. Maturity models offer a simple but effective way to measure the quality of a process. A maturity model can contribute to analysing the As-Is business process model and formulating the requirements for the future To-Be business process model.

The Capability Maturity Model Integration (CMMI) is a maturity model that helps organisations with streamlining process improvement and encouraging efficient behaviours for software, product, and service development. This maturity model has been chosen as it describes both software and system engineering. The maturity model consists of five different levels. From one to five, these levels are initial, managed, defined, quantitatively managed and optimising. For each of the levels, a detailed elaboration can be found below Figure 12.

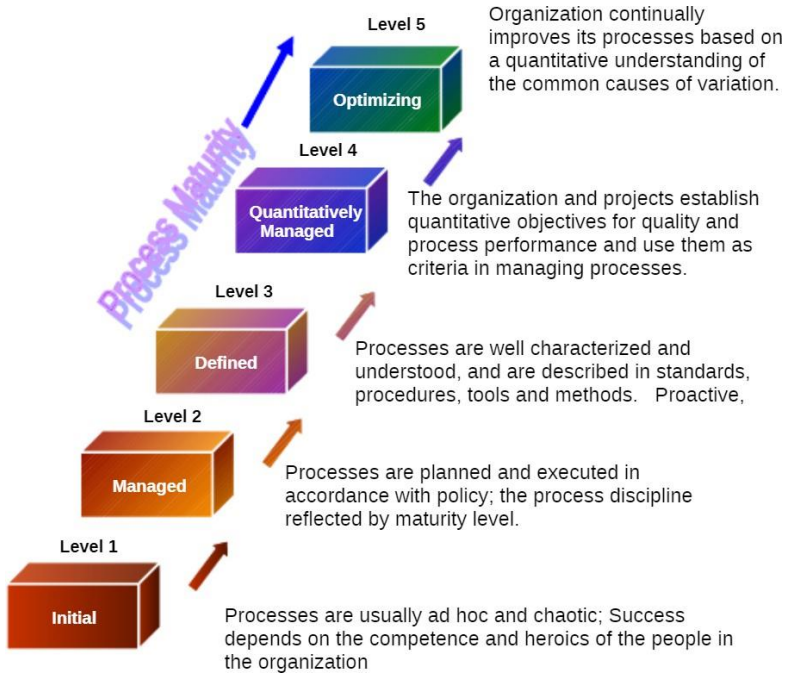


Figure 12: Capability Maturity Model Integration (Mahmood, 2016)

Maturity levels are built on the previous maturity level by adding new functionality. Level one is described as ad hoc, unpredictable, poorly controlled and reactive. Most of the time, the work gets completed. However, the work is often delayed and over budget. The second level is reactive as well. Processes are managed on a project level. The requirements for the third level are that the processes are well characterized and well understood. The organisation is mostly proactive instead of reactive. There should also be standards that guide the process. Level four means that the processes are measured and controlled. This includes measuring the quality. Level five includes continuous improvement. This is enabled by quantitative feedback from the process and piloting innovative ideas and technologies.

The CMMI will be used as a framework to assess the current situation and to formulate the To-Be model. Assessing the current situation using a maturity model will be done by an interview. Within this interview, the interviewees will be asked to what level they think the process belongs.

2.3 Performance measurements

This section provides a framework for the interviews of phase four, the demonstration phase. This includes a detailed overview of which indicators can be used to determine the performance of an Information System (IS). The questions in the interview are formulated in a way to measure these indicators. Measuring these indicators helps to assess the As-Is model and to formulate the requirements for the To-Be model.

According to DeLone and McLean (2003), the success of an information system can be measured using the model in Figure 13. These categories are system quality, information quality, service quality information system use, user satisfaction and net benefits. This research will measure user satisfaction, system quality and information quality. The reason why these three categories have been chosen can be found in the next paragraph. The next paragraph also includes an explanation of all categories and how these can be measured.

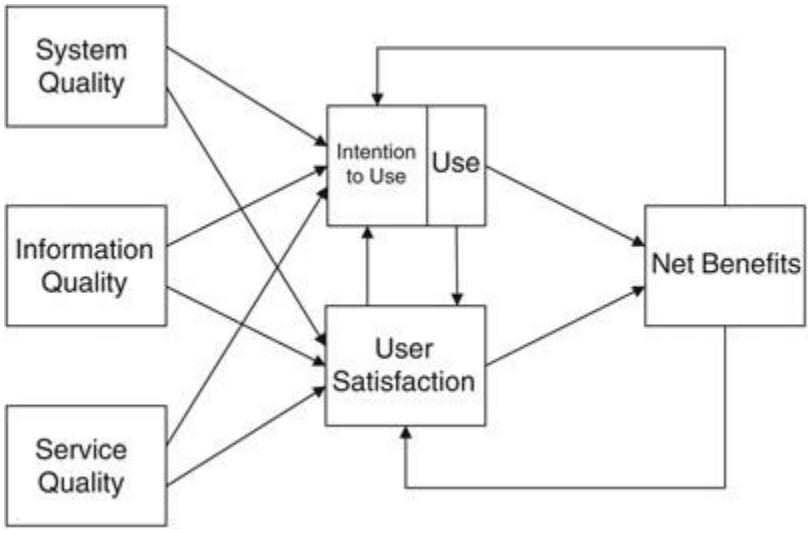


Figure 13: DeLone and McLean IS success model

System quality is a measure of the information processing itself that includes software and data components, and a measure of the technical soundness of the system (Benmoussa et al., 2018). DeLone and McLean (2003) measured the quality of the system in terms of ease of use, functionality, reliability, flexibility, data quality, portability, integration, and importance. Seddon (1997) notes that "system quality is concerned with whether there are bugs in the system, the consistency of the user interface, the ease of use, the quality of the documentation and sometimes the quality and maintainability of the program code".

Information quality means measuring the performance of the output of the information system. This output is primarily in the form of reports. Information quality (IQ) is measured in terms of accuracy, timeliness, completeness, relevance, and consistency (Benmoussa et al., 2018).

Service quality is a measure of the quality of information system services. Service quality is a support of users by the IS department, often measured by the responsiveness, reliability, and empathy of the support

organisation. (Benmoussa et al., 2018). The department information system services are not included in this research. Therefore, this is not measured.

System Use refers to the use and exploitation of outflows from the information system which is an expected future consumption of an IS or its output. (Benmoussa et al, 2018). This is outside the scope of acquiring and delivering data sources and therefore not included in the interviews.

User satisfaction is the degree of user satisfaction. This is measured by how users perceive the system while using it.

An overview of how to measure the performance of WenS can be found in Table 2. In this overview, the most important indicators have been chosen based on WenS. Per category, examples that could occur within the acquisition and delivery of source data for WenS are shown. Together with the maturity model, this overview is a framework for the interviews

Table 2: Categories with their indicators

Category	***Indicators***	***Example***
System quality	Ease of use, reliability, data quality and quality of documentation	- Acquiring and delivering source data is manual, and therefore not scoring high on ease of use - Quality of documentation is lacking
Information quality	Accuracy, timeliness, completeness, and relevance	- Missing data in files delivered to WenS - Too much data delivered
User satisfaction	Satisfaction	- Lots of problems every month which cause lower satisfaction

3. Current situation

This chapter sketches a detailed overview of the current acquisition and delivery of all the source data for the WenS process. To sketch a clear overview of the acquisition and delivery of source data for WenS in a systematic way, it is assumed that every sub-process consists of input, process and output (Long, Keng, & Ling, 2005). This chapter answers the question "What does the As-Is model of acquiring and delivering the source data look like?".

3.1 Infra data

This section describes the infra source data which is delivered monthly. Every month, two files are acquired and delivered for WenS. Section 3.1.1 describes the acquisition and delivery of static information about the tracks while section 3.1.2 describes the acquisition and delivery of infra data about the switches.

3.1.1 Track data

In this section, there is an overview of the current acquisition and delivery of the track source data for WenS. The business process model can be found in Figure 14. Acquiring the track source data is an automatic process done monthly by a SAS script. This is a statistical analysis system (SAS) that can be used for advanced analytics, business intelligence, data management, and predictive analytics. To complete the track source data, several files are needed:

- Branche Breed Monitoring Systeem (BBMS) file
This system is updated once a half year. All known tracks, including tracks not monitored by ProRail, are included. These tracks have a unique name.

- ProRail Monitoring Platform (PMP) file
This file is monthly acquired by a SAS script. The original purpose of the PMP file is to monitor switches that are maintained by ProRail. The file includes a unique code for each switch. This unique code is used to combine switches and tracks.

- Log file
Every month, Materieel Impact receives a log file from WenS. This logfile includes information about the switches that could not be connected to a track or the other way around. If a switch can not be connected to a track, the track is undefined and therefore missing. Materieel Impact reviews this and adds the tracks to the track file manually if needed.'

- Most recent track file
This file is used to complete the characteristics of a track. This track file is the output of WenS from last month.

These files are combined every month to create the track source data. This is a new process and still in progress for improvement. The process is described below in Figure 14. The old process was not updated from January 2020 till April 2021. As a result, Materieel Impact saw a rapid change in missing tracks. Therefore, this new procedure has been set up since April 2021.

The process starts when a log file is received by Materieel Impact. This log file includes missing tracks in the output of WenS from last month. Based on this log file, relevant tracks are added to the most recent track file. Together with the BBMS and PMP file, it is combined into the track data source. If combined correctly, the track data source is sent to WenS via FTP.

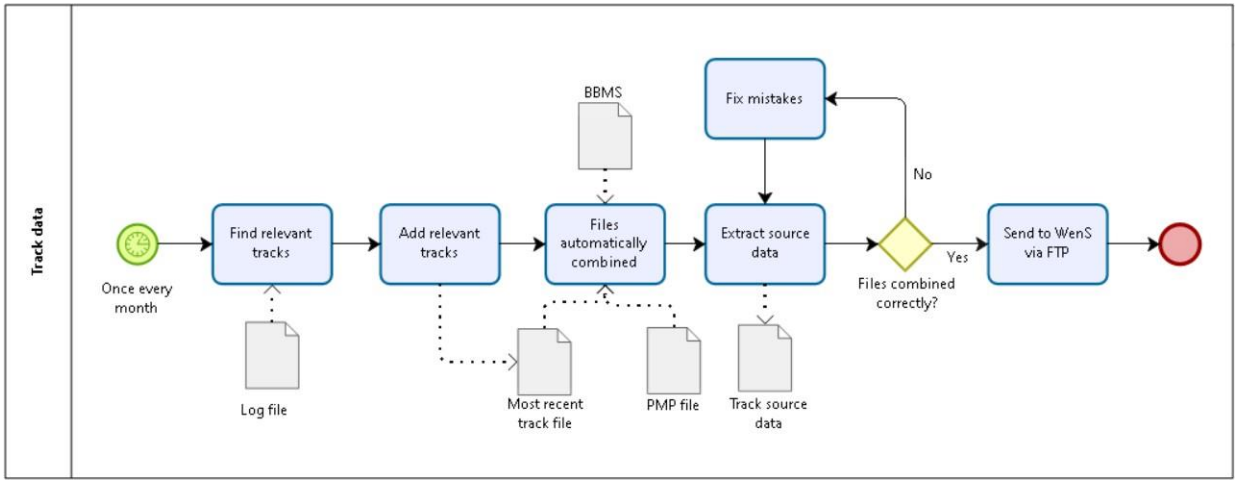


Figure 14: Process to acquire and deliver track source data

3.1.2 Switch data

In this section, there is an overview of the current acquisition and delivery of the switch source data for WenS. The business process model can be found in Figure 15. Acquiring the switch source data is a manual process done every month.

The process starts once every month. The goal of the process is to add and/or remove switches in the switch source data. To create the switch source data for next month, one of the two information specialists within Materieel Impact analyses the switch output of WenS from last month. By focusing on certain columns in the output, it can be defined if a switch is still active or if there are any new switches. Inactive switches are removed while new switches are added to the switch source data. Characteristics of the new switches can be found in several different databases and are manually added to the new switches in the switch source data. Characteristics of the switches which are still active are kept the same and used again for next month. After all the old switches have been removed and the new switches have been added, including the characteristics, the switch source data is extracted and put on the FTP server.

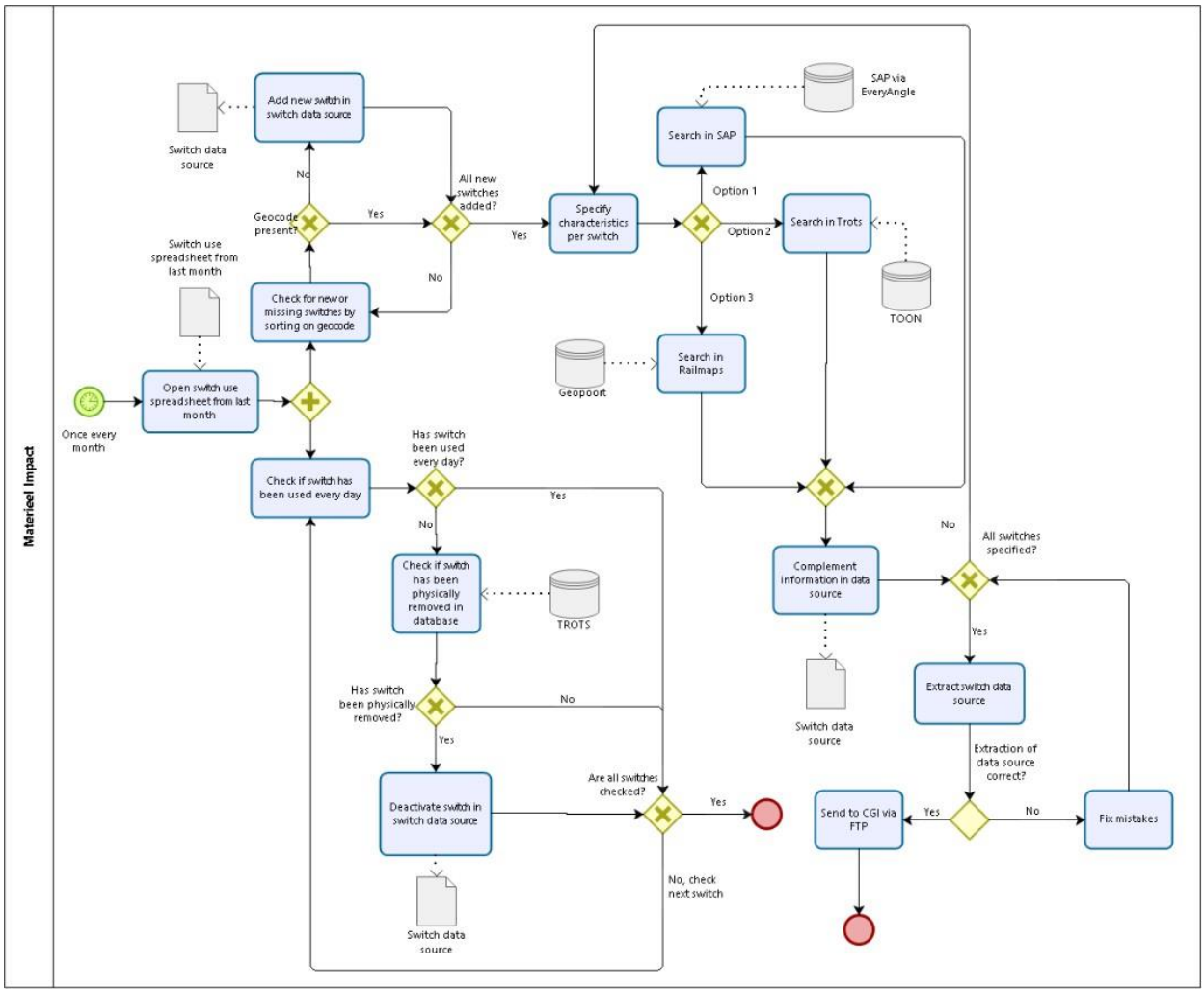


Figure 15: Process to acquire and deliver switch data

3.2 Quo Vadis data

In this section, there is an overview of the current acquisition and delivery of the Quo Vadis source data for WenS. The business process model can be found in Figure 16.

