UNIVERSITY OF VAASA

SCHOOL OF TECHNOLOGY AND INNOVATIONS

ENERGY TECHNOLOGY

Matias Mäkinen

DATA QUALITY IN SMART MANUFACTURING

Master's thesis for the degree of Master of Science in Technology submitted for inspection

Vaasa 19.12.2019

Supervisor	D.Sc. (Tech) Seppo Niemi
Instructors	M.Sc. Rayko Toshev
	M.Sc. (IEM) Joni Hautala
	M.Sc. Pekka Ritari

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

SYMBOLS AND ABBREVIATIONS

Abberevations

AM	Additive Manufacturing
AR	Augmented Reality
CPS	Cyber Physical System
DAX	Data Analysis Expressions
DQ	Data Quality
DS	SAP Data Services
EDW	External Data Warehouse
ERP	Enterprise Resource Planning
IoT	Internet of Things
IT	Information Technology
MES	Manufacturing Execution System
MRP	Material Requirements Planning
NIST	National Institute of Standards and Technology
PoC	Proof of Concept
RPA	Robotic Process Automation
SM	Smart Manufacturing
SoA	Start of Assembly
S&F	Wärtsilä Search and Forms

LIST OF FIGURES

LIST OF TABLES

TIIVISTELMÄ

Datan laatu on tärkeä osa liiketoimintaa 2000-luvulla. Korkealaatuista dataa tarvitaan yhä enemmän yritysten toimesta, että voidaan tuottaa korkealaatuisia tuotteita ja palveluita. Tämän tutkimuksen tarkoitus on löytää ja määritellä työkaluja ja prosessi datan laadun parantamiseksi kohdeyrityksessä. Nämä työkalut ja monitorointiprosessi luotiin yhteistyössä kohdeyhtiön toimitushallinan organisaation kanssa.

Tämä lopputyö perustuu tuotannossa käytettävien materiaalien datan laadun parantamisen tutkimiseen kohdeyrityksessä. Lopputyössä on esitelty muutama eri työkalu datan monitorointia varten ja näistä yksi on valittu, josta on rakennettu prototyyppi datan laadun monitorointia varten. Työkaluja olivat kolmannen osapuolen ohjelmistot SAP:lta ja IBM:ltä, kohdeyhtiön oma ohjelmisto sekä Microsoftin Power BI raportti. Prototyyppi rakennettiin Microsoftin Power BI -ohjelmistolla ja se konfiguroitiin toimitushallinan tarpeiden mukaiseksi tutkimuksen määrittelemän laajuuden perusteella. Tämä määritelty laajuus koostuu muutamasta avainparametrista materiaalidatassa, joilla on vaikutusta tuotantoon.

Tämän tutkimuksen sekä kirjallisuuskatsauksen perusteella hyvä toimintatapa ja prosessi datan laadun parantamiseksi löydettiin. Tämä prosessi koostuu kuudesta yksinkertaisesta askeleesta joita seuraamalla voidaan saada suuria parannuksia datan laadussa. Nämä askeleet ovat: kerättävien parametrien tunnistaminen, monitorointikohteen tunninstaminen, monitorointiprosessin toteutus, lähtötilanteen arviointi, monitorointitulosten julkaiseminen sekä monitorointitrendien seuranta. Parannus datan laadussa havaittiin myös tutkimuksessa. Ongelmat materiaalidatassa pienenivät huomattavasti, kun verrattiin aikaa sekä ennen että jälkeen monitorointiprosessiin käyttöönoton.

ABSTRACT

Data quality is important aspect for the business in 21st -century. High quality data is needed more and more in companies for producing high quality products and services. The purpose of this research is to find and define tools and process for improving data quality in case company. These tools consist of software and monitoring process established in collaboration with delivery management of the case company.

This thesis is based on studying on improvement of data quality of materials used in production in case company. Few tools for data monitoring are presented and one was chosen for building a prototype for monitoring data quality. These tools were 3rd party software from SAP and IBM, own solution from case company and a Microsoft Power BI report. This prototype was built with Microsoft Power BI and configured for the needs of delivery management according the scope presented in this thesis. This scope consisted of few key parameters of material data that have impact on production.

Based on this study and literature review, a process for improving data quality was found. This process consists of six simple steps, that when followed correctly, can yield great improvements in data quality. These steps were: identifying metrics to collect, identify where to monitor, implementing monitoring process, running a baseline assessment, posting monitoring reports and reviewing monitoring trends. Improvement was also found in data quality in this thesis. Issues, for example missing master data parameters, in material master data quality decreased significantly, when comparing to time before data monitoring to time after monitoring process was implemented.

1 INTRODUCTION

Data quality is important aspect for the business in 21st -century. Previously the requirements for production consisted of labor, materials and tools. In the industry 4.0 and smart manufacturing the era, the 4th requirement is starting to be data. Data is needed more and more to producing high quality products and services and for better understanding the needs of clients and markets and how to cater for them.

1.1 Case company – Wärtsilä

Wärtsilä is 1834 founded company and is the global leader in smart technologies and complete lifecycle solutions for the energy and marine markets. Wärtsilä consist of two business lines: Marine Business and Energy Business. Marine Business focuses on marine industry and enhances the business of its marine and oil and gas industry customers by providing integrated solutions and innovative products. These products include engines and systems for marine use. Energy Business is the leading transition towards 100% renewable energy future. Energy Business offers engine-based power plants, hybrid solar power plants and energy storage and integration solutions. Wärtsilä had over 19 000 employees at the end of 2018 which of 20% were located in Finland, 40% elsewhere in the Europe, 24% in Asia, 11% in Americas and 4% in other countries (Wärtsilä 2019a).

1.2 Motivation

Motivation for this thesis subject came from my current position in case company as a trainee in the smart manufacturing and innovation organization. I have been on my current position for almost two years now and I am working with the manufacturing execution system (MES). My responsibilities with MES have mainly been on the production side as in the role of supporting the key users of the production facilities and lines in problematic cases. These cases mostly consist of bugs and configuration issues in MES. Other

responsibilities consist of daily support of production in different topics and MES development according to production's needs.

Data and data quality have begun to intrigue me after encountering and eventually solving numerous issues in MES relating to data. After root cause analysis and with more knowledge, I have noticed that many of these problems could have been avoided if the data would have been correct and of high quality. Usually the issue is because of human error in one step of the process. In many cases there are issues that are noticed too late. This creates additional urgent work that could have been done days, or even weeks before. My motivation for this thesis was to study and create tool that would help us notice and correct these issues before they become urgent.

1.3   Objectives, methods and research questions

The main objective of this thesis is to find solution or tool to monitor the quality of the data and to find solutions to improve it for the future. Other objectives consist of understanding better the needs and requirements that current, 4th, revolution in industry and manufacturing are creating for companies and organizations. These themes consist of industry 4.0, smart manufacturing and data quality. One objective is to find the data quality issues in systems and finding processes to fix them after monitoring system has found them. These issues consist of missing material and traceability data. Full list of issues will be presented later on the thesis.

These objectives can be expressed as research questions:

 "What is data driven manufacturing in practice?"

"What is the appropriate model to monitor and control the data quality in smart manufacturing in case company?"

"What is the aim for the future of data quality monitoring in case company?"

This study focuses on production side of the view of the case company. Study was done using mixed methods during research process. These methods consist of qualitative research with informal interviews around the company, including production, delivery management, information management and master data organization. Other methods consist of applied research, where the goal is to produce and develop techniques, tools and processes for solving a practical problem. Material used in this thesis consist mainly of online articles regarding the subjects addressed and also other literature.

1.4 Thesis Structure

This thesis consists of different sections: introduction, theory, study and building and testing the reporting tool section and finally conclusions and summary. Introduction consist of presenting the case company, methods, motivation and formation and explaining of research questions Theory section consists of mainly literature review. Study section consists of explaining the current situation and issues and how to understand and solve them. Reporting tool section consists of choosing the reporting tool that will be build, choosing issues and parameters to monitor and building the monitoring solution itself.

After building the monitoring solution, comes part where evaluation is done from the monitoring results. The study, solution building and analysis sections are closely connected to the six steps of building monitoring solutions presented in the white paper "Data quality strategy: a step-by-step approach" (Business Objects 2008). These steps are presented and explained more thoroughly in chapter 2.6 Data monitoring. Last parts of this thesis consist of conclusions and summary in the form of answering the research questions and making the conclusions from results of the study and tools that were built for it. These tools were built by me according to the scope defined in this thesis.

2 THEORY

In this section I will be covering the essential terms and systems that are needed for understanding this thesis and its content thoroughly. Theory section consists of presenting the basic theory of different terms and systems and their functions and capabilities. These terms and systems are enterprise resource planning (ERP), manufacturing execution system (MES), smart manufacturing (SM), data quality (DQ), Industry 4.0 and basics of data monitoring.

2.1 ERP

Enterprise Resource Planning or ERP is suite of software that corporations and organizations use to manage day-to-day business activities. These activities are presented in Figure 1: HR Module, Planning, Inventory Management, Reporting, CRM, Sales and Marketing, Finance and Accounting and Production (Skew Infrotech 2019). ERP systems also usually include enterprise performance management that includes tools for planning, reporting, predicting and budgeting organizations financial results (Oracle 2019).

One of the functions of ERP is to present a holistic view of the business from a single information technology (IT) and information architecture. ERP software can exist in three different forms: generic, pre-configured and installed. Software is generic in its most comprehensive form and targets range of industries and must be configured before usage. Pre-configure solutions can be derived from generic software. These templates are tailored for specific industries and their needs and companies for specific size. Full installation of ERP system comes usually after generic or pre-configured package has been selected and used according to requirements on site (Klaus 2000: 141).

Figure 1.   Enterprise resource planning modules.

2.2   MES

Manufacturing Execution System or MES is an information system that controls, moni-
tors and connects data flows and manufacturing systems in factory. The main goal of
MES is to improve output and provide simple and effective execution of manufacturing
operations and products (Rouse 2017). MES is technology that provides application soft-
ware for the needs of company to rely their manufacturing processes on, online. MES can
benefit all types of manufacturing from process flow production to discrete item produc-
tion. As seen on Figure 2, MES has evolved to fill the gap in communication between
control systems used to run plant floor equipment and manufacturing planning systems
(Schubert 2019). While ERP is focused on planning, MES is focused on executing
(McClellan 1997).
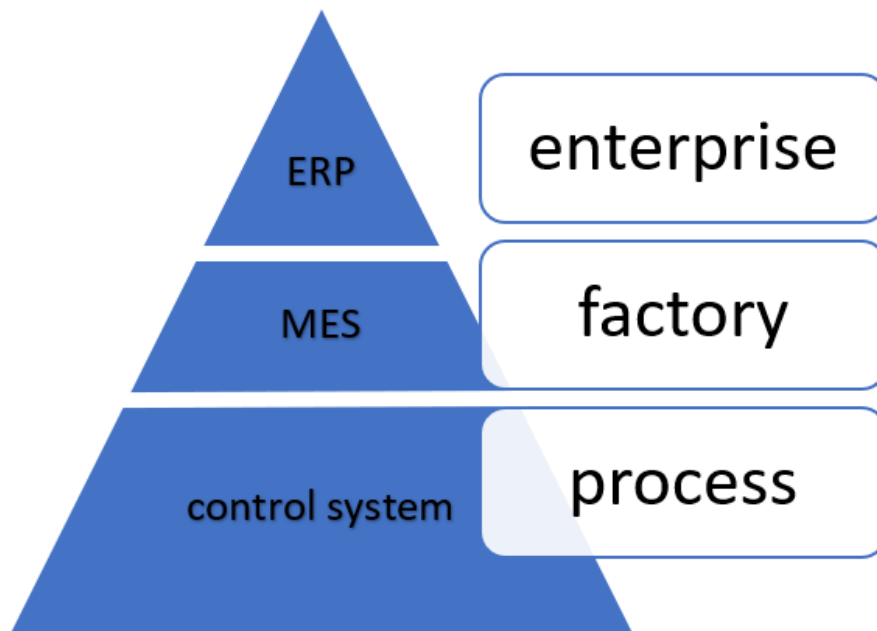
Figure 2.   Vertical integration of ERP and MES systems.

2.3   Smart manufacturing

Smart Manufacturing (SM) is, simply explained, the real-time data and technologies in any form or shape that are needed by people and machines (CMTC 2019). More comprehensive definitions have been done by, for example, National Institute of Standards and Technology (NIST) and by Deloitte. NIST says that Smart Manufacturing systems are fully integrated, collaborative manufacturing systems that respond in real time to meet changing demands and conditions in the factory, in the supply network, and in customer needs (NIST 2019). Deloitte states that smart factory and SM is a leap forward from more traditional automation to a fully connected and flexible system – one that can use a constant stream of data from connected operations and production systems to learn and adapt to new demands (Deloitte 2019). In the global manufacturing scene in the industry 4.0 era, the smart manufacturing is starting to become the focus point in companies and industries. Data provides intelligence for companies and manufacturing big data provides benefit to all aspects

of manufacturing. Manufacturing big data can give huge benefits, for example in the form of accurately predicting requirements and it also can help in identifying errors and bottlenecks. This allows the improvement of manufacturing processes and accelerates product and service innovation (Tao 2019).

Smart manufacturing has few defining technical threads. These threads are synchronization, integrated performance metrics, time and cyber-physical-workforce requirements (Davis 2012). Smart manufacturing can provide great opportunities for companies and organizations in the form of supply chain visibility, automated process management, predictive maintenance and overall quality control (Sikander 2017). In order to take use of their opportunities companies and organizations will be needing a technology strategy to be able to determine the tactics that are needed to be used to ensure success of these opportunities.

Smart manufacturing is about the data collection and going though and using this data to make more informed and better decisions. With better visibility of the operators, shift supervisors and the actual real work in the production line can help make better and more informative decisions. This way smart manufacturing can help in strategic management of technology. These improvements of the visibility can be done for example with display monitors, overhead dashboards or phone messages (Waycott 2016).

2.4   Data quality

Data quality is an important aspect for companies in the 21st century. Poor data quality might cause substantial costs and economic and social impact on firms (Wang 2013). Data quality can be defined in many ways, but basically it means that data is of a good quality and it is suitable for its use case. Quality will always depend on the context of the use case, so it is hard to determine absolute valid quality benchmark for every case (BI Survey 2019). There have been determinations of essential dimensions of data quality. One of these determinations is from Beverly K. Kahn, Diane M. Strong, and Richard Y. Wang as seen on Table 1 (Kahn 2002). Distinguishing data quality dimensions should

help in providing high-quality information. These sixteen dimensions define the different aspects of data quality.

Table 1.   Determinations of essential dimensions of data quality (Kahn 2002).

Dimensions	Definitions
Accessibility	the extent to which data is available, or easily and quickly retrievable
Appropriate Amount of Data	the extent to which the volume of data is appropriate for the task at hand
Believability	the extent to which data is regarded is true and credible
Completeness	the extent to which data is not missing and is of sufficient breadth and depth for the task at hand
Concise Representation	the extent to which data is compactly represented
Consistent Representation	the extent to which data is presented in the same format
Ease of Manipulation	the extent to which data is easy to manipulate and apply to different tasks
Free-of-Error	the extent to which data is correct and reliable
Interpretability	the extent to which data is in appropriate languages, symbols, and units and the definitions are clear
Objectivity	the extent to which data is unbiased, unprejudiced, and impartial
Relevancy	the extent to which data is applicable and helpful for the task at hand
Reputation	the extent to which data is highly regarded in terms of its source at hand
Security	the extent to which access to data is restricted appropriately to maintain its security
Timeliness	the extent to which the data is sufficiently up to date for the task at hand
Understandability	the extent to which data is easily comprehended
Value-added	the extent to which data is beneficial and provides advantages from its use

In addition to these sixteen dimensions, there are six primary dimensions, according to other sources (DAMA UK 2013). These dimensions are Completeness, Uniqueness,

Timeliness, Validity, Accuracy and Consistency, as presented on Figure 3 (Experian 2019).

Figure 3. The six main data quality dimensions.

People can make data-driven decisions only if the data they use is correct. Insufficient data quality means that the data is practically useless and can cause more work and expenses if used incorrectly. On 2005, Fortune 1000 enterprises will lose more money in operational inefficiency due to data quality issues than they will spend on data warehouse and CRM initiatives (Business Objects 2008). There are databases that are not error-free and some of those can have surprisingly low data quality containing large number of errors. To be able to improve the data quality, one needs to understand what that data quality means to the users of the data (Wang 2013).

Data-related problems cost majority of the companies more than five million dollars annually and that one-fifth of companies have estimated that their losses related to data quality are exceeding 20 million dollars per year. 95% of organizations agree that strong information management and data quality are critical for business success (Forbes Insights 2019).

2.5   Industry 4.0

Industry is the part of economy that that produces highly mechanized and automatized materials for the needs of society. Technology has been evolving ever since the beginning of the industrialization and have taken huge steps that have led to a few recognizable eras or industrial revolutions (Lasi 2014). These revolutions can be categorized as 1st, 2nd, 3rd and the current one, 4th, as seen Figure 4 (DataPoint 2018). First revolution consisted of steam powered machines and mechanization in the 1700-century. Second revolution provided the means for mass production by intensive use of electrical energy in the late 1800-century. Third revolution came in the form of widespread digitalization and automation in the late 1900-century. The current revolution, fourth one, consist of cyber physical systems (CPS) and internet of things (IoT). This means continuous data collection from products, machines and production equipment and utilizing this data to improve production and quality of the product.