Acquiring and delivering the Quo Vadis data source is an automatic process done by a query. The process runs every night. The first step is to remove unwanted rows from the input. The enriched match results contain both approved and disapproved measurements and valid and invalid train numbers. The enriched match results are the result of connecting the train numbers to Quo Vadis data, Only the approved measurements and valid train numbers are selected for delivery. After the selection of these rows, only one record is selected due to the same data occurring multiple times. In this record, the distinct measure days are determined. These rows are added to the DMA history file. The distinct measure days are added to the truncated auxiliary table.

To acquire and deliver the Quo Vadis data for WenS, the automatic process makes use of an auxiliary table. Using the auxiliary table, all records with the same measure day are selected from the DMA history file. The purpose of the auxiliary table is to combine the records from measure day "X" in this process with the records from the previously delivered measure day "X". As a result, the same data may be delivered multiple times. The only difference is that the new delivery also contains new data.

After all records are selected in the DMA history file, the next step is to create an XML file that conforms to XDS. One XML file is created per measure day. Using a file distributor, the file(s) are sent via FTP to WenS

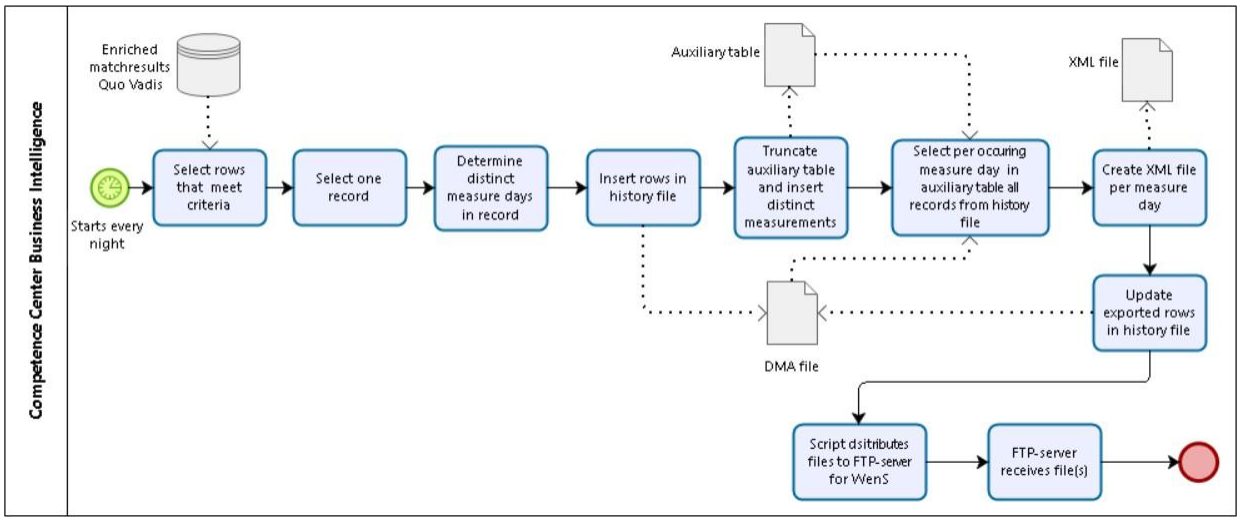


Figure 16: Process to acquire and deliver Quo Vadis data

3.3 TROTS data

In this section, there is an overview of the current acquisition and delivery of the TROTS data for WenS. The business process model can be found in Figure 17. Acquiring the Trots archives is an automatic process done every hour by multiple scripts. Manual actions are needed when delivery is not correct.

The process starts every hour for every traffic control system. As said before, there are thirteen traffic control systems. This means that at least 312 (13x24 = 312) archives are delivered each day. The process starts every hour in the Post21 domain. This domain consists of all the primary ICT activities within ProRail. Therefore, Post21is heavily secured, and archives are first copied to FTP server IPstep before they are distributed. This server is in the office automatization (KA) domain. Then, a script retrieves the archives and puts the archives on a windows server. The script distributes the archives to multiple locations. One of the locations is the FTP server for WenS. If the complete process has been executed correctly, the process stops here. This is not always the case. If not executed correctly, the Central Service Desk (CSD) receives a notification of an incident. The CSD makes sure that the archives are put on the IPstep again. The process restarts.

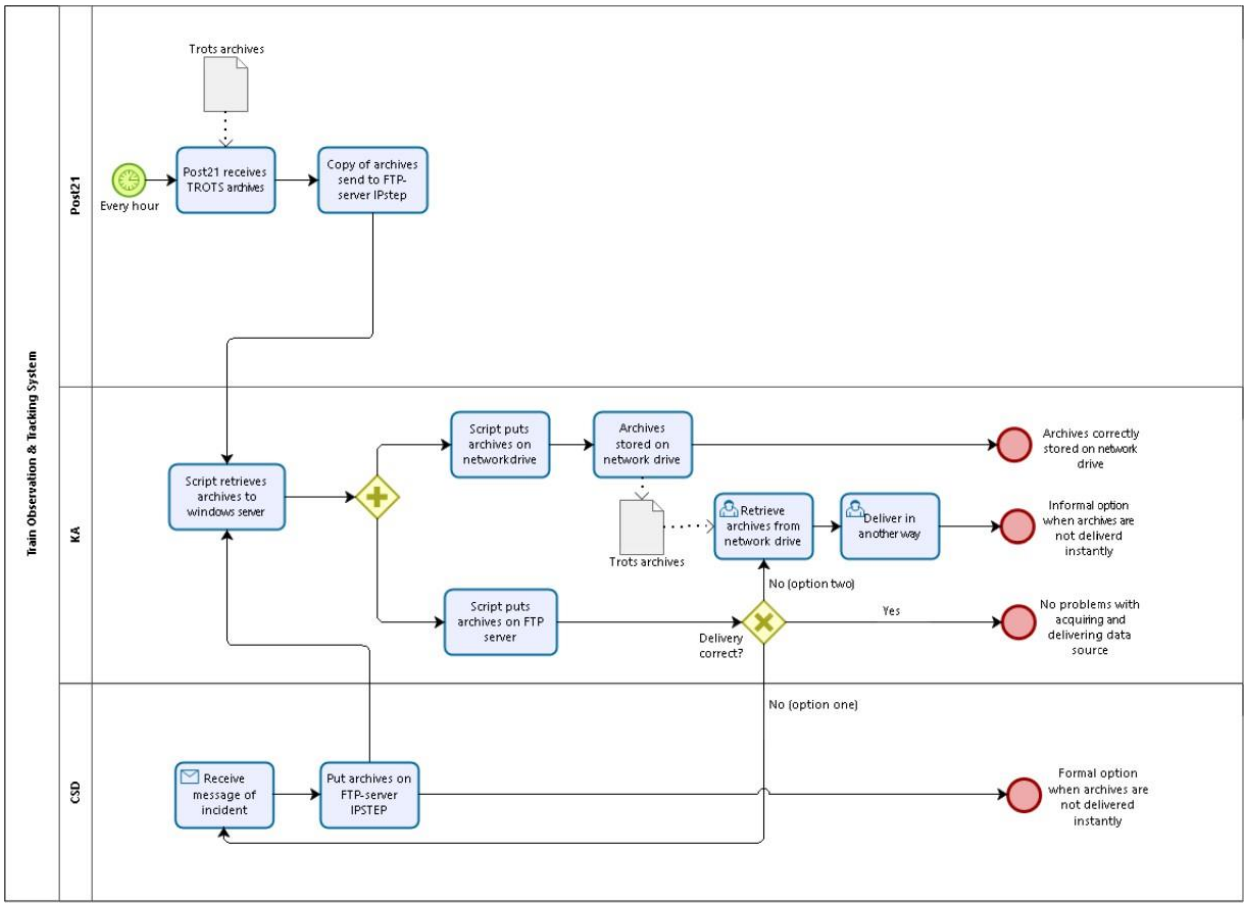


Figure 17: Process to acquire and deliver TROTS archives

3.4 TNR data

This section describes how the train number range (TNR) data is acquired and delivered for WenS. The process of acquiring and delivering the source data can be found in Figure 18. The TNR data is used by WenS to assign what kind of train has driven a switch or track. Within the TNR data source, every train number is assigned to a carrier. Then, a carrier is assigned to a category. These categories can be a passenger train, freight train or other train. A portal provides insight into the used train number. TNR is continuously updated and contains no “frozen” positions.

There are several steps to acquire and deliver the data source. The steps are carried out by one of the two information specialists within Materieel Impact. There is no specified time when this task is carried out. Only when a team member of Materieel Impact feels there is a need, the TNR data source will be updated on the FTP. An estimation is that this happens not more than once a half year.

The process is manual, but not time-consuming. In total, the process takes less than a quarter of an hour. A team member of Materieel Impact goes to the site www.tnr.prorail.nl. The site contains up-to-date information about the train numbers. With a few mouse clicks, an excel sheet can be generated and downloaded. The excel sheet is the data source that is needed for WenS. The data source is put on an FTP server for further use by WenS.

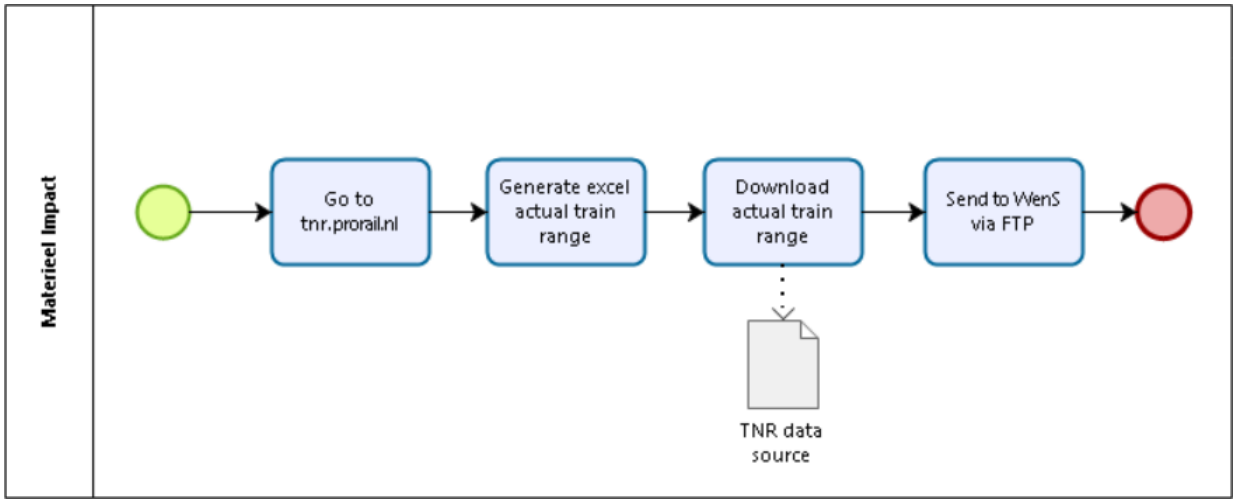


Figure 18: Process to acquire and deliver train number range data

3.5 Configuration data

In this section, there is an overview of the current acquisition and delivery of the configuration data for WenS. Configuration tells us where the tracks and switches are located. The business process model can be found in Figure 19. Acquiring and delivering the configuration data is a process done by RIGD-Loxia, a company working exclusively for ProRail. RIGD and Loxia are two separate companies. These two companies are working together to create what is needed.

RIGD-Loxia creates configurations for the traffic control system of ProRail. This is done by creating a digital map for each traffic control area. Then, the digital map is put into the traffic control system. This makes sure that the traffic control system knows which infrastructure is present.

The process starts after ProRail sets up a project. This can be any kind of project like changing a track or switch. ProRail sends the input specifications of the project to RIGD-Loxia. RIGD-Loxia uses the input specifications to change their designs. The new designs are added to their database. Then, the designs are read into data. The software automatically collects the necessary files into a ZIP file. The last step is to put the ZIP file manually on the FTP server. RIGD-Loxia and ProRail have agreed that this should be done one week before the project is placed in service. This means that there is a delivery for every new project.

ProRail has divided its railway network into thirteen different traffic control posts. Because of this, RIGD-Loxia also divides delivery into thirteen different ZIP files. Each ZIP file is named after the traffic control post and includes a version number. The ZIP File contains four different kinds of files. “RP” files are visualizations of emplacements which the traffic control uses in their traffic control systems. “IA” files are files that contain specific information about the railway infrastructure. “TPSA” files contain measuring points. These measuring points can be used to determine the punctuality of trains. The last file is a TROTS SQL file.

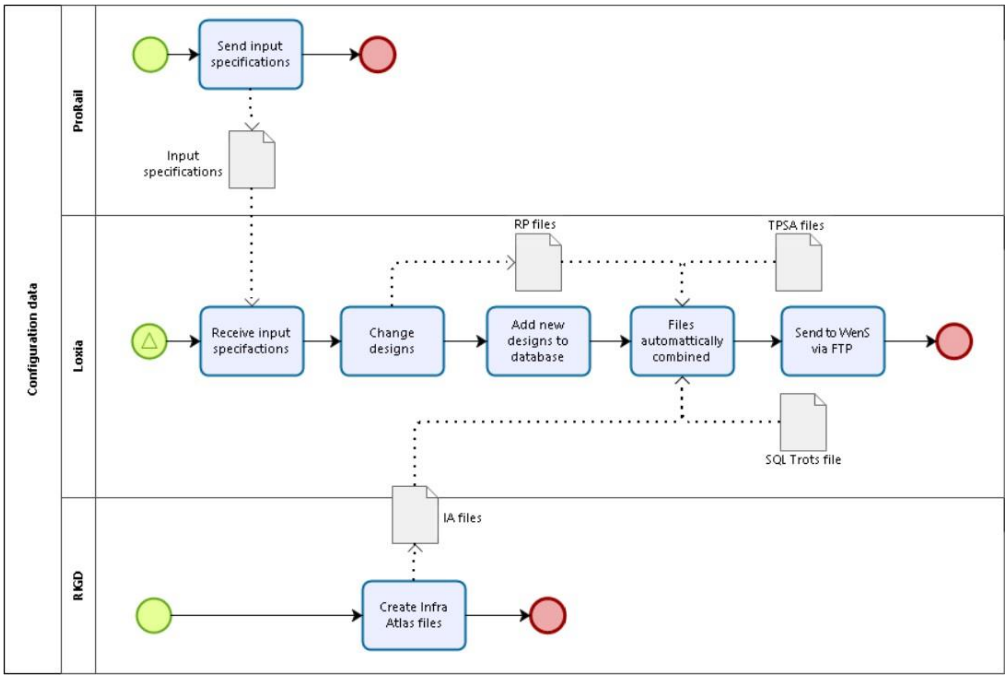


Figure 19: Process to acquire and deliver configuration data

3.6 Problems occurring in the source data

For the interviews in round one, to determine the current situation, interviewees were already asked about challenges in the process. In this section, all challenges mentioned per data source can be found. The interviewee from configuration data could not name any challenges. Also, this data source is outsourced to RIGD-Loxia, a company working exclusively for ProRail. Therefore, improving this data source is less of a priority for ProRail. As a result, this data source is not taken into account for round two.

Track data
The data from the most recent track data source has been used to fill in the columns of the tracks till January 2020. This data has not been reviewed for a long time and can therefore be incorrect or incomplete.

Since the new procedure, tracks that do not exist anymore, have been deleted. The track source data also includes many new tracks. Completing the information on each of the new tracks is challenging. Characteristics of the new tracks are not filled completely as the information is not available in the most recent track file. Materieel Impact is struggling on how to acquire the missing information. When this information is missing for a certain track, the output of WenS does not specify to which load class the track belongs.

Switch data
To complete the characteristics of the switches, the output of WenS is used. This means that the same data is used over and over. Changes will stay unnoticed. This leads to the fact that the switch and track data may contain many errors.

Quo Vadis data
New Quo Vadis data for a certain day is delivered every day. This means that Quo Vadis data for a certain day is sometimes delivered in multiple days if there is new data. Team members are questioning whether data should be delivered every day, as WenS only runs once a month. This could prevent chaotic delivery on multiple days.

TROTS archives
Distributing the archives often goes wrong. Deliveries are incomplete, files are corrupt, or ZIP files are empty. This leads to much frustration within the acquisition and delivery of TROTS archives.

TNR data
Acquiring and delivering the TNR data is an ad-hoc process. The process starts when a team member of Materieel Impact feels there is a need to start the process.

4. Interview results

Team members that were involved in the process on the receiving and delivering end of the data have been interviewed. On the delivery end of the data, four persons were interviewed. One of the persons was interviewed on three different data sources. This means that a total of six interviews has been conducted on the delivery end of the data. For the receiving end of the data, one person was interviewed. This person is employed at CGI. As this person has the complete overview of the receiving end of the data, only one person was considered as enough. To conclude, five different team members have been interviewed by conducting seven interviews in total.

4.1 Method

The set-up for the semi-structured interviews can be found in Appendix E. Reasoning on why semi-structured interviews have been used can be found in section 1.4.3. The interview contains four parts and contains both open and closed-ended questions. All the interviews had the same structure. First, interviews were asked about their employment at their company. These questions included their function, their department, how many years they are employed at their company and the amount of time involved in the WenS process. These were the same for all interviews. Then, open-ended questions were asked. This part was semi-structured for all interviews. Follow-up questions changed depending on the answers of the interviewees. The goal of this part was to determine different challenges and bottlenecks in the WenS process. For analysing this part, Atlas.ti was used. Using Atlas.ti, answers from the interviewees were coded and organised. Then, for the next part, a maturity model was showed to the interviewees. The interviewees pointed out to which level they think the process belongs. If not given immediately, an explanation was asked.

4.2 Interviewees

For this research, five different team members have been interviewed according to the interview setup. As one of the team members is responsible for three data sources, this team member has been interviewed for each of the three data sources. All of the information of the interviewees can be found in Table 3.

Table 3: Information on all of the interviewees

Team member:	Function:	Department/Company:	The number of years employed at the company:	Amount of time involved in WenS process*
Quo Vadis	Product owner advanced analytics	Competence Center Business Intelligence	11	7
Switch	Information specialist	Materieel Impact	5	5
Switch, TNR and Track	Information specialist	Materieel Impact	4	1,5
TROTS	Product manager WenS	Logistics	13	1
CGI	Developer and tester WenS	CGI	2	2

*The name of the process changed over the years. Since 2018 it is called WenS. The data covers the process since it has been set up in 2004.

In Table 3 can be seen that the longest active team member has been involved for seven years. However, the switch use and track load data exist since 2004. This was set up and managed by an ex-team member till 2018. In 2018, this team member stopped due to other duties. This means that there are no active team members with the original knowledge at this moment. Current active team members do not have complete knowledge of how the WenS process has been set up. Due to this, the interview results might not completely reflect the real situation. However, with an experience of at least one year for each interviewee, it is assumed that the results are close to reality.

It can also be said that three out of five team members are relatively new. These team members have between one and two years of experience. Two out of five team members have been working with the process since they started working at ProRail. The other three team members started working with the process three and a half to twelve years after they started working at ProRail. This shows us that there is a rotation of team members. Therefore, transferring knowledge is important. Knowledge can be transferred by using documentation. Later in this research, documentation per data source will be analysed.

4.3 Interpreting and coding the respondents' answers

The set-up for the questions asked can be found in appendix E. The exact questions asked depended on the answers given by the interviewees. Appendix F can be used to find the exact questions which were asked per interview. Relevant answers from the respondents were coded. This table has been set up by open and axial coding. First, the data was broken down into discrete parts and codes were used to label each of these parts. This is also known as open coding. Then, axial coding was used. All different codes were connected into different categories. The results from the interviews can be grouped into the

following categories: information need, automation, availability of data, data quality, communication and data delivery.

Switch

Table 4 provides an overview of the most important quotes in the switch interview. The quotes will be explained further below the table.

Table 4: Important quotes from switch interview

Interview	Quote	Open coding	Axial coding
#1	"It once ran via an Access database. This database was linked to all other kinds of Access databases to collect input data. That part is not clear. It is not clear whether this still works properly. This was once set up like this, it is not clear to us whether this works well and how it works. That's why it is difficult to execute the process for us."	Unclear process	Information need
#1	"The process is not clear, which means that the automatic process cannot be performed. A work-around for this is to enter it manually."	Manual work-around	Information need
#1	"We should create the switch data source in another way, automatic."	Process change	Automation
#1	"These are the switches (in output WenS, from TROTS), what information do we need about these switches? Then, we should extract this from the file (output WenS) every month. In this way, you always get an up-to-date situation."	Use output WenS	Automation
#2	"The ex-product owner has a lot of knowledge and experience. So far, we've lingered a lot in the past. That is why I think it could be a lot more efficient by, for example, extracting data directly from systems."	Extracting data directly from systems	Automation
#2	"We use our own database in which we store all data. We do this while this same data is actually also centrally available and maintained."	Centrally available data	Availability of data
#2	"Updating the data source, every month, would be better. At the moment, we now only look for new switches or removed switches. We do not look at every single switch to see whether, for example, the speed, angle ratio, etc. is still good. We don't check that."	Obsolete data	Data quality
#2	"We should actually deliver the same data source as what we are delivering now, but we have to find out where we can obtain all the data."	Unknown location of data	Information need

The interviewee from interview one indicates that the current process is not clear. The interviewee does not know whether this current process is still working properly. Due to not knowing how the process exactly works, the interviewee found a manual work-around to complete the process each month. Interviewee two indicates that the current data in the data source might not be correct anymore. Certain characteristics of switches might have changed over the years. This is not checked in the current process.

Both interviewees indicate that the process should be automated. Interviewee one suggests extracting all the switches from the output of last month. These switches are kept up to date by TROTS. After

extracting all the switches, the needed corresponding characteristics per switch should be found and added. Interviewee two indicates that they store the data in their database. According to interviewee two, this is not efficient as the same data is also centrally available.

However, to achieve this, both interviewees do not have the full knowledge yet on where to acquire all data. Interviewee one is wondering what information is needed about the switches. Interviewee two indicates that the team has to find out where all data can be obtained.

Track

The table below Table 5 provides an overview of the most important quotes in the track interview. The quotes will be explained further below the table.

Table 5: Important quotes from track interview

Quote	**Open coding**	**Axial coding**
"The data source which is built depends on many other systems where we acquire data. It is difficult to track down where this data is precisely located and how qualitatively high the data is. For example, it is unclear how many times the data is updated. There are also formulas for the fictive day tonnage, these formulas are pretty old and the background on this is also not completely clear. This is the case because it has been transferred from the previous product owner to us. The previous product owner was a specialist and was deep in it. Because of this, it is a steep learning curve for us."	Uncertainties about data quality	Data quality
The previous product owner had a certain system for compiling this data source, this system worked perfectly for this product owner. But the system was difficult to transfer. A lot of knowledge is required, and this knowledge was mainly in hands of the previous product owner. We tried a lot of times to keep going with the system of the previous product owner, this included many manual actions, but we are now looking to do it in another way. All data is available, it just needs to be added together properly.	Transferring of knowledge	Information need
"We are now doing it by using SAS scripts, automatic scripts which create the data source. At the moment, however, it is not complete. Making the data source complete is what we are doing right now, and that is actually the hardest part. Completing includes the formulas, speeds etc."	Incomplete data	Data quality
"We are looking where the data comes from and whether we understand the data well. Actually, we don't understand the data very well. We know what the values are and how the values should be shown, but we do not know why it is that way. That is still a tricky part."	Understanding the data	Information need
"We are still missing speeds and other standards. Those are difficult to find. We are still adding those manually. We would like to automate this"	Missing data	Data quality
"We have to ask ourselves; how can we acquire the data? Or is there another location we could acquire the data? At this point, things can be downloaded manually, but you want to have direct access to them."	Direct access to data	Automation

After axial coding, three different categories were found. These categories are current situation, information need and automation.

For the current situation, the interviewee is unsure about the data quality. The interviewee does not know where the data is originally from due to a change of product owner. Also, the formulas for fictive day tonnage are old and there is no knowledge on the background of these formulas. The interviewee also says that the track data source is missing data.

For the information need, the interviewee indicates that he has not the original knowledge of the track data source. Because of this, it was difficult to keep working with the previous system. Therefore, a new system was set up. However, the data is not understood very well which makes it difficult to complete the new system at this moment.
In the end, according to the interviewee, there should be an automated process where data is directly acquired to create the data source for WenS.

TNR
Table 6 provides an overview of the most important quotes in the TNR interview. The quotes will be explained further below the table.

Table 6: Important quotes from TNR interview

Quote	**Open coding**	**Axial coding**
"WenS uses the train number ranges to see what kind of train it is. These train number ranges are somewhere in a system. The train number ranges are extracted manually once in a while and then used for WenS. There is no fixed time when this should happen. The train number ranges are quite static but do change."	Obsolete data	Data quality
"The data which we need is available for ProRail. We can just extract it from the TNR environment, download it to Excel and then upload it to WenS. In principle, this could be automated. The ideal situation is that a connection would be built from TNR to WenS. WenS would then directly extract the train number ranges from TNR and not via a half-hearted action from an employee."	Directly acquiring data	Automation
"It should be looking into whether WenS can directly access it. Can WenS directly extract data from TNR? Or is an in-between step needed, and should it be done via a Business Intelligence environment? These are the different steps that should be looked into. Then you would look into a solution where the data is monthly retrieved from BI and then put on the FTP."	Possibilities for directly acquiring data	Information need

After axial coding, three different categories were found. These categories are current situation, automation, and information need.

The interviewee tells us that if the TNR data is acquired, the data is good. However, the data is not acquired every month. As a result, the data used by WenS might be out of data and therefore obsolete.

According to the interviewee, the ideal situation is a connection between TNR and WenS. This would mean that the TNR data is automatically sent to WenS every month. However, to achieve this, it should be looking into whether WenS can do this and how. An option suggested by the interviewee is to do this via the Business Intelligence environment.

Quo Vadis

Table 7 provides an overview of the most important quotes in the Quo Vadis interview. The quotes will be explained further below the table.

Table 7: Important quotes from Quo Vadis interview

Quote	Open coding	Axial coding
"During the project, the project team asked for some kind of shortcut to get that data. Apparently, we (CC BI) said: "yes that is possible, but only temporary." But then nobody actually took the step to convert that into a permanent situation. Then, at a certain point, there was simply no more data available."	Responsibility	Communication
"Data of 1 day is spread over 2-3 days, not entirely clear whether this always went well. Maybe it is better to no longer deliver daily because WenS only needs it monthly. So, delivering weekly or monthly might be better."	From daily delivery to monthly delivery	Data delivery
"There has not been much contact between CGI and CC BI, so there is not much understanding of what is being done with the data, how important it is, and the wishes are wishes that will arise in the future. There is no process to look further into this together. Materieel Impact could or should play a role in this. However, as an example, it is unclear to me how the organisation of that team looks like and how it all happens on that side. It is not always necessary, but it can be useful to think along with each other to prevent problems."	Mutual understanding	Communication
"Over the years, little has changed in the structure of the data source. So, few people had hands-on experience with it. The risk of this is that you don't really understand how it was originally conceived."	No experience with the current process	Information need
There was not really a nice way to do an after-delivery of data, so if data was missing, we kind of hacked into the system, so to speak. That could be better and automated"	Work-around after-delivery	Automation

According to the interviewee, there has been a situation where the Quo Vadis data delivery stopped. The interviewee said that this happened because no one changed a temporary process into a permanent process. A cause of this might be unclear responsibilities for each team member.

The interviewee suggests that an increase in mutual understanding between the different teams in the WenS process would be beneficial. According to the interviewee, this would help in the future for improving the WenS process. If teams understand each other better, it is easier to get to a solution together.

Within the Quo Vadis, the team members have not had much hands-on experience with the process. The reason for this is that the process has not changed much over time. Therefore, team members did not have to investigate the process and understand it.

The interviewee tells us that sometimes data is not delivered. If this is the case, an after-delivery has to take place. Now, there is no efficient way of doing so. The interviewee thinks that this could be done better and automated.

TROTS

Table 8 provides an overview of the most important notes in the TROTS interview. Due to confidentially issues, notes have been made instead of quotes. The notes will be explained further below the table.

Table 8

Table 8: Important quotes from TROTS interview

Notes	Open coding	Axial coding
It all went wrong in mid-June. Different responsible persons in the chain, everyone pointing at each other. It is solved now, nobody knows how. Because of this, priorities are low again. Responsibilities are not clear.	Responsibility	Communication
TROTS to post21 works perfectly. To IPSTEP too. IPSTEP to Windows server usually. Less stable from KA.	Unstable process steps	Technical issues
AM communication and logistics must be improved.	Improving communication	Communication
Get rid ff IPSTEP. Unlock data from ESB. ESB is Enterprise Service Bus. From scratch. AM provides requirements, Logistics builds after that. The initiative is now from logistics.	Unlocking of data	Data delivery
Different processes, but deliver data sources in a uniform way?	Uniform data delivery	Uniform data delivery
Bug in Windows server. Script/scheduler (for new archives) starts up multiple times and does not work properly. The system must be restarted manually regularly.	Bugs in windows server	Technical issues

To acquire and deliver the TROTS archives, several parties are involved. These parties can be found in Figure 17: Process to acquire and deliver TROTS archives (chapter 3). The interviewee indicates, due to more than one party, that the responsibilities within acquiring and delivering the TROTS archives are not always clear.

The interviewee tells us about two technical issues within the process of acquiring and delivering TROTS data. The first issue is a stability problem. The data is usually correctly transferred from IPSTEP to the windows server, but this is not always the case. The office automation environment is even less stable. The second issue is a bug in the windows server. The schedular, for new archives, starts up multiple times and does not work properly. As a result, the system has to be restarted manually regularly.