Industry 1.0			
Mechanization, steam power, weaving loom	Industry 2.0		
	Mass production, assembly line, electrical energy	Industry 3.0	
		Automation, computers and electronics	Industry 4.0
1784			Cyber physical systems, internet of things (IoT), networks
	1870		
		1969	
			Today
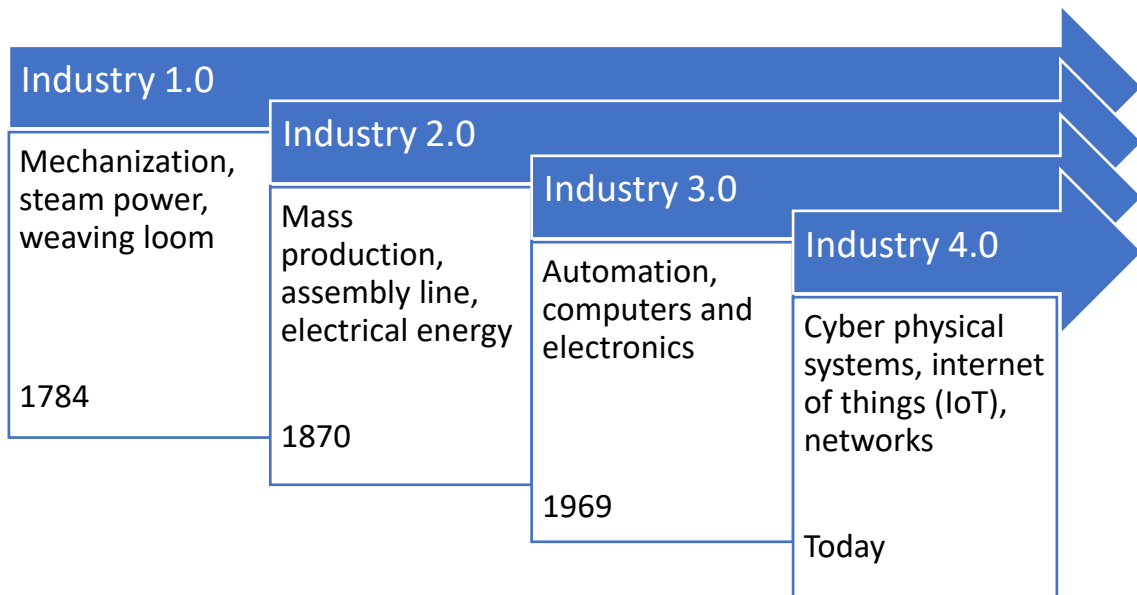
Figure 4.   The development and revolutions of the industry.

Industry 4.0 was originally initiated in Germany and it is close related with the internet of things, enterprise architecture, enterprise integration, information and communications

technology and cyber physical system (Lu 2017). In some instances, industry 4.0 can even mean the same as smart manufacturing and vice versa, but this can be case relative.

There are total of nine technology trends that from the building blocks of Industry 4.0 (see Figure 5). These technologies are autonomous robots, simulation, horizontal and vertical system integration, the industrial IoT, cybersecurity, the cloud, additive manufacturing (AM), augmented reality (AR) and big data and analytics (BCG 2019).
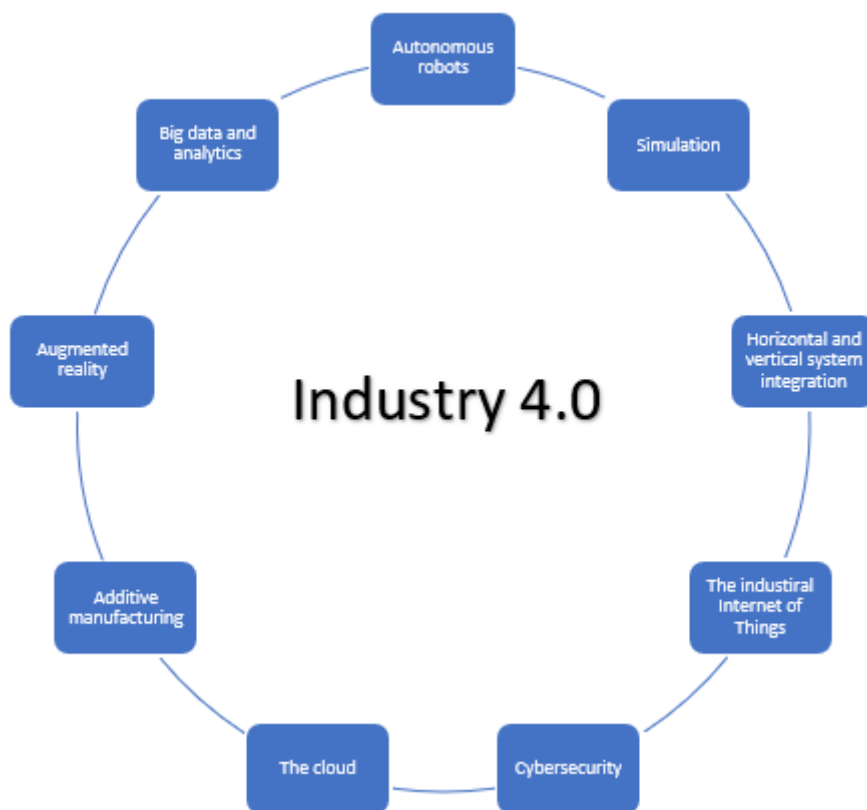
Figure 5.   Nine technologies of the Industry 4.0.

2.6   Data monitoring

Continuous data monitoring is a critical aspect in keeping high data quality. Monitoring means analyzing, measuring and then improving the system or data continuously. This process allows a company or an organization to gauge and measure the data, so it can detect harmful events such as corrupt, wrong or missing data. Continuous monitoring consists of six different areas: identifying measurements and metrics to collect, identifying when and where to monitor, implementing monitoring process, running a baseline assessment, post monitoring reports and scheduling regular data steward team meetings to review monitoring trends (Business Objects 2008). These six areas will provide guideline for creating the monitoring process and they will serve as steps in building a monitoring system.

Continuous monitoring helps users to detect and act on defective data earlier and help them notice which actions are or are not affecting data quality. Monitoring and reporting also quantifies the effectiveness of the actions on data improvement and most importantly it continually reinforces end users' confidence in the usability of the data (Business Objects 2008).

Many systems become unused because of low quality data and stay this way even tough the data is cleaned and enhanced afterwards. The reason is perception and bad first impression. The users of the system have already seen the state of the system in the form of misinformation or slow performance. A few badly timed defects can destroy the users' reliability to the system. To regain the users' trust to the system, a steady stream of monitoring and progress reports needs to be published from monitoring program (Business Objects 2008).

2.6.1   Identifying measurements and metrics to collect

Starting point is defining the goals and the data that support those goals. Measurements need to be focusing on this data to achieve the best results possible. Series of attributes (mentioned in Chapter 2.2) can be measured and these measurements can be aggregated

or rolled up into metrics that combine two or more measurements. A data quality dashboard can be formed from a group of measurements. This dashboard can be used for various of systems in smart manufacturing to aid in data quality issues. For example, invalid address, incorrect email or wrong personnel titles can represent on data field of correct contact data. This monitored data can now be used to make decision for the coming plans, for example postpone planned marketing campaign, because the client data was not in the needed level of quality (Business Objects 2008).

2.6.2 Identifying when and where to monitor

After selecting the goals and data to monitor, we need to know where the monitored data is in the databases. After finding the selected data we need to know when and how frequently the monitoring needs to be done. Even though data monitoring is implemented, organizations should be aware that simple approach to data monitoring most likely won't optimally fit in an organizations goals. Nevertheless, this small step can be viewed as suffice initial implementation. After this first implementation, data monitoring plans should be enhanced to better serve the organizations goals (Business Objects 2008).

2.6.3 Implementing monitoring process

After selecting criteria for the monitoring process, the process itself needs to be implemented. This implementing involves configuring a software solution for testing selected specific data elements against selected specific rules or criteria. After testing the results should be saved for later analysis and correction. Configuring the software requires establishing specific business rules to test against. For example, these rules could be something like setting rules that specific field or column can only have values AAA###, where A means any alphabet and # means any number on its place. Simpler rule could be that the software checks if the value is empty or null and returns error if so. After creating the rules, the user can run them either in regularly scheduled automated monitoring sets or as ad-hoc queries (Business Objects 2008).

2.6.4   Running a baseline assessment

Baseline assessment is the first set of tests conducted to which the future results of the monitoring will be compared. Building the software with rules and configurations is where most of the work is required. Baseline assessment servers as prototype for the monitoring software. Initial rules and requirements in the prototype will need to be changed in the future, because they may and will not evaluate the data criteria effectively. Other tweaks will also be done in the future in the form of generating new rules or adding entirely new data metrics to monitor. Baseline assessment will serve as comparison for future versions of the monitoring program (Business Objects 2008).

2.6.5   Posting monitoring reports

Main purpose of the data monitoring program is to provide information and to cause actions to correct the low-quality data. Common issue in these kinds of programs is poor distribution of the results. Restricting access from user to view the monitoring results are usually very counterproductive. Program that posts monitoring results frequently to an open forum can be very effective in boosting up communication and productivity. The presented monitoring reports should be selected by the viewing group. For example, managers usually like review only higher-level reports whereas people who are more working with the actual data would like to see data from more specific items and reports (Business Objects 2008).

2.6.6   Reviewing monitoring trends

Another typical failing point in data monitoring programs is that there is no follow-through on results and trends. The information has been gained but it isn't acted upon. Therefore, it is important to regularly review monitoring trends and results. Even if tremendous value can be achieved from just knowing that the data is defective and thus avoiding the defects, the greatest value of the monitoring program comes from fixing those defects and even finding their root causes from process and curing them. Without regular reviewing, this cannot be done (Business Objects 2008).

3 STUDY

In the study part of the thesis I will present the issues that will be focused on, requirements for the monitoring software and the software will be selected on this section. Data quality strategy from Business Objects will be used to create continuous data monitoring system that includes process and system for monitoring the data.