The interviewee also suggests removing IPSTEP. This would result in fewer steps within the process of acquiring and delivering TROTS data. A decrease in steps means that there are fewer possibilities for the process to fail. Therefore, this would be beneficial. The interviewee brings up how the new situation should look like. The new situation would be to unlock data from the Enterprise Service Bus (ESB). Materieel Impact should provide the requirements so that the logistics team can build the new process based on these requirements.

CGI

Table 9 provides an overview of the most important quotes in the CGI interview. The quotes will be explained further below the table.

Table 9: Important quotes from CGI interview

Quote	Open coding	Axial coding
‘’FTP is actually a pretty outdated standard and I think CGI has stopped supporting FTP since June of this year. So, we said to ProRail: ‘guys, we have to switch to …’ Well, then we offered a few alternatives, basically FTP over tls, so with encryption or SFTP, the successor to FTP. In first instance, ProRail said: ‘okay we are going to use SFTP.’ We got an issue to convert WenS to SFTP. Well, after we did so, ProRail came back to us with: “we are not going to build an SFTP server anymore”, so now I have to convert it again.’’	Miscommunication on changes in process	Communication
The ProRail FTP sometimes has the problem that if there are a lot of people on it at the same time, it kicks out other clients. This happens to us sometimes. For hours long, it says that we can not connect.	Overload FTP server	Technical issue
“We are told that very big changes are about to happen while there are not. A lot of words were spoken on changing the direction of the tracks in SAP. Those are things we run into. Yes, they do not really have to do with the technical process itself, but they do create a certain amount of stress for us.”	Miscommunication on amount of work	Communication
“What could help, and what we probably should do, is to take ProRail through the process to show: ‘look, it is quite robust, we can deal with some change in this way.’”	Mutual understanding	Communication
“The problem, and we've had that a number of times, is when you can't access the (FTP) server for a while. Then, if both the original file and the after-delivery are on the server in different folders, it can happen that the after-delivery imports first and then the original. So, the after-delivered data is overwritten instead of the original data. As a result, we actually get outdated data here.”	Outdated data due to overload of FTP-server	Technical issue

CGI indicates that the communication between the delivering and receiving end of the data should be better. Several situations have indicated that this should be improved. For one situation, there was a planned change for the FTP server. CGI worked on implementing the change, while in the end, this change was not executed. This resulted in unnecessary extra work. Also, CGI indicates that there are misunderstandings in the expected amount of work that a change in the process will take. On the delivering end of the data, changes are generally expected to be more work than they are. This negates the planning of CGI and brings a certain amount of stress. The interviewee indicates that it might help to show the product owner the process on the receiving end.

There are also some technical issues where CGI is dealing with. One of them is an overload of clients on the FTP server. If this happens, there is a possibility of CGI being kicked out of the server. Normally, this is not that big of a problem. However, if this happens at the time when CGI has to combine all data sources and create the output of WenS, the output of WenS might be delivered late. Also, when the FTP-

server can not be accessed, the original data may overwrite the after-delivered data. When this happens, the data is incorrect.

General
Analysing all interviews together, some general themes can be found. These themes are communication, automation and information need. Below, all these themes are explained.

The theme communication has been mentioned frequently by the interviewees, both on the receiving and delivering end of the data. CGI, on the receiving end, indicates that changes are made bigger than they are. As a result, the planning of CGI is not accurate anymore. Interviewees also indicate that there is not much understanding between the delivering and receiving end of the data sources. The interviewee from CC BI, the delivering end, says that the process can improve by sharing knowledge. CGI also indicates that communication could improve by sharing knowledge between the different parties.

Automation has been brought up multiple times in both interviews of Switch data, Track data and TNR data. Interviewees in these interviews argue that how data is retrieved can be optimised. Currently, the switch data is not directly retrieved from its source. In contrast to Switch data, TNR data is retrieved directly from its source. However, an interviewee believes that the process should be automated. Now, the data is retrieved when a team member feels there is a need to do so. This does not happen every month. As a result, data gets outdated. For the Track data, the interviewee tells us that they are working on automating the process. However, according to the interviewee, there is still data missing.

The theme “information need” has also been mentioned often. Interviewees do not exactly know the quality and background of data. The following quote from the track data covers the load: “The data source which is built depends on many other systems where we acquire data. It is difficult to track down where this data is precisely located and what the data quality is. There are also formulas for the fictive day tonnage, these formulas are old and the background on this is also not completely clear. This is the case because it has been transferred from the previous product owner to us. The previous product owner was a specialist and was deep in it. Because of this, it is a steep learning curve for us.”

4.4 Maturity level

All interviewees have given their opinion about the maturity of the delivery of their data source. The interviewees determined the maturity level by themselves in the interview. The maturities levels shown to the interviewees can be found in Figure 20. The description per maturity level was also translated to Dutch. This can be found in appendix E.

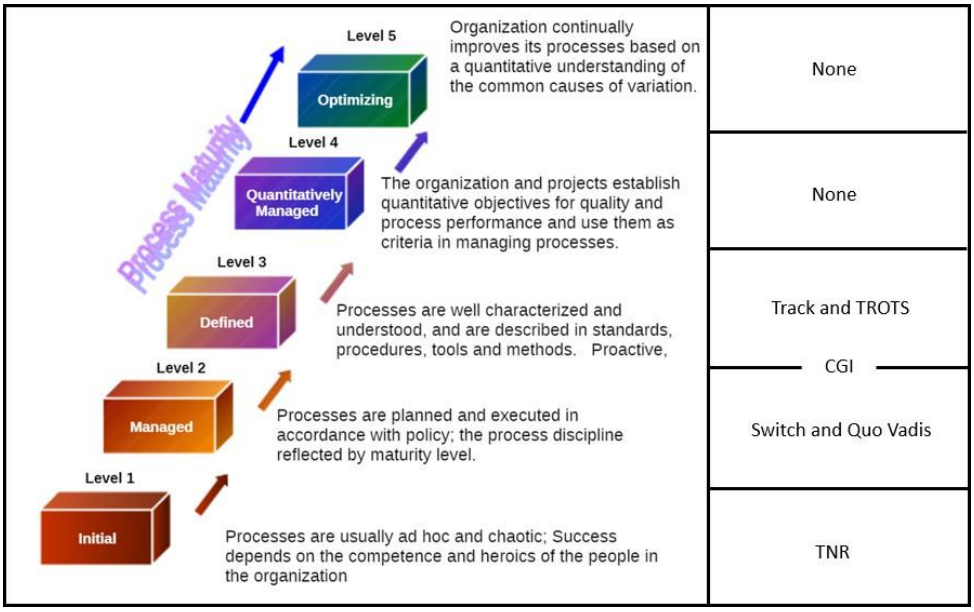


Figure 20: The maturity model shown to the interviewees

None of the processes scores a level four or higher. Interviewees also explained why they assigned a certain maturity level to a data source. The keywords per data source can be found in Table 10. Table 10 shows that for the Switch data, two persons have been interviewed. They both assigned level two as the maturity level. Each keyword is assigned the colour green or red. Red represents a reason for not assigning a higher maturity level whereas green is assigned when it is a reason for assigning a certain maturity level. No colour was assigned when it was neither positive nor negative.

Table 10: Maturity level and keywords per data source

Data source:	**Maturity level(s):**	**Explanation:**
Switch	2 (2x)	Out-dated documentation (2x), repeatable, no improvement but improvement possible, no improvement
Track	3	Good documentation, managed, much space for increasing efficiency process
Quo Vadis	2	Repeatable, no improvement, stable, bad documentation
TNR	1	Ad-hoc, no documentation, low risk
TROTS	3	Repeatable, consistent, improved when incidents occur
Delivery of all data sources (CGI)	Between levels 2 and 3	Repeatable, good documentation, no development

TNR

For the TNR process, the interviewee assigned level one as the maturity level. There is no documentation available on this process. The interviewee argues that the process is ad-hoc but also low risk. The interviewee says: “I do it when I think about it and when I have the time to do it. I do not think anyone else would do it if I would not do it. This is not even that bad, which makes it worse.” This quote shows us that it is a small process and that there is not much risk when it is not done. Based on the maturity model, level one is the right maturity level for TNR.

Switch

For the Switch process, level two was assigned by both interviewees. The interviewees argue that the documentation is old and unclear. Both interviewees claim that there is no improvement. One of the interviewees says: “At this moment, we are working with the idea that we are doing it in a certain way, so we keep doing it that way.” One of the interviewees indicates that for the future, a maturity level between level three and level four would be sufficient. Based on the maturity model, level two is the right maturity level for Switch.

Quo Vadis

For the Quo Vadis process, the interviewee assigned level two. One of the reasons for assigning level two is that there is no documentation. Another reason is that the process is not reviewed. The interviewee says: “It is not something which is frequently tested functionally to see if more can be achieved or to support other processes.” It also shows us that the interviewee thinks that there might be more options to use Quo Vadis data in other processes. Based on the maturity model, level two is the right maturity level for Quo Vadis.

Track

For the Track process, the interviewee assigns level three. The documentation is available and good. The process is under control and the data is delivered every month. The interviewee finds it difficult to say if there is improvement happening. The interviewee also says that there is space for developing more efficient processes. Based on the maturity model, it is true that level is the right maturity level for TNR. Based on the maturity model, level three is the right maturity level for Track. However, we must take into consideration that the data which is delivered is not complete (section 3.6).

TROTS

For the TROTS process, the interviewee assigns level three. The process is repeatable and consistent. However, the process is only improved when incidents occur. Based on the maturity model, level three is the right maturity level for TROTS.

CGI (Receiving end of the data)

CGI indicates that the maturity level is between levels two and three. Closer to two than to three. We must take into consideration that this level is not given to the acquisition and delivery of a certain data source. CGI only receives the data and can therefore only tell something about the delivery of all the data. There is no knowledge on the acquisition of data

The interviewee indicates that most of the data sources are always delivered. If not, they just have to send an e-mail and an after delivery will take place. There are not many problems with this. CGI has good

documentation. The interviewee says: “The process is in a phase where it exists, it works, and we do not change it as we are scared that we will break something.” The interviewee says that to achieve level three, there should be more improvement. However, it is difficult to improve as the process is mostly automated.

General

Figure 21 shows that the main reason for interviewees to give a certain maturity level is that the processes have not been improved since their creation. There has not been any, or little, improvement. This has been said by six out of seven interviewees (86%). To reach level five of the maturity model, the model describes that the process should be continuously improved. This is not the case. Another reason for not giving a higher maturity level is that the quality of the documentation is low. This has been said by four out of seven interviewees (57%). Well-described documentation is needed to reach level three. A reason for giving maturity level one was that the process is ad-hoc. One out of seven interviewees said so (14%).

Positive reasons for assigning a process to a certain level were high quality of documentation, the process being repeatable and/or managed. It was also said that the delivery of TNR data is a low-risk process. This should be considered as the maturity level for TNR is one. This is a matter of prioritization and whether the efforts are worth it for improvement.

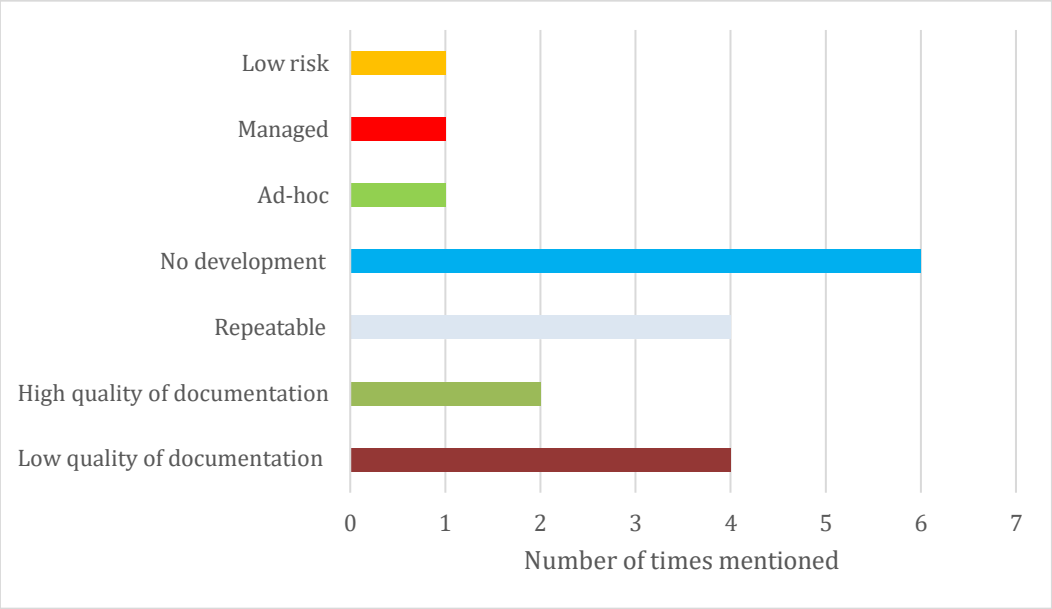


Figure 21: Explanation by interviewees for maturity model

4.5 Statements

Three different categories were established in the theoretical framework of this research. For the category’s information quality and process quality, four statements were presented. For the category user satisfaction, one statement was presented. The number of statements depends on the number of relevant indicators found in the theoretical framework, Chapter 2.3. The statements can be found in Table 11. All the answers from the interviewees to the statements can be found in Appendix F.

Table 11: Statements that were presented to the interviewees

Information quality
1. The delivered data is accurate (accuracy)
2. The delivered data is complete (completeness)
3. The data is delivered on time (timeliness)
4. All the data which is delivered is used by WenS (relevance)
Process quality
5. The quality of the documentation for acquiring and delivering data source x is high (quality of documentation)
6. The quality of the delivered data is high (data quality)
7. The process of acquiring and delivering the data source could be done easier (ease of use)
8. The process runs smoothly and is reliable (reliability)
User satisfaction
9. I am satisfied with the current process (user satisfaction)

For scaling the statements, the Likert scale has been used. To create the Likert scale, symmetry and balance were considered. Symmetry means that the scale exists of equal numbers of positive and negative reactions and is therefore symmetric to the neutral/zero value. Balance means that the value between each value is the same and therefore allowing averaging. As a result, the following five-level format was used:

1. Strongly disagree
2. Disagree
3. Neither agree nor disagree
4. Agree
5. Strongly agree

The interviewee could strongly disagree, disagree, be neutral, agree or strongly agree with a statement. For each of these options, the points dedicated to the choice of the interviewee respectively were one to five points. This excludes the statement: “The process of acquiring and delivering the data source could be done easier”. For this statement, reverse ordering of the Likert scale was needed. It can be argued that the interviewees do not realise that question is intended as opposite to the rest. However, as the non-negative statement was discussed at a calm pace by the interviewer and interviewee, including an explanation by the interviewee for the chosen value, it can be excluded that there were misunderstandings.

In Table 12, the descriptive statistics of the statements can be found. . In this table can be seen that some of the statements were not answered. These statements have been indicated by an ‘x’. The main reason for a statement not being answered was that the interviewee did not have the knowledge to answer the statement. A consequence is that the mean is calculated by fewer numbers and therefore less reliable. Table 12 was used to create Figure 22, a bar chart. To be able to create a good bar chart and draw fair conclusions, the sum of the points was calculated. Then, the sum was divided by the number of statements answered per category. Also, the average for each category between interview Switch 1 and interview Switch 2 was taken.