3.1 Requirements for monitoring system

Requirements for software come from the needs of company. The first requirement was that the monitoring system resulting this thesis needs to be simple and easy to use. The second requirement that the output of this thesis needs to be system that doesn't need to be maintained but rather it can be upgraded. The third requirement is that the result needs to be easily accessible across the company if needed. The fourth requirement was that the monitored data needs to show the coming eight weeks from current date on rolling basis in the form of start of assembly (SoA) dates.

3.2 Scope of the report

First step of the data monitoring process was to identify measurements and metrics to collect. In this case this means identifying the issues that are causing problems in manufacturing. These issues usually occur because of bad data and these data points are the measurements and metrics that are to be found and corrected.

One of the objectives in this thesis was to find the issues in data that cause errors in manufacturing. The methods used for finding these issues were personal knowledge of the systems and processes and informal interviews of personnel across the organizations. Personal knowledge provided me the issues mainly related to MES. These issues occur mainly to the assembler in production.

The usual case is that assembler is doing his work and using MES when an error pops up that prevents execution. Assembler then calls their supervisor which then contacts MES key user. MES key user are person who are responsible for their own departments MES. Key user tries to solve the issue but if they are unable to, they contact MES main user. MES main user is person who is responsible for MES

These issues can be divided to three categories: Issues in source system (ERP), issues in target system (MES) and misalignment of data between source and target systems. These issues are listed on Table 2.

Table 2.	Found issues in different systems categorized.

Issue	Category
Supervisor validation in ERP	Issue in source system
MRP controller validation in ERP	Issue in source system
Material group validation in ERP	Issue in source system
Set code validation in ERP	Issue in source system
Routing validation in ERP	Issue in source system
ZMES-class validation in ERP	Issue in source system
Serial number profile validation in ERP	Issue in source system
Traceability validation in ERP	Issue in source system
Operation validations between MES and ERP	Issue in alignment between systems
Master data validation between MES and ERP	Issue in alignment between systems
Configuration validation in MES	Issue in target system

After analyzing the issues, six of them were selected to focus on more deeply. The chosen issues to fix were selected based on the workload and effect on production of fixed issue. The selection of these issues was done in collaboration with delivery management organization.

All of these issues were selected from category 'Issue in source system' because this simplifies the scope. Now the focus can be on only one system and thus one database instead of multiple. This helps in building the process and monitoring system because of simpler configuration. Selected issues are seen in Table 3.

Table 3.   Selected issues for monitoring system.

Issue	Category
Supervisor validation in ERP	Issue in source system
MRP profile validation in ERP	Issue in source system
Material group validation in ERP	Issue in source system
Set code validation in ERP	Issue in source system
Serial number profile validation in ERP	Issue in source system
Traceability validation in ERP	Issue in source system

3.3   Choosing monitoring parameters

The second step in data monitoring process is to identify when and where to monitor the data. Where to monitor data can be selected quite easily in this case. Data is in enterprise data warehouse (EDW). Data tables used are located in Oracle database WBR_ODS. From this database I am using few tables. Choosing where to monitor is second phase in this step. As mentioned in requirements of the system the data needs to be monitored eight weeks in future from the current date.

Choosing the monitoring parameters was done according to previously selected issues. For example, for supervisor and MRP profile validation in ERP, it is needed to monitor

those parameters respectively. This means monitoring MRP profile and production supervisor data from materials to find if there are issues in them. On the other hand, traceability validation in ERP requires more than one parameter to be monitored. This means that every issue in defined scope requires different parameters to be monitored constantly.

4 DATA MONITORING PROCESS

After selecting issues and parameters to will be focus on, next step is implementing the data monitoring process. Implementing starts from choosing the tool that will be used in monitoring. There are many 3rd party options available in markets. For this thesis four will be presented and from these one will be chosen. Options are: complete system from SAP, complete system from IBM, using Wärtsilä Search & Forms or self-made Power BI report.

4.1.1 SAP Data Services

SAP offers product for data management and fixing issues with data quality. SAP Data Services (DS) helps to transform, integrate and improve enterprise data and making it available for real-time usage. DS offers on-premise deployment, data quality and integration, intuitive user interfaces and simplified maintenance.

Key benefits of the DS are: Ability of getting complete view of information by accessing data of any size and from any source. Ability to standardize, correct and match data to identify relationships, reduce duplicates and correct quality issues proactively to achieve excellence in information management. Ability to connect data on premises, cloud or Big Data to discover insights.

One of the key capabilities of the SAP Data Services are data quality dashboards (see Figure 6). These dashboards provide for the impact of data quality issues across all downstream systems and applications. Other capabilities are universal data access and high performance. Universal data access provides access to all enterprise data sources from SAP and non-SAP locations. High performance enables parallel processing, bulk data loading and grid computing for high-volume data handling (SAP 2019).
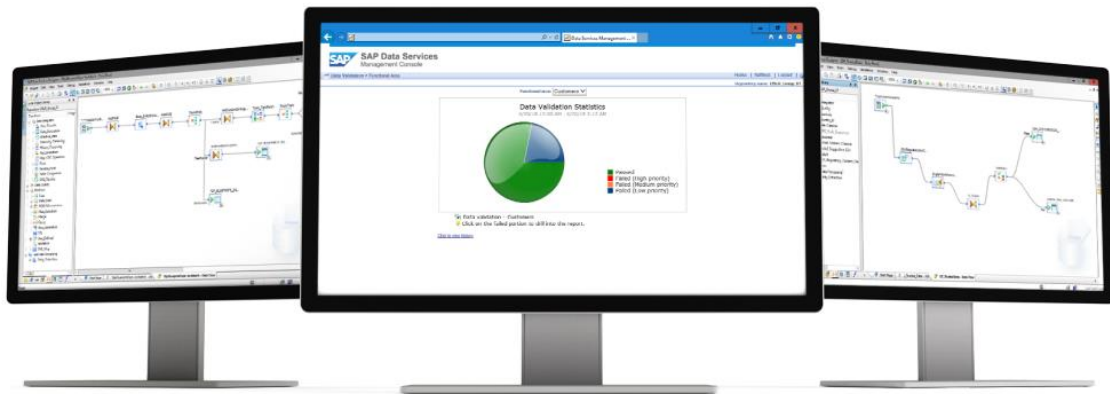
Figure 6.   View of SAP Data Services UI and dashboards (SAP 2019).

4.1.2   IBM InfoSphere

IBM InfoSphere Information Server for Data Quality is software that provides tools for cleaning data and monitoring data quality on ongoing basis. InfoSphere offers end-to-end data quality tools to maintain data lineage; cleanse, standardize and match data; analyse and monitor data quality continuously and tools to help understand data and its relationships with simple and effective user interface (see Figure 7).

One of the key features in InfoSphere are related to data quality and management. InfoSphere uncovers data quality issues and establishes a remediation plan for correcting bad data in systems. IBMs solution continuous data monitoring reduces the proliferation of incorrect and inconsistent data (IBM 2019).
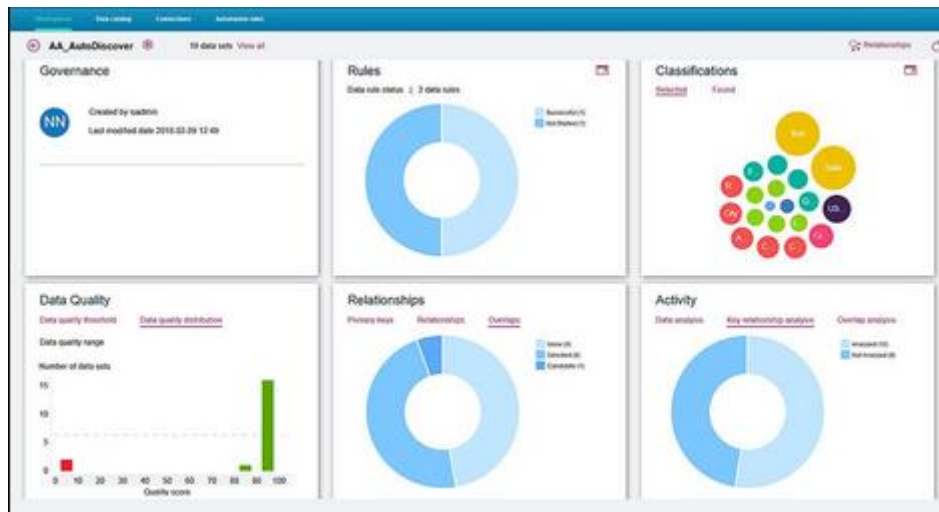
Figure 7.   View of IBM InfoSphere for Data Quality UI (IBM 2019).

4.1.3   Wärtsilä Search & Forms

Wärtsilä Search & Forms (S&F) provides data visibility and advanced automation tools to end users and system developers. Forms is a platform that provides tools for building forms used in creating, modifying or deleting data. Search is an application that acts as a jumping point to modify individual data objects and aims to provide generic search functionality into various internal and external datasets.

The applications that make up Search & Forms are built using the microservices approach. So what users see is a collection of microservices that make up the applications called Search & Forms. They share many of the underlying microservices, and these microservices are also shared with other system developers and services.