Table 12: Descriptive statistics of statements

Data source / Statement	Switch (1)	Switch (2)	Track	Quo Vadis	TNR	TROTS	CGI
Information Quality							
1	3	2	4	x	5	5	x
2	3	3	2	5	5	4	4
3	5	5	5	4	3	4	4
4	4	4	4	x	4	3	4
SUM	15	14	15	9*	17	16	12*
MEAN	3.75	3.5	3.75	4.5*	4.25	4	4*
Process Quality							
5	2	3	3	1	1	3	4
6	3	3	2	4	3	5	x
7**	2 (4)	1 (5)	2 (4)	4 (2)	1 (5)	2 (4)	2 (4)
8	3	3	2	4	2	3	4
SUM	10	10	9	13	7	13	10*
MEAN	2.5	2.5	2.25	3.25	1.75	3.25	3,33*
User Satisfaction							
9	2	2	3	4	2	3	4
SUM	2	2	3	4	2	3	4
MEAN	2	2	3	4	2	3	4

*Not all statements have been answered for this category. The mean has been calculated by only using the answered statements.

** For this statement, the Likert scale was inverted. The number between brackets is the original score given by the interviewees.

Figure 22 shows the information quality, process quality and user satisfaction within the current WenS process. The categories have been grouped per data source. The Figure has been set up by using the mean from Table 12. For the Switch data source, the average between Switch 1 and Switch 2 has been taken.

.

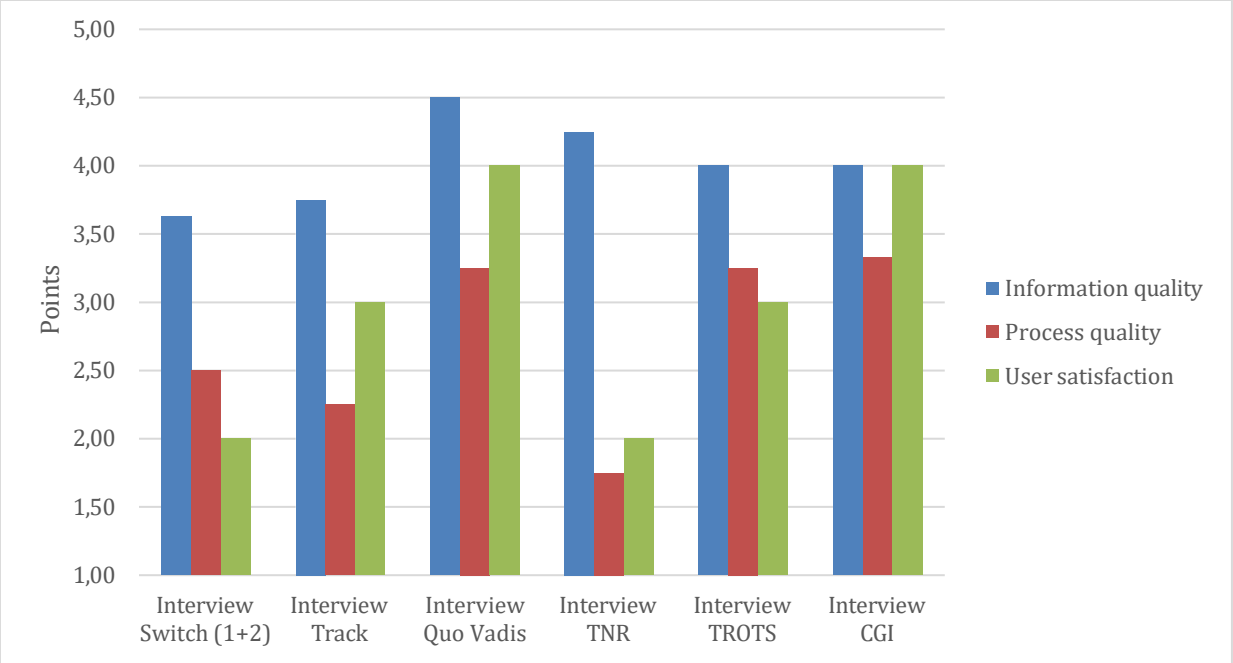


Figure 22: Information quality, process quality and user satisfaction on each data source

According to the statements, the information quality of Switch and Track are the lowest. Only these two data sources score lower than four out of five. Reasoning from the interviewees shows us that the switch data is not accurate and that the track data is not complete.

For process quality, TNR scores 1.75 due to no documentation, no structure and no reliability. According to the interviewee, it can also be done more easily. Track scores 2.25 due to outdated documentation and missing data. It can also be done easier and more smoothly. Switch scores 2.75 due to incoherent documentation and the possibility of higher quality data. The process can also be done more easily using automation. Quo Vadis and TROTS both score 3.75. For Quo Vadis, only the quality of documentation is low. For TROTS, it can be done more easily. CGI scores 3.0 and mentions that it can be done more easily by better communication and resolving technical issues.

User satisfaction divided the different data sources and CGI into three groups. Group 1, Switch and TNR are not happy with the process. A score of two out of five is given on the statement whether they are satisfied with the process. Interviewees from the Switch argued that it should be more accurate, more clearly and easier. The interviewee from TNR argued that it works but can be done more easily. Group 2 exists of Track and TROTS. Interviewees gave a three out of five score on satisfaction. The interviewee from Track argued that he is not satisfied due to the amount of time it takes to complete the process. The interviewee from TROTS argues that it depends. Sometimes, the process works well for a certain amount of time, while at other times, everything fails. Then we have group 3, who are satisfied with the process. Group 3 gives a score of four out of five based on satisfaction. CGI indicates that the user problem should be solved and Quo Vadis states that there should be more thoughts on what they exactly want and whether the data can be improved, expanded or used for other projects before full satisfaction can be reached.

5. Proposed solution

To optimise the WenS process, a proposed solution for each data source was formulated. To achieve this, the main challenges must be tackled. These challenges have been determined per data source in Chapter 4.

The proposed solution tackles these challenges by addressing solutions and explaining how these solutions can be realised for the WenS process per data source. The solutions are formulated by using the semi-structured interview with the interviewees, including their opinions on information quality, process quality, satisfaction and maturity level. First, some general solutions are proposed for the WenS process. Then, each data source is discussed individually.

5.1 General

Analysing all interviews together, some general themes can be found. These themes are communication, automation and information need. These categories can not be linked to a specific data source but are important for the overall WenS process. Communication has been mentioned several times by the interviewees and is important since data needs to be shared within the WenS process. Coordination is good to coordinate how data is being acquired and delivered. Security is also an important aspect, as the WenS process is dealing with sensitive data. Within this section, there is more elaboration on these aspects.

5.1.1 Communication

Two teams, CGI and Quo Vadis, have indicated that there is a need for improved communication. CGI wishes to show Material Impact their process for better understanding between the two teams whereas Quo Vadis wishes to communicate more with CGI about what is being done with the data, how important it is and certain wishes that will arise in the future. A meeting once every amount of time, which should be agreed on by the teams themselves, could help to achieve these wishes. Materieel Impact, as a product owner, should play a key role in achieving this. One meeting that definitely should take place is between CGI and Materieel Impact. Due to limited knowledge of the process of the other department, miscommunications have taken place in the past and will keep taking place in the future. These two departments should take each other through their process to avoid miscommunications.

5.1.2 Coordination and overview

In addition to the paragraph about communication above, it would be beneficial to make sure there is good coordination between the different teams of each data source. The WenS output relies on all the data sources. Without one of the data sources, CGI can not create the output. Therefore, good coordination and a clear overview are crucial.

This can be established by assigning a WenS process owner and a point of contact for each data source. It would be good to document this. To succeed, a process owner must be a permanent role with real responsibility for and authority over designing the process, measuring its performance, and training the frontline workers who perform it (Hammer & Stanton, 1999). In addition, the continuous review and update of performance measurement system should also be constituted as a process with a defined process owner, who is in charge of development of the required skills (Kuwaiti, 2004). In such a way process owners are in charge of assuring the dynamic improvement of the capabilities of business processes.

The process owner should be the one keeping track of the development of the WenS process. A relatively simple method to do this is to make the process owner responsible for planning meetings within the WenS process. By doing so, the process owner gets to know new wishes, new challenges and new thoughts within the WenS process. These insights will contribute to the continuous improvement of the WenS process. Another party that could be considered for those meetings are the users of the output of WenS. In the current digital age, the wishes of the users can change every moment. The wishes of the users determine which source data is needed. Therefore, it might also be an idea to set meetings with the users of the output of WenS.

The process owner can use the BPMN figures including the written description from Chapter 3 as guidance for understanding all the different data sources. The results of the interviews are also useful. The information about the interviewees can be used to determine who the points of contact might be. These interviewees have been interviewed to acquire knowledge about the data source. It is unsure whether these interviewees are the best points of contact. This should be determined by the process owner. The given maturity level might also be useful for understanding each data source. It helps the process owner to determine the priority level for each data source. ‘’Interpreting and coding the respondents’ answers’’ can already be used to find out the thoughts of the interviewee before going into the meeting. Figure 22, about information quality, process quality and user satisfaction, can be used to see how each data source scores. This helps the process owner to see where the most improvement is possible.

The process owner does not have to work on the WenS-process full time. The exact amount of time spent on the process depends on the number of meetings per year that the process owner wants to schedule. Assuming that the meetings will take place once a half year, it is expected that the process owner only needs a couple of working days per year. Therefore, the costs are kept low.

5.1.3 Security

WenS uses a File Transfer Protocol (FTP) server for the exchange of files between ProRail and CGI. One of the disadvantages of FTP is that the connections are not encrypted. As a result, the delivery of data sources is less secured and therefore vulnerable. An alternative is SFTP, also named FTP over SSH. Using SFTP secures the transfer of data by adding a layer of security – SSH encryption. SSH (secure shell) encryption makes sure that two computers can establish a secure channel. This is done before the actual transfer occurs. SFTP also uses a host key, which should be validated before the transfer. Once the connection is established, all data is securely transferred via the encrypted channel. ProRail and CGI should discuss this together if this is needed.

5.2 Infra data

In this section, the data sources switch and track are grouped. The infra data which is delivered for WenS consists of the same structure, but different data. The exact location where the data has to be acquired needs to be determined by Materieel Impact. As a result, the desired new process for track and switch should be the same and can therefore be visualised in one business process model (Figure 23).

It is advised to first determine all the actual switches and tracks. This can be done by multiple options. It should be investigated whether using the WenS output of last month or using another source system is the best option. For the characteristics of the tracks and switches, it is strongly discouraged to use the output of WenS from last month. Doing this results in not up to data characteristics for tracks and switches. Then, the file should be delivered to the S(FTP) server, as discussed in section 5.1.3. Elaboration for these two data sources can be found below the figure.

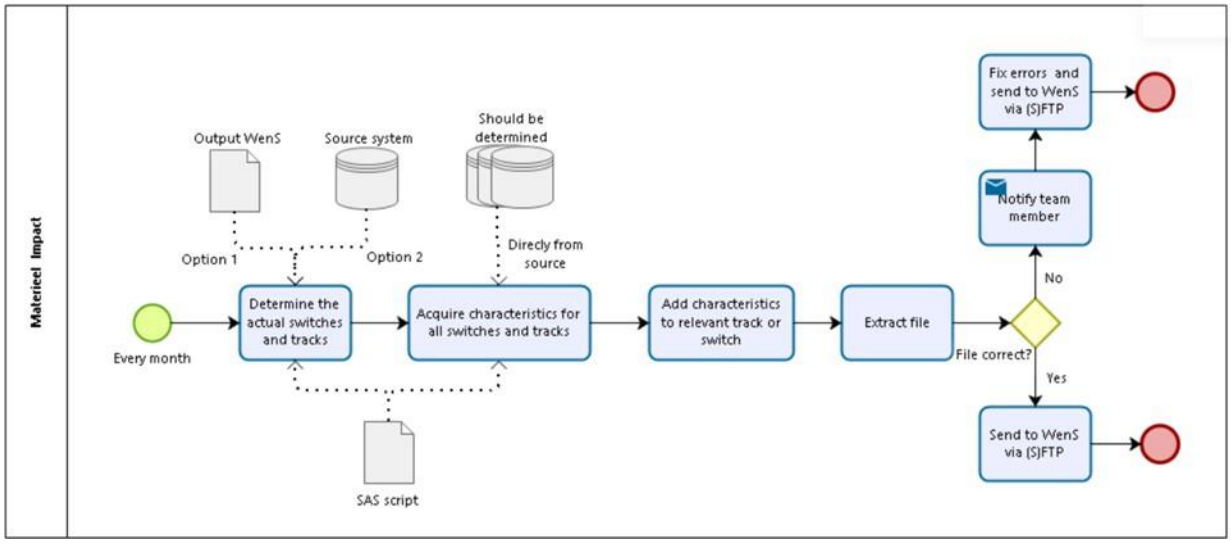


Figure 23: New process for switch and track

5.2.1 Switch

Based on the statements, the main problem for creating the switch data is the process quality. At the moment, new switches are added and removed manually in the Access database. There already exists an idea by interviewees to use a SAS script for automating this process. It would be best to use this script to directly extract the data from the source system. This would mean that every month, a new up-to-date data source is created instead of editing the old data source. Doing this would also improve the data quality, as all the characteristics have not been checked since the beginning.

5.2.2 Track

The track data is not complete yet. Characteristics of some tracks are missing. This causes a decrease in the quality of the WenS output. At the moment, the track file from the month before is used to complete the characteristics of a track. This means that the characteristics of new tracks are missing, and the characteristics of existing tracks are not checked for changes. It would be better to retrieve the

characteristics directly from databases within ProRail. Materieel Impact should find out where the needed data is located. This might be some work, but in the end, the data will be complete and kept up to date automatically.

5.3 TNR

The TNR data is not sent to WenS every month. TNR data changes every month which means that outdated data is used for WenS if not send every month. The process of acquiring and delivering the TNR takes less than fifteen minutes. Although outdated TNR data has a low risk on the quality of the output of WenS, it is fairly easy and not time-consuming to schedule a timeslot every month to acquire and deliver TNR data. The best time to do this would be as close as possible to the deadline of delivering all data sources. This would mean that the best up-to-date data is delivered to WenS. Doing this is the first step for increasing the maturity level. This would already bring the maturity level to level two, as this change would mean that the process is not ad-hoc anymore.

As can be seen in chapter 4.2, there is a high circulation of team members within the WenS process. Together with the knowledge that there does not exist any documentation on this process, creating documentation about the process is valuable for the future. According to Bae, documentation helps to preserve key knowledge. The documentation should be updated when changes occur. According to Bae (1993), standardization is an important benefit of process documentation. “Standard operating procedures will help to achieve consistency in operations. In addition, they provide some administrative advantages. These include reducing conflict among current employees and training new hires regarding how a job should be performed (Ungan, 2006).”

In the end, it would be best to have a completely automated process for TNR. This would mean that no manual actions are required for the acquisition and delivery of TNR data. A set-up of how this process should look can be found in Figure 24. By using BPMN, it can be seen that this process should start once every month. Then, the TNR data should be automatically acquired. This must be done via a script that acquires the TNR data from the TNR application. The next step is to automatically put the Excel sheet on the (S)FTP server. SFTP or FTP depends on the decision of ProRail on what kind of network protocol to use. To make sure that these steps are completely executed, a validation check should take place. If not executed right, the team should be automatically notified so that errors can be fixed.

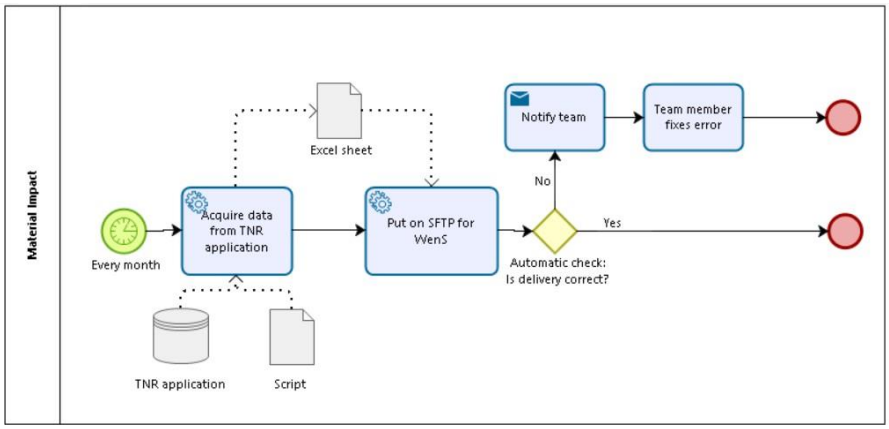


Figure 24: New process for TNR

5.4 Quo Vadis

The Quo Vadis data is delivered to WenS daily, while WenS only runs once a month. Therefore, it might be better to deliver the Quo Vadis data once a month. The data should be delivered at the last moment possible so that WenS has all the available data from last month. A problem that might occur with monthly delivery is that the amount of data is too big. If this is the case, data can also be delivered in smaller time units, e.g. every week.

Delivering the data every month also makes the process easier. Now, after-deliveries have to take place because data from a certain day is spread into multiple days. As an example, the data (enriched match results) from day one is spread to days one, two and three. As a result, most deliveries to WenS are not complete yet. This would be solved by doing a monthly delivery. This would make sure that after-deliveries do not have to take place. In Figure 25 can be seen what part of the process is not needed when after-delivery does not take place. This part has been outlined in red.

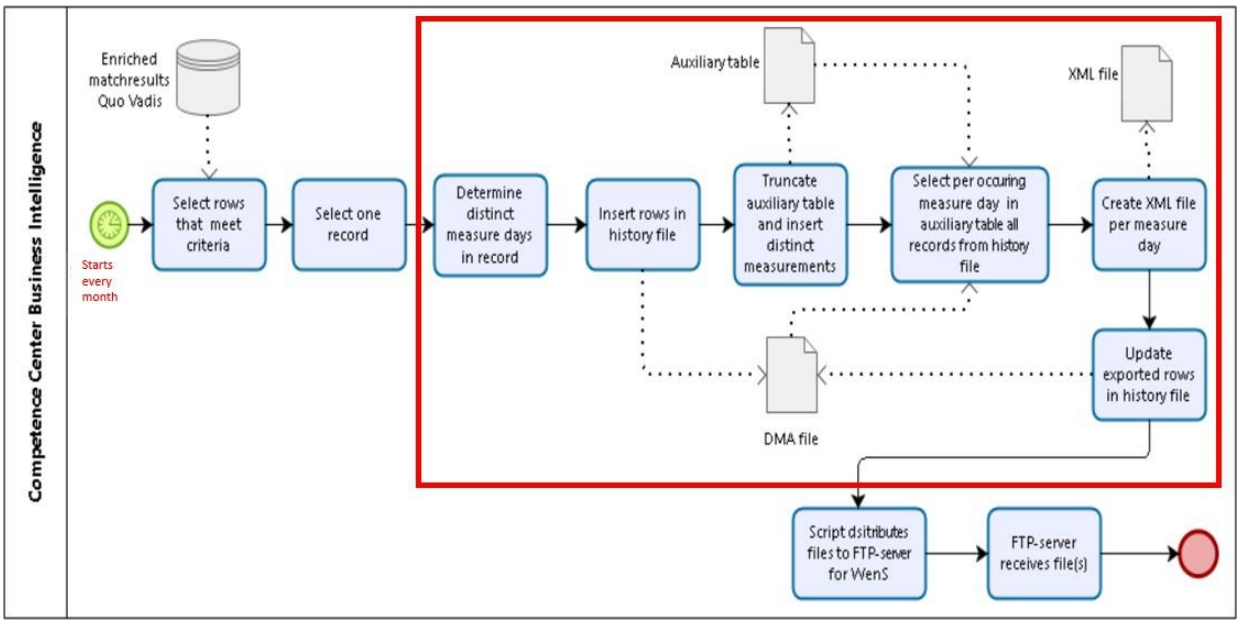


Figure 25: New process for Quo Vadis

5.5 TROTS

To acquire and deliver the TROTS archives, many activities in a short amount of time are needed. Reducing the number of activities will decrease the chance of failing to acquire and deliver the data source. Reducing the activities will also make sure that there are fewer parties involved, and therefore reducing those responsible. This can be achieved by removing IPSTEP server and directly sending the TROTS archives to the Windows or Linux server. More on Linux in the next paragraph.

The windows server was indicated as troublesome in the interviews. These troubles include bugs in the windows server. The scheduler for new archives starts up multiple times. The scheduler has to be restarted multiple times to prevent this from happening. A suggested solution is Linux. Linux has a

reputation for being fast and smooth while Windows is known to become slower over time. Linux is also known to be very stable and not prone to crashes. Another advantage is software cost. Linux can be downloaded for free, making it cheaper than other options. Upgrades are also for free, which means that there are no costs to keep up with the latest version of Linux, unlike Windows. (Dedrick, J., & West, J., 2003)

5.6 CGI

An option that CGI could take into consideration is to try to offer their customer(s), in this case ProRail, an analysis of the input for WenS before combining it. Cleaning the data, detecting and correcting (or removing) corrupt or inaccurate values from all the input might be a good service to offer. This could be done by using the input from the months before and should be done in good consultation with ProRail.

6. Conclusion, recommendations, and discussion

This research was executed to optimise the WenS process at ProRail. Focusing on the whole WenS process was not realistic within the given timeframe. Identification of the core problem showed us that the focus should be on the acquisition and delivery of source data for WenS. As a result, the following research question was formulated: *'How can the acquisition and delivery of source data for WenS be improved?*' The research question is answered in this chapter by briefly answering all of the sub-questions.

6.1 Conclusion

The first sub-question ''What kind of business process modelling technique is suitable for the improvement of the acquisition and delivery of source data?'' was answered in Chapter 2.1. This question needed to be answered so that within this research, the best available process modelling technique was used. To gain this knowledge, a systematic literature review was executed. This systematic literature review can be found in Appendix D. First, the most popular techniques were listed including an explanation and example for each of them. Then, criteria for choosing a business process modelling technique were defined. This was done by combining the knowledge from different researches. Finally, these criteria were used to determine the chosen business process modelling technique, Business Process Model and Notation (BPMN).

The second sub-question "Which type of maturity model can be used for business process modelling?" was answered in Chapter 2.2. After researching different maturity models, The Capability Maturity Model Integration (CMMI) was chosen. This is a maturity model that helps organisations with streamlining process improvement and encouraging efficient behaviours for software, product, and service development. This maturity model has been chosen as it describes both software and system engineering.

The third sub-question "What does the As-Is model of acquiring and delivering the source data look like?" was answered in Chapter 3 and Chapter 4. In Chapter 3, the processes of acquiring and delivering all six data sources were described by using both BPMN and text. This has been done by conducting interviews (round 1) with team members of each of the data sources. This round also indicated that for the next round, one of the six data would not be taken into account. The reason for this was that the interviewee from configuration data could not name any challenges. Also, this data source is outsourced to RIGD-Loxia. For a better understanding of one or more of the data sources, please refer back to Chapter 3. Chapter 4 goes deeper into the current situation by conducting another round of interviews (round 2, set-up for this interview can be found in Appendix E). Within this chapter, answers from the interviewees were coded and interpreted. For an overview of this, please refer back to the tables in Chapter 4.3. In round 2, the maturity level per data source was also defined. Please refer back to Figure 20 in Chapter 4.4 for a summary of the given maturity level per data source. Finally, the information quality, process quality and user satisfaction were defined by using statements. For a summary of this, please refer back to Figure 22 in Chapter 4.5.

The fourth sub-question "What alterations are needed to improve the acquisition and delivery of the chosen source data for WenS?" was answered in Chapter 5. The challenges named in Chapter 3 and Chapter 4 are tackled by addressing solutions and explaining how they can be realised per data source. Below, a brief summary is given:

Switch

- Use a script to automatically acquire and deliver the data each month
- Create the source data from scratch. Use the output of WenS from last month or another source system (needs to be determined) to determine the actual switches. Locate the source of the characteristics for switches.

Track

- Complete the characteristics for each track as soon as possible. Locate the source for all the characteristics. This will increase the quality of the output of WenS.

TNR

- Make sure to schedule a timeslot every month to acquire and deliver TNR data. The best time to do this would be as close as possible to the deadline of delivering all data.

Quo Vadis

- Change the delivery from a daily delivery to a weekly/monthly delivery.

TROTS

- Look into the options for removing IPSTEP. This would make the chain shorter, and therefore reducing the risks.
- Analyse whether to use Linux or Windows in the process of acquiring and deliver TROTS archives.

CGI

- Offer analysis of input data to increase the quality of the output of WenS.

Furthermore, ProRail should assign a process owner to the complete WenS process and a point of contact for each of the data sources. This should improve the communication and coordination within the WenS process. Another way of improving communication is to schedule a regular time to meet. One meeting that definitely should take place is between CGI and Materieel Impact, the product owner of this research. Both departments should take the other department through their process to increase the understanding of the other department. The number of meetings that should take place in a year should be determined by the process owner.

Another point to look into is whether an FTP server is still sufficient in this digital age. WenS is dealing with sensitive data and FTP is not the most secure way of transferring data. An alternative could be SFTP. Implementing these changes should lead to an improved acquisition and delivery of source data for WenS and therefore answering the main research question.

The fifth and last sub-question "How to evaluate the To-Be business process model?" will be answered in this chapter, section 6.3. The evaluation showed that this research has value for ProRail and that the solutions are perceived as useful. Therefore, the proposed solution can be used to answer the main research question: *'How can the acquisition and delivery of source data for WenS be improved?*'.

6.2 Recommendations

ProRail is recommended to take a look into the proposed solutions in Chapter 5. This chapter is an answer to the main research question: *'How can the acquisition and delivery of source data for WenS be improved?*'. The acquisition and delivery of source data for WenS will be improved if the proposed solutions are implemented.

ProRail is also recommended to stimulate information sharing between all the different teams. This research has shown that team members from a certain department have limited knowledge of the processes that take place at another department. This leads to miscommunications. ProRail can use this research, namely the BPMN models from Chapter 3, to increase the knowledge of team members on other departments. CGI and Materieel Impact, the product owner of WenS, should increase information sharing to prevent miscommunication. It is advised that Materieel Impact increases their knowledge on the process of combining all data sources which take place at CGI.

ProRail should use both the maturity model and the performance measurements after the solutions have been implemented. The maturity model and performance measurements can be used to measure the actual improvement of the processes for acquiring and delivering each data source after the solutions have been implemented.

A bigger idea to investigate in the future are the formulae in the algorithm for combing all the data sources in WenS. The algorithm, and therefore the formulae, were outside the scope of acquiring and delivering a data source but have been mentioned by different persons in this research. The formulae may be outdated and therefore affecting the quality of the WenS process.

6.3 Discussion

This section discusses the evaluation, the assessment of validity and reliability, the research's limitations, future work and contributions to practice.

6.3.1 Evaluation

An evaluation of this research was conducted with the manager of Materieel Impact, the problem owner of this research. For the evaluation, the manager was taken through this research. Then, questions were asked. These questions were prepared beforehand but deviated based on the answers given.

The evaluation showed that the business process models made in this research are a good representation of reality. The manager indicated that these business process models mapped the complete process which helps to find improvements. Also, the defined maturity level and statements were marked as useful. The manager said: "If we start improving the processes, we can measure if we score higher on these aspects after the solutions are implemented". The solutions were also perceived as useful.

The manager also indicates that it is important to keep monitoring the WenS process. This research should not be a one-off project. In the evaluation is said that it should not be the case that another

independent individual research takes place in a couple of years. This research should be fundamental for further improvement. Continuous improvement after this research should be the reality.

6.3.2 Assessment and validity and reliability

According to Le Comple and Goetz (1982), validity is concerned with the accuracy and truthfulness of scientific findings. The research should measure what is intended to be measured. The findings should be a true representation of the actual situation. According to Cooper and Schindler (2014), there are two major varieties for validity. Therefore, these types of validity are considered most important.

The first type of validity is internal validity. Within this research, data collection is done by semi-structured interviews. When doing qualitative data collection, bias may occur. Respondents are asked to give their insight. To decrease the bias and therefore increasing the internal validity in this research, several actions were taken. First, respondents were familiarized with the research. By making sure the respondents knew the nature of the research; the reason behind the research, what the research was about, data collection methods and what would be done with the data, the validity increased.

The second type of validity is external validity. External validity is the extent to which you can generalize the findings of a study to other situations, people, settings and measures. This research was conducted specifically for the WenS process. This makes it hard for other researchers to use this research as a generalization for other situations. Other situations will not be identical to this research. Due to this, external validity is limited.

Drost (2011) defines reliability as "the extent to which measurements are repeatable when different people perform the measurement on different occasions, under different conditions, supposedly with alternative instruments which measure the construct or skill" Within this research, reliability may have been a key concern when conducting interviews. The truth of responses for qualitative research may have varied due to bias. Respondents may have wanted to worsen or better a certain process for their favour. To mitigate this risk, questions were kept simple and words that might introduce bias were avoided.

To increase the reliability of this complete research, the research should be repeatable by other researchers. For other researchers to be able to do repeat this research, every step taken in the process was documented. Although this research is difficult to repeat as it is not generalisable, documenting every step still helps to see if this research was reliable. More on this in the research's limitations.

6.3.3 Research's limitations

An important limitation for the execution is time. This research's time span was ten weeks. Due to this limitation, the in-depth investigation of each of the data sources may vary. Another limitation of this research is that it is done in the time of a pandemic due to Covid-19. For this reason, the phase of data collection was troublesome. In a normal situation, the research would have taken place at the headquarters of ProRail. Reaching out to different departments for data collection and/or questions in a normal situation is more efficient. However, this limitation will be dealt with professionally by the means of online alternatives.

Due to confidentiality issues, not all interviews could be recorded. As a result, notes had to be taken. This made it more difficult to create an accurate BPMN model for the current situation, and coding and interpreting the interviewee's answers.

Another limitation of this research is that the WenS process is unique. Therefore, the research is specifically executed for ProRail and the WenS process. As a result, the research is not easily generalisable and repeatable.

None of the interviewees has the original knowledge of how the WenS process was set up. The WenS process was set up by an ex-team-member long time ago, and therefore some knowledge has been lost. This means that the team members who have been interviewed do not have complete knowledge of how the WenS process has been set up. Due to this, the interview results might not completely reflect the real situation.

For all data sources, except one, only one interviewee was interviewed. Due to this, conclusions were made based on one person's idea only. This might decrease the power of the study, with less accurate interview results as a consequence.

6.3.4 Future work

A research topic that might be of interest for the WenS is the formulae used within this process. The background on the formulae is not clear. These formulae have been composed a long time ago. It should be checked whether these formulas are still up to date. If not, the output of WenS is less accurate.