Using S&F would give us the benefit of owning the product. S&F is quite new product, but it has wide usage spreading globally and good support internally. Choosing an internal solution would give us quite free hands to select our approach to data monitoring and it would allow more customization for the data monitoring program.

S&F can fetch data from multiple different data sources, as seen on Picture 6. In this case vessel data has been fetched from Wärtsilä vessel database, HIS vessel database and Clarkson vessel database (see Figure 8). This allows user to compare the data and fix it accordingly. S&F also supports the use of Robotic Process Automation (RPA) so it can be utilized for finding and fixing data issues and thus increasing data quality. For example, RPA can check if all the three databases are showing the same data and fix it according if information between different databases are not in line.

	Wärtsilä vessel data	IHS vessel data	Clarksons vessel data
	Vessel information		
Name	SILJA EUROPA	SILJA EUROPA	Silja Europa
IMO number	8919805	8919805	8919805
Status	Operating	In Service/Commission	In Service
Type	Passenger & Cargo Vessel	Passenger/Ro-Ro Cargo Ship	Pass./Car Ferry
Flag	EE	Estonia	ES.
Classification society	-	Bureau Veritas 1993-00	Bureau Veritas
Shipyard	MEYER WERF	Meyer JL GmbH & Co - GEU	403047
Hull number	627	627	627
Delivery date	1993-03-05	1993-03-05	1993-03
Commissioning date	1993-03-05	-	-
Records updated	-	2019-04-22	-

Figure 8.   Wärtsilä Search & Forms UI from search results (Wärtsilä 2019b).

4.1.4   Power BI report

Microsofts software, Power Bi is a business analytics solution provides tools for visualizing and presenting data and sharing it online across organization. Power BI is a collection of software services, connectors and apps that can be used to present data more visually. Data can be extracted from for example, Excel spreadsheet or from multiple cloud-

based and on-premises data warehouses. Data from sources can be presented in a created Power BI report that can then be distributed across organization. In addition, data can be modified before and after fetching it from database. This can shorten the load times and quicken the usage of the reports greatly (Microsoft 2019a).

Power BI report consists of three main parts: Report, Data and Model. Report tool is for creating the actual report and UI, selecting filters and overall presenting the wanted data. Data tool helps user handle the loaded data tables, pre-filter them and for example create custom columns. Model tool presents every data table of the model. In modelling too, user can connect columns between different tables to create connections between them so information more information can be presented. Overall Power BI offers great tools for creating custom reports that can be easily spread across organizations and thus can be used widely (Microsoft 2019b).

Power BI user two programming language: Data analysis expressions (DAX) and Power Query Formula language or "M". These two languages differ from each other quite a bit and they are used in different environments in the Power BI. DAX is the native query language and native formula for Power BI. DAX is used in the "Data" section in the Power BI and it will be applied to the tables after they are loaded from data base. With DAX, user can combine columns from different tables to create custom columns to serve certain purposes. M is powerful mashup query language optimized for building queries. M is used in the "Power Query" section of Power BI. M only works in the moment when the tables are loaded from the data base. After loading the M-codes don't run anymore and this creates certain issues which can be corrected with DAX. Microsoft Power BI offers multiple possibilities for visualisation of the reports and very user-friendly interface (see Figure 9).
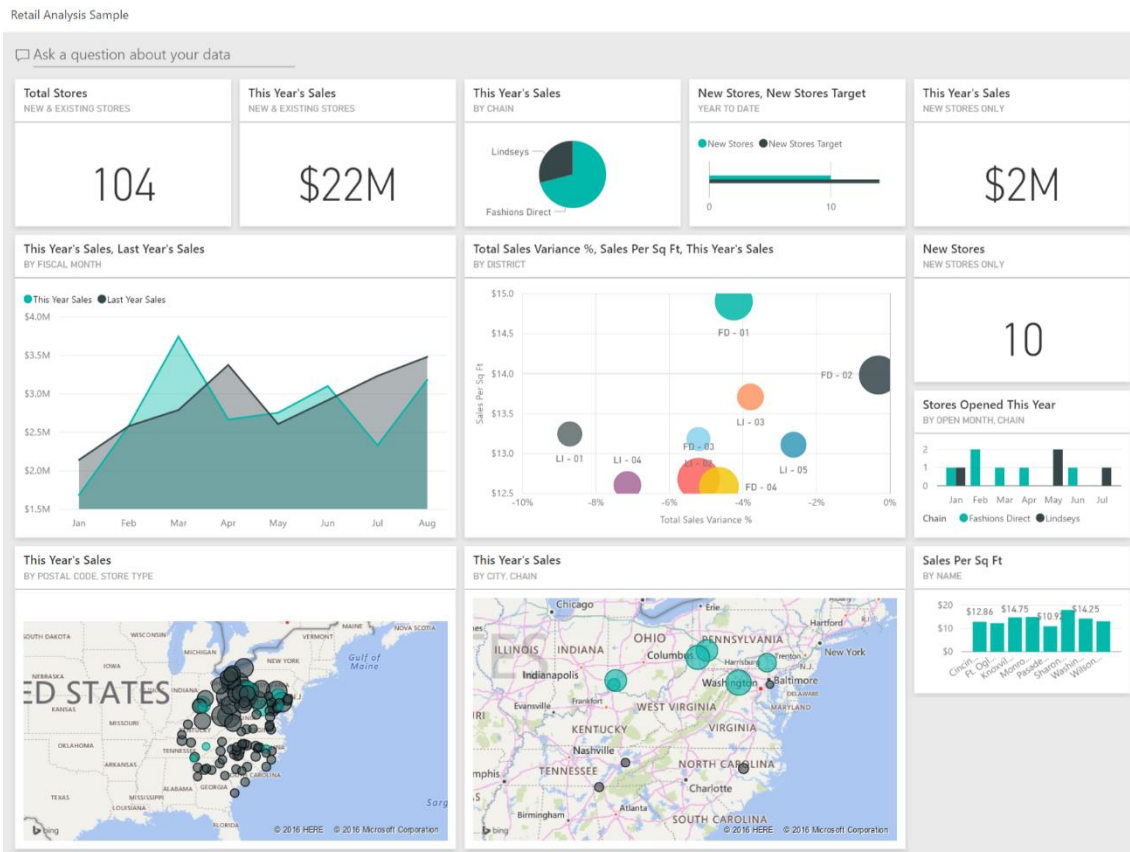
Figure 9.   Example of a Power BI report (Microsoft 2019c).

4.2   Choosing data monitoring solution

SAP and IBM software's are complete solutions that provide good tools for company to start monitoring their data quality. Downside of these solutions are price. As products of large multinational companies, pricing of these products tends to get quite high sometimes. Second drawback is the learning needed for the use of software.

Wärtsilä Search & Forms could provide exact tools for exact needs for Wärtsilä since it is internal software. S&F is free to use, but the needed features for data quality monitoring are not implemented yet to it. This means that to get S&F working for data monitoring, is needs still quite a lot work to get it to version that support data quality monitoring. This work would be done externally so it creates costs for this option.

Power BI is Microsoft reporting tool that is currently in wide use in Wärtsilä and therefore doesn't cause any costs to use. Power BI provides great tools for creating detailed reports for specific needs to monitor and present data quality. Other upside in Power BI is its wide adoption in Wärtsilä which means that there are many Power BI experts and they are easy to contact and available to help.

The output of this thesis will be proof of concept (PoC) that present on the high level the possibilities of data quality monitoring and its impacts on production. Therefore, I will be choosing the Power BI option, since it is free and easy to use to create this kind of PoC. It is highly recommended to switch from basic reporting tool to more complete software to furthermore improve the data monitoring process.

5 MONITORING SYSTEM

After selecting the system that will be implemented, it needs to be built. This step on the thesis contains system building from choosing data bases and data tables to posting monitoring results. Using Power BI as monitoring solution means that some manual work is needed to be done to post monitoring results for data quality.

5.1 Choosing databases

Production External Data Warehouse (EDW) in Wärtsilä is located in Oracle database "wdw2-scan.wartsila.com:1521/edwprod.wartsila.com". In this database, all of the data tables are located in subfolder WBR_ODS. This folder contains source tables that contain data from EDW, dimension tables that are data tables created for storing data. Mandatory data tables for this monitoring software are: V_TS_MARC, V_TS_RESB, V_TD_MATERIAL_PLANT_FI06 and V_TS_AUSP.

V_TS_MARC is standard SAP data table which contains plant data for material table and data. MARC contains metrics for example material numbers, plant numbers, material statuses and MRP profiles and controllers (SE80 2019a). V_TS_RESB is standard SAP table which contains data for reservation and dependent requirements. RESB contains metrics for example reservation numbers, requirement types and quantities (SE80 2019b). V_TD_MATERIAL_PLANT_FI06 is a custom build dimension table that contains data for every material used in SAP in Delivery Center Vaasa (DCV) Finland. One row in this table is one material used in DCV. V_TS_AUSP table is standard SAP table for storing characteristic values for materials, for example classification values regarding traceability (SE80 2019c).

First thing to do was filter the data tables so every time this monitoring report loaded, it wouldn't load the whole tables from EDW. Pre-filtering tables provides faster load times and better user experience. Filtering irrelevant data also improves data quality, as said in one of the six main data quality dimensions: validity.