Another research topic that would be of interest for the WenS process is to investigate if it is possible to deliver all data sources identically. This could be an additional improvement to the WenS process and perhaps lead to an easier to use and understand process for every party involved.

This research lacks a cost-benefit analysis. As there is a limited amount of time available to execute this research, this research design is only focused on finding solutions for the existing challenges. The cost-benefit analysis would make it easy for the management team to decide on which recommendations to implement. Also, the cost-benefit analysis allows ProRail to measure the benefits of a recommendation. It involves measuring whether the costs of a recommendation save enough time to acquire and deliver the data, or whether it significantly improves the data quality. The cost-benefit analysis should be considered as future research.

6.3.5 Contributions to practise

Business Process Modelling and Notation (BPMN) was used in this research. However, before concluding that BPMN was the best technique, a theory was established. By using three different papers, Aldin & De Cesare (2009, Aguilar-Savén (2004) and Luo & Tung (1999), a list of five different criteria for choosing a business process modelling technique could be obtained. These criteria can be used by others to determine the best suitable business process modelling technique. The criteria can be found in Table 1.

One of the limitations was that this research is difficult to generalise and repeat. However, one part of this research can be used for other processes within ProRail, or by other companies. By combining the knowledge of Benmoussa et al (2018), DeLone and McLean (2003), and Seddon (1997), a way of measuring the information quality, process quality, and user satisfaction was established. This was done by finding indicators for each of these three categories. The indicators were turned into statements so that a score on each of the categories could be determined. This way of measuring could be used by others to gain knowledge about their information quality, process quality and user satisfaction.

References

Aldin, L. & De Cesare, S. (2009). A Comparative Analysis of Business Process Modelling Techniques. Paper presented at the UKAIS 14th annual conference.

Artino, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical Teacher*.

Bae, H.M. (1993), "Process flow modeling and analysis: a practitioner's perspective", Industrial Engineering, Vol. 25, pp. 54-5.

Benmoussa, Majid, Khoulji & El Yamami. (2018) Impact of System Quality, Information Quality and Service Quality on the efficiency of information system

Boell, S. K., & Cecez-Kecmanovic, D. (2015). What is an information system? C3 - Proceedings of the Annual Hawaii International Conference on System Sciences.

Code of conduct, ProRail, 2017. Retrieved by: https://www.prorail.nl/sites/default/files/gedragscode_prorail.pdf

Code of conduct, University of Twente, 2019. Retrieved by: https://www.utwente.nl/organisatie/over-de-ut/integriteit/gedragscodes/code-of-ethics.pdf

Cooper, D. R., & Schindler, P. S. (2014). Business Research Methods (12th ed.).

Curtis, B., Kellner, M.I., and Over, J., 1992. Process modelling. Communications of the ACM, 35 (9), 75–90.

Dedrick, J., & West, J. (2003, December). Why firms adopt open source platforms: a grounded theory of innovation and standards adoption. In *Proceedings of the workshop on standard making: A critical research frontier for information systems* (pp. 236-257).

DeLone, W. H. and Mclean, E.R. (2003). The DeLone and McLean Model of Information Systems Success: A Ten-Year Update

De La Vara, J. L., Sanchez, J., & Pastor, O. (2008). Business process modelling and purpose analysis for requirements analysis of information systems C3 - Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics).

Drost, E., A. (2011). Validity and reliability in social science research. Education Research and Perspectives, 38 (1), 105-124

Hammer, M., & Stanton, S. (1999). How process enterprises really work. HarvardBusiness Review,77(6), 108–118

Heerkens, H., & van Winden, A. (2017). Solving Managerial Problems Systematically

Krippendorff, K. (2004). Content analysis: An introduction to its methodology: Sage.

Kuwaiti, M. E. (2004). Performance measurement process: Definition and owner-ship.International Journal of Operations & Production Management,24(1), 55–78

Lecompte, M. D., & Goetz, J. P. (1982). Problems of Reliability and Validity in Ethnographic Research. *Review of Educational Research*.

Mendling, J., & Strembeck, M. (2008). Influence factors of understanding business process models C3 - Lecture Notes in Business Information Processing.

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*.

Ungan, M. (2006). Towards a better understanding of process documentation. *The TQM Magazine*

Weske, M. (2012). Business Process Management (2nd ed.). Heidelberg, Germany: Springer.

Shen, H., Wall, B., Zaremba, M., Chen, Y., & Browne, J. (2004). Integration of business modelling methods for enterprise information system analysis and user requirements gathering. *Computers in Industry*.

Appendix

Appendix A: Output of WenS (example)

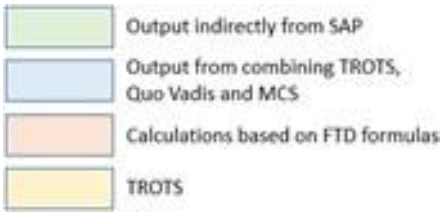
The output data generated by Wens is in the form of spreadsheets. There are 6 different spreadsheets. Three spreadsheets for switch use and three spreadsheets for rail load. Users of the spreadsheets can choose which spreadsheet fits their needs the best. The following three spreadsheets are available for switch use:

A. Switch overview: This spreadsheet shows all the switches including the 25 most important parameters
B. Left and right movement: This spreadsheet consists of spreadsheet A but also shows how many times a switch has been driven left- and right-handed. In total, there are more than 150 parameters.
C. Straight and diverging movement: This spreadsheet consists of spreadsheet A but also shows how many times a switch has been driven straight or diverging. In total, there are more than 100 parameters.

Output	Description
TROTS_Wisselnaam	TROTS switch name
REGIO	ProRail region: NO, RN, RZ of Z
ProcesContractgebied	Proces contract area
Geocode	Geocode
Emplacement	Emplacement indication according to ProRail
Wisselnaam_Kort	Switch name short
Type	Type of switch
Wisselsoort	Kind of switch
Hoek	Angle of switch
Equipmentnummer_Wissel	8 digits equipment number of switch
Equipmentnummer_Steller	8 digits equipment number of switch machine
Bouwdatum	Building date as dd-mm-yyyy
Omlopen	Number of movements of switch
Totaal_Treinen	Total number of trains in all directions
Totaal_Tonnen	Total number of tonnage in all directions
Totaal_Assen	Total number of axes in all directions
Totaal_Treinen_Gemeten	Total number of trains where the tonnage and axes have been measured
Totaal_Treinen_Geschat	Total number of train where the tonnage and axes that have been estimated
Totaal_Treinen_Berekend	Total number of trains where the tonnage and axes have been calculated
Totaal_Treinen_Onbekend	Total number of trains where the tonnage and axes are unknown
Wb	Dimensionless number to calculate Belasting_Baan
Ws	Dimensionless number to calculate Belasting_Seinwezen
Belasting_Baan	Load Baan (A, B, C or D)
Belasting_Seinwezen	Load Seinwezen (I, II, III or IIII)
Aantal_Gemeten_Dagen	Number of days measured
Is_wissel	Other constructions can be measured as switches. If Is_Wissel is "Yes", only then construction information is delivered in the switches spreadsheet.
Periode_Van	Start time of all measurements
Periode_Tot	End time of all measurements

Table 2: Output spreadsheet A switch use

An elaborated description on spreadsheet A can be found in this appendix. Both spreadsheets B and C are too extensive for a detailed elaboration on each of the columns. Spreadsheet A consists of 25 columns. All of these columns are also found in spreadsheets B and C. The rest of the columns of spreadsheet B and C can be found in appendix 2.



Appendix B: Output of WenS

Spreadsheets B and C are more detailed. The number of movements of a switch are divided into different categories depending on the kind of train, how the weight is determined and how the switch has been driven.

Switch use per switch	Kind of train	Realisation	Side of crossing frog	Direction	
Number of trains	Total Passenger trains Freigh trains Other trains	Measured Estimated Calculated Unknown	Along Against	Left Right	Diverging Straight
Tonnage Number of axes	Total Passenger trains Freigh trains Other trains	Measured Estimated Calculated Unknown	Along Against	Left Right	Diverging Straight

Table 1: Output WenS switch use

TROTS, TNR and MCS

TROTS, Quo Vadis, TNR and MCS

As shown in the table, the number of trains are divided into total measured along left, total measured along right, total measured along diverging, total measured along straight, etc. In total, the number of trains gives us 64 columns, tonnage gives us 64 columns and number of axes gives us 64 columns.

The track load is almost the same. Track load also consists of three different spreadsheets with different information. The spreadsheets are similar to the switch use spreadsheets but do not include side and direction. The total output of each of the track load spreadsheets combined can be found in the table below

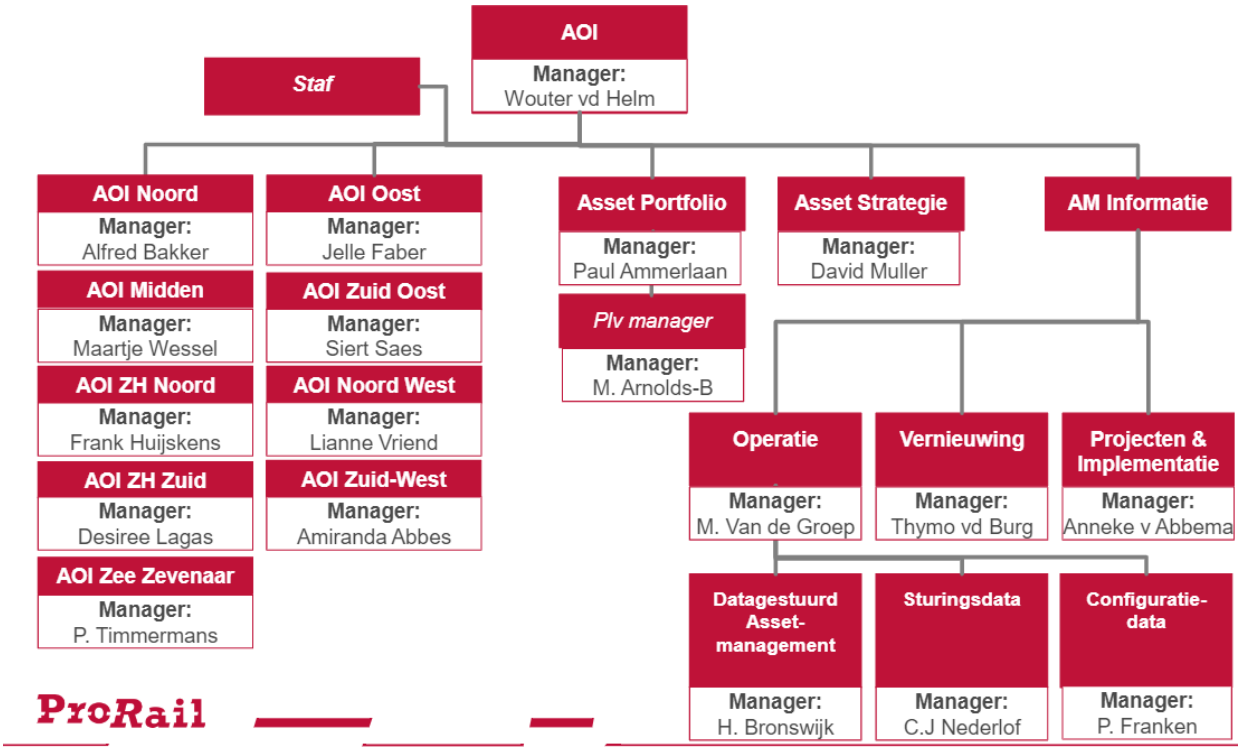
Track load per track	Kind of train	Realisation
Number of trains	Total Passenger trains Freigh trains Other trains	Measured Estimated Calculated Unknown
Tonnage	Total Passenger trains Freigh trains Other trains	
Number of axes	Total Passenger trains Freigh trains Other trains	Measured Estimated Calculated Unknown

TROTS, TNR and MCS

TROTS, Quo Vadis, TNR and MCS

Table 2: Output track load WenS

Appendix C: Organogram of the organisation structure of ProRail



Appendix D: Systematic Literature Review

As part of my problem analysis, I am asked to answer one knowledge question by conducting a *systematic literature review*.

Definition of knowledge problem

During the research, I will have to use business process modelling to get to a solution. Business process modelling can be done in several ways. As I have no knowledge on this, the following knowledge question has been constructed: "What kind of business process modelling technique is suitable for the improvement of the acquisition and delivery of source data?"

Integration of theory

This section provides an integration of all the relevant theory found on business process modelling for improvement of acquiring and delivering source data. The goal is to find out which business process modelling technique to use in this research. To achieve this, knowledge question one must be answered. The knowledge question is as follows: "What kind of business process modelling technique is suitable for the improvement of the acquisition and delivery of source data?" There are several steps needed to answer this question. The first step is to give an overview on the different techniques. Second, the overview should be analysed to find the most suitable technique for this research.

Inclusion and exclusion criteria

Inclusion	**Argumentation**
Literature discussing information systems	Articles relevant to information systems will be helpful.
BPM or business process modelling in abstract or title	Articles without these criteria are not relevant for research
Exclusion	**Argumentation**
Not mentioning "Information system" in abstract	When this keyword is not mentioned in abstract, it may be assumed that article is not relevant
Articles focused on employee behavior in information systems	Not relevant for research question
Article is not freely accessible	No budget for research

Databases used

For a good systematic literature review, at least two different databases should be used. Google scholar will be excluded, as these article are not all peer-reviewed. The two databases that are used are Scopus and Web of Science.

Search terms and used strategy

The following search terms have been used for getting a quick overview of all the literature. These search terms can be combined in different ways to find the most relevant articles:

Information systems

Techniques/methods
Data collection
Design
BPM / business process modelling
Data management
Process mapping
Analysis

After using all these search terms, the following most useful combinations will be used:
"BPM" OR "business process model*" AND "information system*" AND "process mapping"
"BPM" OR *"business process model*"* AND *"information system*"* W/2 *"analysis"*
"improv*" W/2 "information system*" AND bpm OR "business process model*"

Overview articles

Search string	**Date of Search**	**Data range**	**# of articles**
Scopus			
"BPM" OR "business process model*" AND "information system*" AND "process mapping"	26/10/2020	All	13
Improv* W/2 "information system*" AND "bpm" OR "business process model*"	27/10/2020	All	20
"BPM" OR *"business process model*"* AND *"information system*"* W/2 *"analysis"*	27/10/2020	All	32
TOTAL			**65**
Removing duplicates			-1
Using exclusion criteria			-38
Removed after reading abstract			-18
Removed after reading complete article			-5
END TOTAL			3

Conceptual matrix

Article title	Authors	Year	Key findings
Business Process Modelling and Purpose Analysis for Requirement Analysis of Information Systems	Jose Luis de la Vara Juan Sánchez Óscar Pastor	2008	The paper describes an approach to prevent common problems causing ineffective information systems. The paper uses BPMN for business process modelling. Notation consists of Business Process Diagram (BPD)

Integration of business modelling methods for enterprise information system analysis and user requirements gathering	Hui Shen Brian Wall Michal Zaremba Yuliu Chen Jim Browne	2004	Data Flow Diagrams (DFD) is an effective tool as it can describe information flow clearly, from source to destination. Combining DFD with IDEF3 and IDEF0, the As-Is model can be designed after a step-by-step approach.
Influence Factors of understanding business process models	Jan Mendling Mark Stremback	2008	According to this paper, one should put focus on text length, separability and theory when creating a business process model. These factors contribute to the understandability of a process model.
A Comparative Analysis of Business Process Modelling Techniques	Aldin, L. De Cesare, S.	2009	Insight in which techniques can be used for modelling an information system and requirements to choose a technique
BusinessProcessManagement	Weske, M.	2012	Business process modelling techniques

Integration of theory

This section provides an integration of all the relevant theory found on business process modelling for improvement of acquiring and delivering source data. The goal is to find out which business process modelling technique to use in this research. To achieve this, knowledge question one has to be answered. The knowledge question is as follows: “What kind of business process modelling technique is suitable for the improvement of the acquisition and delivery of source data?” There are several steps needed to answer this question. The first step is to give an overview on the different techniques. Second, the overview should be analysed to find the most suitable technique for this research.

Business process modelling techniques

Business process models specify the activities, with their relationships, that are performed within a single organization and therefore specify process orchestrations. Process orchestrations provide a detailed view on the activities of processes and their execution constraints (Weske, 2012). Aldin & De Cesare (2009) give an overview of several popular modelling languages that can be used for visualizing process orchestrations. For each of the techniques, an example can be found below the explanation.

Flow charts

A flow chart is a graphical representation that shows the flow of control throughout a process by providing a step-by-step illustration of what occurs given a specific situation. Flow charts are used predominantly in software engineering, but their simplicity and ease of use have enabled managers and business owners to adopt this technique for organizational purposes as well. (Aldin & De Cesare, 2009).

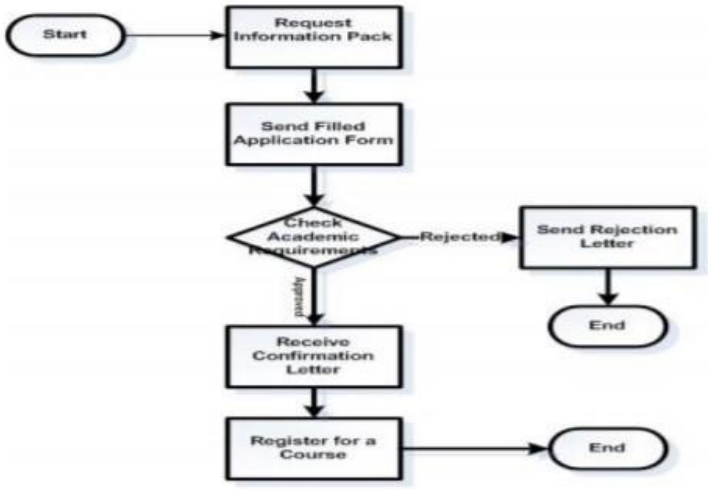


Figure 4: Example of a flow chart (Aldin & De Cesare, 2009)

Petri nets

Petri nets is a well-known technique for specifying business process in a formal and abstract way. Petri nets consist of places, transitions and directed arcs connecting places and transition (Weske 2012). According to Aldin & De Cesare (2009), a petri net is a mathematical/graphical representation that is appropriate for modelling systems with concurrency. Petri nets are used for modelling computer software, hardware, control and business processes.

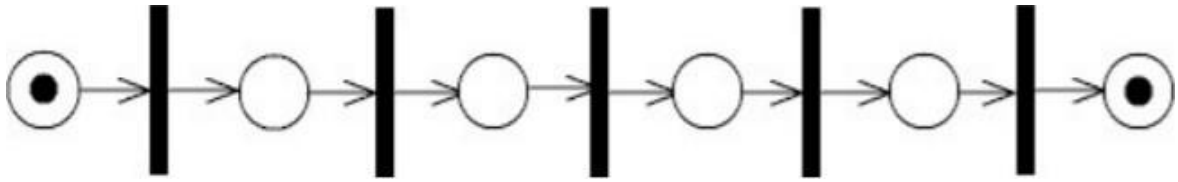


Figure 5: example of a petri net (Aldin & De Cesare, 2009)

Data flow diagram (DFD)

A data flow diagram is graphical representation used to show system functionality. This includes underlying processes and flow of data. It is mainly used for studying systems analysis and design in software engineering. (Aldin & De Cesare, 2009)

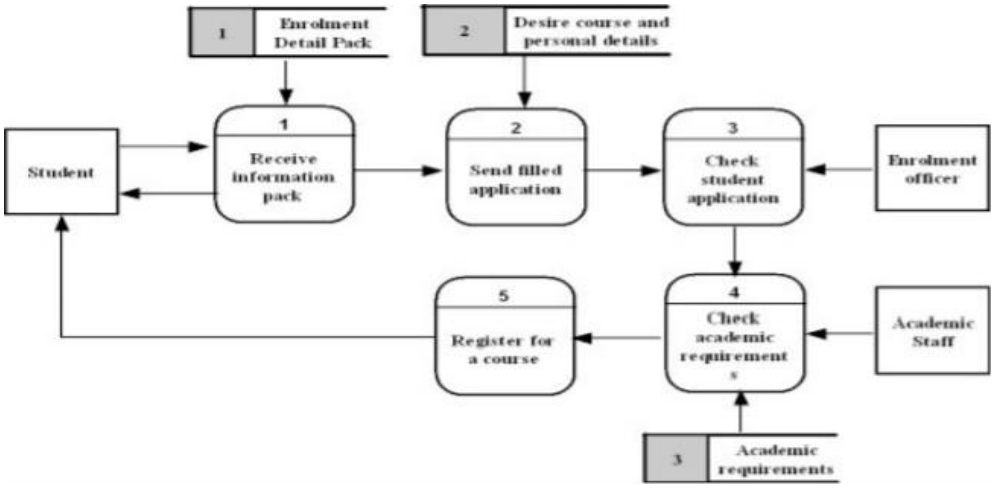


Figure 6: Example of a data flow diagram (Aldin & De Cesare, 2009)

Role activity diagram (RAD)

RADs are a graphical representation of processes in terms of the roles presented within these processes, their component activities and their interactions, together with external events and the logic, which determines what is the sequence of those activities (Aldin & De Cesare, 2009)

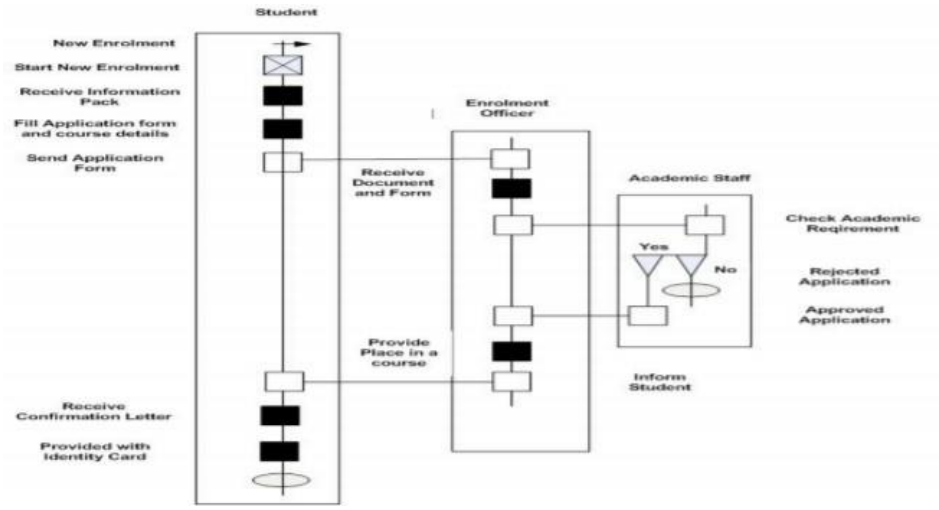


Figure 7: Example of a role activity diagram (Aldin & De Cesare, 2009)

Business Process Modelling Notation (BPMN)

The Business Process Modelling Notation at aims at supporting the complete range of abstraction levels, from a business level to a technical implementation level. The primary goal of this notation is a high understandability for all users. To conclude, BPMN creates a standardized bridge for the gap between the business process design and process implementation (Weske, 2012).

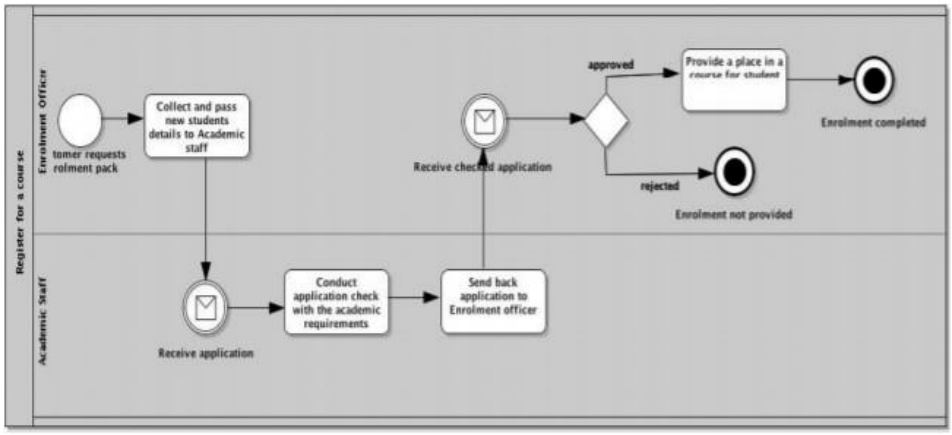


Figure 8: Example of business process modelling notation (Aldin & De Cesare, 2009)

Choosing a business process modelling technique

As the most popular business process models have been described, a decision has to be made on which technique will be used within this research. To do this, the criteria for choosing a business process model have to be defined. According to Aldin & De Cesare (2009), there are five criteria. These criteria include scope, flexibility, easy of use, understandability and simulation.

For this research, the criteria simulation and easy of use are not taken into consideration. There is no need for the business process model to be capable of dynamically simulating a business process. There is also not a great need for stakeholders being able to apply the technique. As a result, the criteria scope, flexibility and understandability remain.

In this research, it is important for business stakeholders to understand the technique. Understandability is important is because the business stakeholders need to understand the business process models for implementation of the recommended changes. A petri net is an abstract technique and therefore more difficult to understand. As a result, petri nets are excluded. Flexibility is another important criterion. It should be possible to change the model without completely replacing it. After reading documentation and conducting semi-structured interviews, the created model has to be validated by a team member. It is possible that the created business process model is not correct and that changes are necessary. Therefore, the business process model should be easily changeable. The last criterium is scope. The scope can be defined by the extent to which the process modelling elements are represented by constructs of the technique. The elements include process, activity, service and product, role, goal, event and rule.

Based on these criteria, Business Process Modelling Notation (BPMN) has been selected. The BPMN technique is flexible, easy to understand, and all elements are included. Although some other techniques fulfil these criteria as well, BPMN has been chosen because of its flexibility. BPMN is a well-structured technique for modelling the different aspects of processes in an organisation. The source data is provided by different kind of processes with fundamental differences. An example of a difference is that some source data is acquired and delivered automatically while other source data is acquired and delivered manually. Therefore, BPMN is suitable for all of these processes.

Appendix E: Setup for semi-structured interview in round two

Start with repeating what has been done in round one. After, permission is asked to record online meeting. If permission is not given, answers are noted down.

Intro

What is your current role at your organisation? What is your organisation?
How long have you been involved in the WenS process?
What is your current role within the WenS process?
Which data sources for WenS are within your domain?

Questions about bottlenecks:

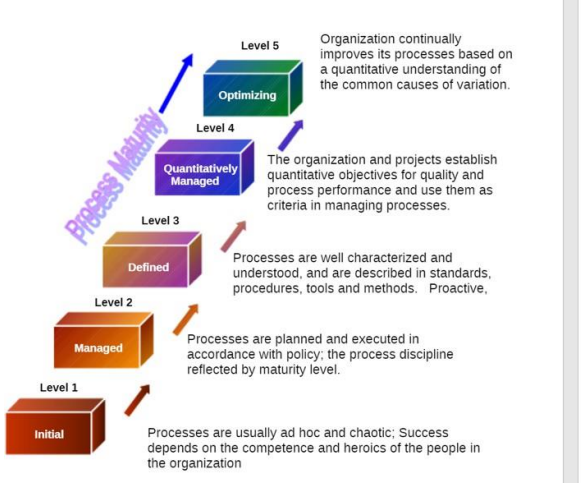
In the last round we discussed the current situation, now I would like to know more about the bottlenecks and how these bottlenecks can be solved. Therefore, I would like to ask you the following questions:

What do you think of the current process, what could be done better?
Why is the process running in this certain way?
What are the bottlenecks with acquiring and delivering data source x?
Where are these bottlenecks?
Zijn er bepaalde problemen met sommige bornbestanden?
How are the bottlenecks influencing the process?
How can this be solved?
Why do these bottlenecks exist?
Do you have any other suggestions for improvement of the acquisition and delivery of all data sources?

Maturity level

Here you can see the different maturities levels for this process. According to you, to which level does acquiring and delivering data source x belong? Can you explain your decision for me?

Levels were presented in both Dutch and English to the interviewees in the following way:



Level 5: Het verkrijgen en leveren van bronbestand x voor WenS wordt continue verbeterd. Deze verbetering is gebaseerd op een kwantitatieve feedback van het proces en het testen van innovatieve ideeën en technologieën.

Level 4: Er zijn kwantitatieve doelen voor de kwaliteit en de prestatie van het verkrijgen en leveren van bronbestand x. Deze doelen worden ook gebruikt als criteria voor het managen van het proces.

Level 3: Het verkrijgen en leveren van bronbestand x is goed beschreven aan de hand van documentatie. Hierdoor is het proces onder controle. Er is sprake van ontwikkeling van processen omdat zij verbeterd kunnen worden d.m.v. eerder behaalde resultaten.

Level 2: Het verkrijgen en leveren van bronbestand x is herhaalbaar en er worden consistente resultaten gehaald. Beslissingen worden genomen op basis van ervaring.

Level 1: Het proces is ad-hoc en soms zelfs chaotisch. Succes voor het verkrijgen en leveren van het bronbestand hangt af van de competentie en daden van medewerkers binnen de afdeling.

Figure x: The maturity model which was presented to the interviewee.

The levels in Dutch can be translated to English as follows:
Level 5: Acquiring and delivering data source x for WenS is continuously improved. This improvement is based on a quantitative feedback of the process and piloting innovative ideas and technologies.
Level 4: There are quantitative goals for the quality and performance of acquiring and delivering data source x. These goals are also used as criteria for managing the process.
Level 3: Acquiring and delivering data source x is well described by means of documentation. This keeps the process under control. Processes are being improved by using previous results.
Level 2: Acquiring and delivering data source x is repeatable and achieves consistent results. Decisions are made based on experience.
Level 1: Acquiring and delivering data source x is ad hoc and sometimes chaotic. Success depends on the competence and heroics of team members within the department.

Statements

Lastly, I have some statements which I would like to get your opinion on. If possible, explanation is very helpful.

Information quality

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
The delivered data is accurate					
The delivered data is complete					
The data is delivered on time					
All the data which is delivered is used by WenS					

System quality

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
The quality of the documentation for acquiring and delivering data source x is high					
The quality of the delivered data is high					
The process of acquiring and delivering the data source could be done easier					
The process runs smoothly and is reliable					

User satisfaction

	Strongly agree	Agree	Neutral	Disagree	Strongly disagree
I am satisfied with the current process					

Thank you for the interview. I have gathered all answers that I need. Do you still have any comments?