Next step in implementing monitoring system is connecting the data tables so they can be accessed and used in Power BI. These tables were connected using built-in tool in Power BI (see Figure 10).
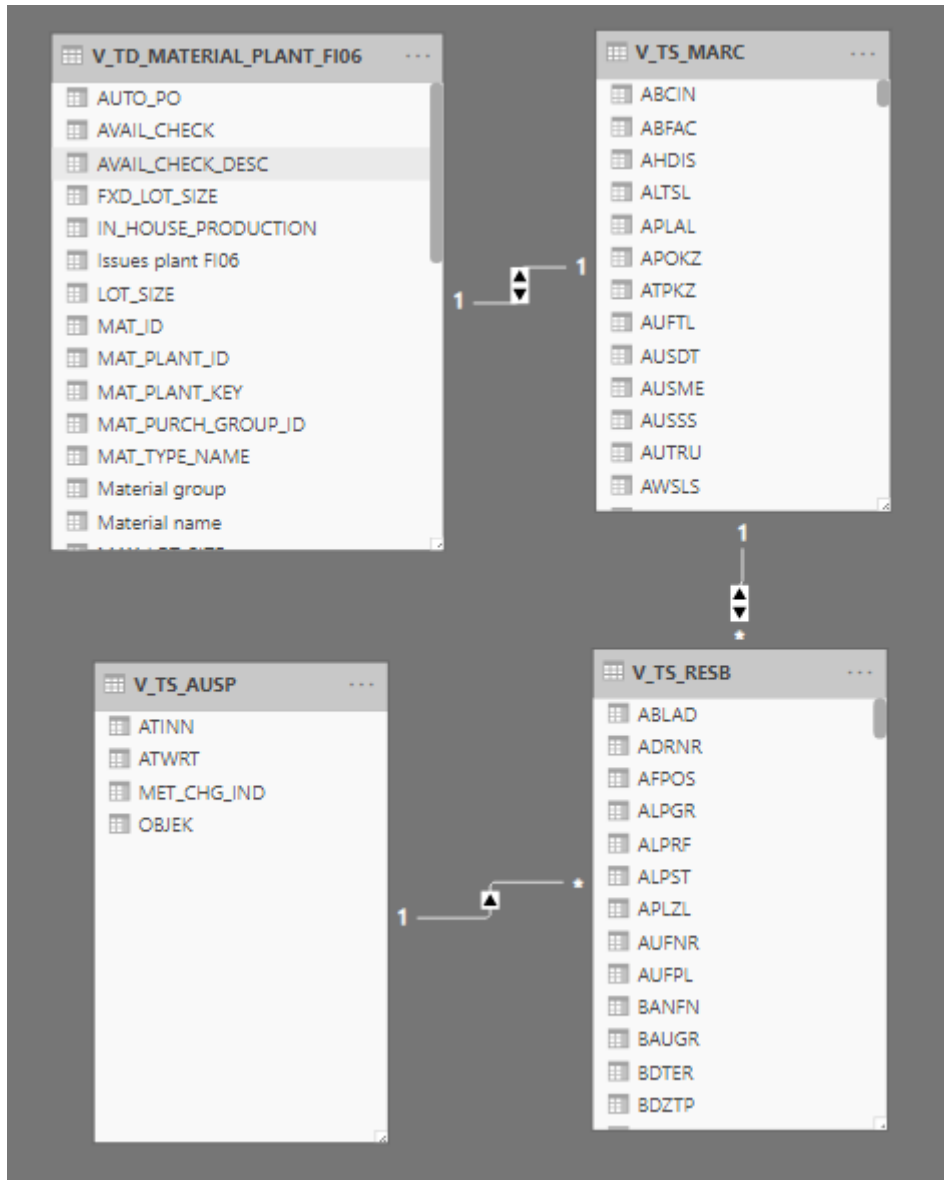
Figure 10.   Data table connections in Power BI.

Table 4.   Relations between data tables in Power BI report.

Source table	Metric	Target table	Metric
V_TS_RESB	Material number	V_TS_MARC	MATNR
V_TS_MARC	MATNR	V_TD_MATERIAL_PLANT_FI06	MAT_ID
V_TS_RESB	Material number	V_TS_AUSP	OBJEK

5.2   Building the report

Using the Power BI report tool, the next step is to build the report itself. Report is named "Material Master Data Check" and it consists of rows of materials, which main orders SoA-dates are in nine weeks in the future and one week in the past. This creates a time window of 10 weeks that will be monitored. Every row contains data of SoA-date of the order, order number, material number or ID, material name and columns of metrics selected in previous chapters that will be monitored. These columns are: material group ID, Material Requirements Planning (MRP) profile, production supervisor, serial number profile, sort string and traceability. Last column of the report is programmed column, that indicates the possible issue in every row and displays it according to issue.

5.2.1   Issues-column

Because of the limitations created by the two programming languages of the Power BI, DAX and M, I needed to do certain steps in order to create the final "Issues"-column that would indicate issues in the material and so that it would be filterable. This means that I had to do several custom columns both in DAX and in M and finally combine them in to one column with DAX. First version of this validation is done only by checking if there

are any values in cells or if certain criteria are met. Future version will show more specific issues in materials with more complex logic behind it.

In table V_TD_MATERIAL_PLANT_FI06, there is one custom column done with M. That column, "Issues plant FI06", is a programmable column that indicates if material has any issues that would affect production based on parameters that exist in that table. Issues plant FI06 -column contains the following code that is commented with #-markings.

```
if ([MRR_PROFILE_ID] = "SA01" or [MRR_PROFILE_ID] =
"SA00" or [MRR_PROFILE_ID] = "PA02" or [MRR_PROFILE_ID]
= "PA04") and [PRODUCTION_SUPERVISOR] = null
then "Missing production supervisor"

#This function checks if there are any material that are
self-made by the plant and don't have production super-
visor. Every material in self-made engine or module needs
to have production supervisor

else if [MRR_PROFILE_ID] = null
then "Missing MRP profile"

#Simple check for MRP profile

else if [MRR_PROFILE_ID] = "FP00" and ([MAT_GROUP_ID] =
"TOBECHECK" or [MAT_GROUP_ID] = null)
then "Missing material group"
#Simple check for missing material group on materials
that's MRP profile isn't "FP00"

else    if    Text.Contains([MAT_ID],    "PAAF")    and
([MAT_GROUP_ID] = "TOBECHECK")
then "Missing material group"

#Check for materials that's material number start with
"PAAF" if they have needed material group

 else "No issues"
```

Table V_TS_MARC contains two simple custom columns with M-language. First column is called "Production scheduling profile.1" and it simply checks if production scheduling profile is missing from materials with MRP profile of "SA00" and "SA01".

```
if ([DISPR] = "SA00" or [DISPR] = "SA01") and [Production
scheduling profile] = null

then "Missing production scheduling profile"

else [Production scheduling profile]
```

The other column "Bulk indicator" checks if the bulk indicator is missing from needed materials. Bulk indicator needs be on every material with MRP profile "PA01".

```
if [DISPR] = "PA01" and [SCHGT] = null
then "Missing bulk indicator"

else [SCHGT]
```

V_TS_RESB contains three custom columns made with DAX. First one is check if material has missing sort string -value. Sort string is custom value in SAP which is used, in Wärtsiläs case, for allocating set code to material. Set code defines which materials are allocated to which activities in MES. This sort string value needs to be on every material that is used in MES. This simple function checks that every material without MRP profile of "FP00" has sort string value.

```
Sort  String  =  IF(RELATED(V_TD_MATERIAL_PLANT_FI06[MRP
profile]) <> "FP00"
&&  V_TS_RESB[Sort  string2]  =  "Missing  sort  string";
"Missing sort string"; "Ok")
```

Other column is the "Issues" column that sums up all the other columns and their outputs for a one single column.

```
Issues = IF(RELATED(V_TS_MARC[Bulk indicator]) = "Miss-
ing bulk indicator"; "Missing bulk indicator";

IF(RELATED(V_TD_MATERIAL_PLANT_FI06[Issues plant FI06])
= "Missing material group"; "Missing material group";

IF(RELATED(V_TD_MATERIAL_PLANT_FI06[Issues plant FI06])
= "Missing MRP profile";  "Missing MRP profile";
```

```
IF(RELATED(V_TD_MATERIAL_PLANT_FI06[Issues plant FI06])
= "Missing production supervisor"; "Missing production
supervisor";

IF(RELATED(V_TS_MARC[Production scheduling profile]) =
"Missing production scheduling profile"; "Missing pro-
duction scheduling profile";

IF(V_TS_RESB[Trace] = "Missing traceability parameters";
"Missing traceability parameters";

IF(V_TS_RESB[Sort  string]  =  "Missing  sort  string";
"Missing sort string"; "No issues")))))))
```

Third custom column is called "Trace". This column checks that every material, that are under traceability, have both serial number profile and traceability parameter. Check is done for inhouse made materials only, these are materials with MRP profile of SA00, SA01, PA02 and PA04.

```
Trace = IF(((RELATED(V_TS_AUSP[ATWRT]) <> ""

&& RELATED(V_TS_MARC[Serial number profile]) = "")

||   (RELATED(V_TS_AUSP[ATWRT])   =   ""   &&   RE-
LATED(V_TS_MARC[Serial number profile]) <> "")) && (RE-
LATED(V_TD_MATERIAL_PLANT_FI06[MRP profile]) = "SA00"

||  RELATED(V_TD_MATERIAL_PLANT_FI06[MRP  profile])  =
"SA01"

||  RELATED(V_TD_MATERIAL_PLANT_FI06[MRP  profile])  =
"PA02"

||  RELATED(V_TD_MATERIAL_PLANT_FI06[MRP  profile])  =
"PA04"); "Missing traceability parameters"; "Ok")
```

If either serial number profile or traceability parameter is missing, this column outputs "Missing traceability parameters". Output is "Ok" if no data is missing.

5.2.2   Rolling date on monitoring results

This reporting tool needs to show materials from the upcoming nine weeks, so that needed actions can be taken well in advance before SoA date. The main idea is that every material with issue needs to be shown if it has SoA date smaller than current date plus 63 days, and 7 days to past from current date. The main idea is that these issues have now 10 weeks to be visible and cause correcting actions.

Rolling date was done by creating custom column from SoA date for each material and formatting it to from of DD/MM/YYYY. After that I created another custom column that indicated current date in the same date form as SoA date. Next step was to make column that indicated the difference of days between these two columns. It was done by using simple line of code in DAX in the "Data" tab of the Power BI report:

```
DateDifference = DATEDIFF([Today]; [Date5]; DAY)
```

In this line of code, the [DateDifference] is the name of new column. Using Power BI function "DATEDIFF" to calculate the difference in in days between [Today] and [Date5], which is the formatted date of the SoA date. Using this date difference, I was able to create another simple line of code in DAX that presented the final date that would be used as filtered SoA date in the report:

```
Date = IF([DateDifference] >= -7 && [DateDifference] <=
63; [Date5]; 0)
```

This function serves as a filter that filters out dates that aren't in the wanted time window. Function checks if the date difference is greater or equal than -7 (one week backwards) and less or equal than 63 (nine weeks to the future). If this function is true, the output in that specific line is the SoA date. If this function is not true, the output is 0 which in this case is 30.12.1899. After this the final filtering to the report was done that it will only shows dates that aren't 30.12.1899 and this way the report will show rolling date of one week to the past and nine weeks to the future.

5.2.3   User interface creation

After creating on column to indicate issues and doing rolling date for report, the next step is to build the reports interface. I started creating interface by selecting parameters of materials to show to user. These parameters consist of: SoA date, Order number, Product hierarchy name, Material number, Material name, MRP profile, Material group, Serial number profile, Production supervisor and custom columns: Bulk indicator, Production scheduling profile, Sort string and Issues.

After selecting and columns and adding them to report, I created filter for custom column "Issues" and "Product hierarchy name". Whit these filters users can filter needed information. For example, user can only be interested in issues in his/her are of responsibility and with "Product hierarchy name" user can easily filter wanted product, for example Wärtsilä 32 (see Figure 12). Issues-column works the same way as users can filter issues they are interested in and thus focus on wanted issues (see Figure 11).

Issues
- ☑ Missing material group
- ☑ Missing MRP profile
- ☑ Missing production schedulin…
- ☑ Missing production supervisor
- ☑ Missing sort string
- ☑ Missing traceability parameters
- ☐ No issues

Figure 11.   Multiselection filter for issues.
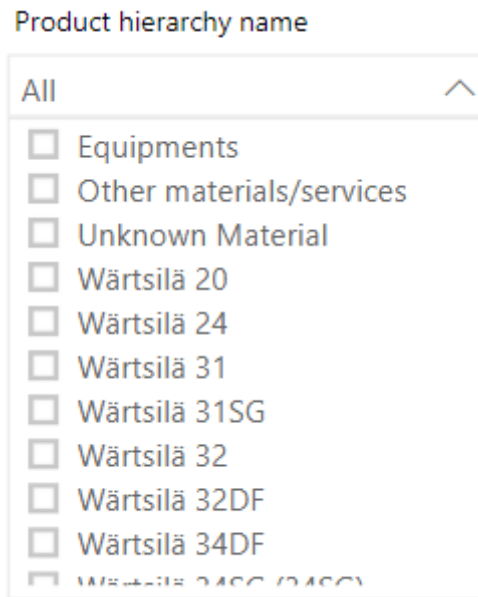
Figure 12.   Multiselection filter with drop down column for product hierarchy names.

Report in final form includes previously mentioned filters for issues and products and includes list of materials that are filtered accordingly. This list updates as user changes selections on filters. Due the size of the report, documentation of the report results is done in sections of the report, see Figures 13, 14 and 15. The full report is seen in Appendix 1.

Date	Order	Product hierarchy name	Material number	Material name
15.11.2019	9108525	Other materials/services	PAAG271954	CYLINDER, RIGHT - 6951FP-22-21
18.11.2019	2536021	Other materials/services	PAAG001260	BEARING - B71932E.TP4S.UL
18.11.2019	2536021	Other materials/services	PAAG271002	BEARING - B71934E.T.P4S.UL
18.11.2019	7721311	Wärtsilä 24	PAAF958982	CONNECTING ROD SHAFT & HOUSING
2.12.2019	1936061	Unknown Material	XAAB682786	24CH001 CH PARTS FREEZING
10.12.2019	2301201	Wärtsilä 20	XAAB623105	W6L20 SP/04606.1.S.FS1-P11 OKSKAYA 5902
11.12.2019	4386364	Unknown Material	WQDCRES13	MAY:W6L20,E,TIANJIN XINGANG HAIJING 1111
18.12.2019	4386364	Unknown Material	WQDCRES03	MAY:W9L20DF, B, BUNKER V.SEMBCORP
9.1.2020	2141462	Wärtsilä 31	PAAF862467	GUIDE BLOCK SERVICE ASSEMBLY (DF)
22.1.2020	4386364	Unknown Material	WQDCRES12	MAY:1520W8L20,D4,PENNA CEMENT 2
22.1.2020	4386364	Unknown Material	WQDCRES21	MAY:W6L20,C6,KRASNOE 06022,108631

Figure 13.   First section of the report results.

Material name	MRP profile	Material group	Serial number profile	Traceability	Production supervisor	Bulk indicator
CYLINDER, RIGHT - 6951FP-22-21		230110				
BEARING - B71932E.TP4S.UL		230110				
BEARING - B71934E.T.P4S.UL		230110				
CONNECTING ROD SHAFT & HOUSING	SA00	102140			402	
24CH001 CH PARTS FREEZING	SA00	TOBECHECK	ZTCB		312	
W6L20 SP/04606.1.S.FS1-P11 OKSKAYA 5902		190100	ZTCB	A+	503	
MAY:W6L20,E,TIANJIN XINGANG HAIJING 1111		190100			503	
MAY:W9L20DF, B, BUNKER V.SEMBCORP		190100			503	
GUIDE BLOCK SERVICE ASSEMBLY (DF)	SA00	150610			402	
MAY:1520W8L20,D4,PENNA CEMENT 2		190100			503	
MAY:W6L20,C6,KRASNOE 06022,108631		190100			503	

Figure 14.   Second section of the report results.

Bulk indicator	Production scheduling profile	Sort String	Issues
		Ok	Missing MRP profile
		Ok	Missing MRP profile
		Ok	Missing MRP profile
	Missing production scheduling profile	Ok	Missing production scheduling profile
	Z001NP	Missing sort string	Missing traceability parameters
	Z001NP	Missing sort string	Missing MRP profile
	Z00003	Ok	Missing MRP profile
	Z00003	Ok	Missing MRP profile
	Missing production scheduling profile	Ok	Missing production scheduling profile
	Z00003	Ok	Missing MRP profile
	Z00003	Ok	Missing MRP profile

Figure 15.   Last section of the report results.

5.3   Deployment and usage

Material master data check -report was deployed 26.9.2019 for team in delivery management to monitor issues in material master data. Deployment was done by publishing the Power BI report to existing Power BI workspace. Weekly meetings were arranged for meeting with delivery management for discussion and improvements for the report. Based on received comments, changes are updates were made to the report. These changes included adding filters and algorithms for finding all the issues that were defined in the original scope of this study.

Monitoring process has now been running for few months and the feedback is mostly positive. Few fixes and tweaks have been made to the report, but main functions defined in the original scope has been kept the same.

6 ANALYSIS

This chapter consist of the last three steps of the data quality strategy used in this thesis. These steps include baseline assessment, posting monitoring reports and reviewing monitoring trends.

6.1 Baseline assessment

After building the report, the next step is setting a baseline for future comparison of the results. As mentioned before, baseline assessment is the first set of tests conducted to which the future results of the monitoring will be compared. Furthermore, the baseline assessment will serve as prototype for the monitoring software.

Baseline assessment was done in 25.9.2019 with certain scope that is not to be confused with the original scope of the whole process that included all parameters defined earlier in chapter 3.2. This scope for the prototype version and baseline of the report contained following issue checks: Missing material group, missing MRP profile and missing production supervisor. These issues summed total of 381 in a defined rolling ten-week time period. This means different material from different planned and production orders that are missing master data.

6.2 Monitoring reports and trends

Last part of the six-step process was to continuously post monitoring reports and be up-to-date on monitoring trends. One critical part of these kinds of monitoring programs is to make sure that the process continues after deployment. Continuity and success of this report can be improved by continuously developing the process and report and to continuously following the monitoring reports and trends overall.

Comparison analysis for the baseline assessment was done 25.11.2019 when the report was complete with the final scope. This includes all the issues that were defined earlier in this thesis. For comparison to baseline assessment, second assessment was done with same parameters and issues that baseline assessment was done with. The result was 1 issue in different material in a rolling ten-week time window. This means over 99% drop in issues comparing time before report and after deployment of the final version the report.

The report serves as work list for the team in delivery management but it also servers as monitoring tool for management which indicates the amount of type of issues in materials. This way management can follow the monitoring trend closely, since the report shows a rolling ten-week time windows with nine weeks to the future and one week in the past. This way they can easily address upcoming issues and get them fixed before SoA-date.

In chapter 3.1 Requirements for the monitoring system, four requirements were presented. These requirements were:

- Monitoring system resulting this thesis needs to be simple and easy to use.

- Monitoring system doesn't need to be maintained but rather it can be upgraded.

- Result of the monitoring process needs to be easily accessible across the company if needed.

- Monitored data needs to show the coming eight weeks from current date on rolling basis in the form of start of assembly (SoA) dates.

These requirements were met in the final version of the monitoring report and process.

7 CONCLUSIONS

The findings of this thesis were:

- Continuous data monitoring can greatly increase data quality in monitored processes.

- In a ten-week time window, defined issues decreased by over 99% compared from baseline assessment to present time.

- Future of the data quality monitoring processes in case company could be done with company's own solution.

Research questions and their answers are presented below.

"What is data driven manufacturing in practice?"

Data driven manufacturing in practice is that data is the main driver in manufacturing. There are few key requirements for manufacturing to be operated properly. These requirements are: materials, work force and tools. Data driven manufacturing adds fourth requirements to this equation: data. Today and in the future manufacturing processes, data is and will be a key component when producing goods and services. Bad data quality can and will lead in to troubles if it is not managed properly.

If data is not up to date or completely missing, production may halt. For example, manufacturing execution systems may not work properly when some critical data is missing. This may cause delay in the start of the production which could delay the whole delivering process of the manufactured goods.

"What is the appropriate model to monitor and control the data quality in smart manufacturing in case company?"

Currently in case company, the best solution is to start monitoring with free and easy to use Power BI report. This way the monitoring will be maintained locally, and this the customization can be kept agile, cheap and easy. Other options were also considered in this study, but Power BI was chosen because other options would have caused costs. This monitoring model will be updated and upgraded after completion of this thesis and will serve as tool for delivery management for monitoring and fixing issues related to material data quality.

This report serves as prototype with certain scope for specific needs. Report will be used by certain team. The key to success is to implement this prototype for wider use with multiple parameters so it could benefit other organizations as well. The main goal is to take this prototype and turn it in to globally supported process that would fix data quality issues not only on organization or on plant level, but on global level.

"What is the aim for the future of data quality monitoring in case company?"

Currently in case company the main point of the material data quality monitoring is to give transparent view on state of quality of the material information on systems. Currently there are team and plant specific dashboards and reports serving specific group or organization. Filters in these dashboards provide users easy way to focus on material information they are interested in, regardless of the material type. Another main target of these reports is to improve the data quality. Reports help interested parties and data product data personnel by highlighting potentially false data in materials.

Future of data quality monitoring in case company could be company's own solution, according to Master Data organization. Power BI reporting could serve wider user base and even multiple plants with data quality issues, but better choice for future and in the long would be completely own system for data quality monitoring. This system provides tools for not just material data quality monitoring but many other aspects of data quality too. This includes for example customer and supplier data.

Steps for coming years in case company is to widen the concept of data monitoring and increase the transparency of data for better understanding the quality state of different attributes. This means faster reaction for possible issues. Way of the future is to proactively monitor data and not just to fix the bad data but to focus on root causes and fix them instead. Goal is to create system that can work both globally and locally.

8 SUMMARY

The purpose of this thesis was to study data quality and its effect on smart manufacturing. This thesis consisted of literature review as theoretical part, studying and building process and tool for measuring and improving data quality in defined scope. Best and fastest way to monitor data quality was to create report using Microsoft Power BI. This report serves as monitoring tool and work list for delivery management to find and fix issues in material data quality.

Based on this study and literature review, a good process for improving data quality was found. This process consists of six simple steps, that when followed correctly, can yield great improvements in data quality. This study showed that continuous data quality monitoring can cause great improvements in data quality. In study that included few parameters defined in the original scope for baseline assessment, a decline of 99% in issues in material master data was noted. This decline indicates well how simple and inexpensive steps and tools can affect data quality greatly. This is a good way to continue addressing and fixing data quality issues in the future and most importantly upgrading systems and processes to further improve data quality.

REFERENCES

BCG (2019). *Embracing Industry 4.0 and Rediscovering Growth* [Online]. Available at https://www.bcg.com/en-ao/capabilities/operations/embracing-industry-4.0-redis-covering-growth.aspx [Accessed 1.9.2019]

BI Survey (2019). *Data Quality and Master Data Management: How to Improve Your Data Quality* [Online]. Available at https://bi-survey.com/data-quality-master-data-management#def [Accessed 4.8.2019]

Business Objects (2008). *Data quality strategy: a step-by-step approach* [Online]. Available at https://pdfs.seman-ticscholar.org/3a5e/eca956aea600f28826dc983ccfd589e63855.pdf [Accessed 20.5.2019]

CMTC (2019). *What is Smart Manufacturing? (Part 1A)* [Online]. Available at https://www.cmtc.com/blog/what-is-smart-manufacturing-part-1a-of-6 [Accessed 5.7.2019]

DAMA UK (2013). *The Six Primary Dimensions for Data Quality Assessment* [Online]. Available at https://www.whitepapers.em360tech.com/wp-content/files_mf/1407250286DAMAUKDQDimensionsWhitePaperR37.pdf [Accessed 15.8.2019]

DataPoint (2018). *Industry 4.0: Importance of Data and Signals and Triggers* [Online]. Available at https://www.presentationpoint.com/blog/data-signals-triggers-industry-4-0/ [Accessed 22.8.2019]

Davis, J., Edgar, T., Porter, J., Bernaden, J. & Sarli, M. (2012). *Smart manufacturing, manufacturing intelligence and demand-dynamic performance* [Online] Available at https://www.sciencedirect.com/science/article/pii/S0098135412002219 [Accessed 1.8.2019]

Deloitte (2019). *The smart factory* [Online]. Available at https://www2.deloitte.com/content/dam/insights/us/articles/4051_The-smart-factory/DUP_The-smart-factory.pdf [Accessed 10.7.2019]

Experian (2019). *What is a Data Quality Dimension?* [Online]. Available at https://www.experian.co.uk/business/glossary/data-quality-dimensions/ [Accessed 15.8.2019]

Forbes Insights (2010). *Managing Information in the Enterprise: Perspectives for Business Leaders* [Online]. Available at https://i.forbesimg.com/forbesinsights/StudyPDFs/SAP_InformationManagement_04_2010.pdf [Accessed 19.8.2019]

IBM (2019). *IBM InfoSphere Information Server for Data Quality* [Online]. Available at https://www.ibm.com/us-en/marketplace/infosphere-info-server-for-datamgmt [Accessed 5.9.2019]

Kahn, B.K., Strong, D.M. & Wang, R.Y. (2002). *Information Quality Benchmarks: Product and Service Performance* [Online]. Available at http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.5.4752&rep=rep1&type=pdf [Accessed 5.8.2019]

Klaus, H., Rosemann, M. & Gable, G.G. (2000). *Information Systems Frontiers 2: 141* [Online]. Available https://doi.org/10.1023/A:1026543906354 [Accessed 21.5.2019]

Lasi, H., Fettke, P., Kemper, H-G., Feld, T. & Hoffmann, M. (2014). *Industry 4.0* [Online]. Available at https://i.forbesimg.com/forbesinsights/StudyPDFs/SAP_InformationManagement_04_2010.pdf [Accessed 20.8.2019]

Lu, Y. (2017). *Industry 4.0: A survey on technologies, applications and open research issues* [Online]. Available at https://www.sciencedirect.com/science/article/pii/S2452414X17300043 [Accessed 25.8.2019]

McClellen, M. (1997). *Applying Manufacturing Execution Systems* [Online]. Available at https://www.taylorfrancis.com/books/9780367805272 [Accessed 1.7.2019].

Microsoft (2019). *What is Power BI?* [Online]. Available at https://powerbi.microsoft.com/en-us/what-is-power-bi/ [Accessed 6.9.2019]

Microsoft (2019b). *Power BI Overview* [Online. Available at https://docs.microsoft.com/en-us/power-bi/fundamentals/power-bi-overview [Accessed 6.9.2019]

Microsoft (2019c). *Retail Analysis sample for Power BI: Take a tour* [Online]. Available at https://docs.microsoft.com/en-us/power-bi/sample-retail-analysis [Accessed 6.9.2019]

NIST (2019). *Smart Manufacturing Operations Planning and Control* [Online]. Available at https://www.nist.gov/programs-projects/smart-manufacturing-operations-planning-and-control-program [Accessed 7.7.2019]

Oracle (2019). *What is ERP?* [Online]. Available at https://www.oracle.com/applications/erp/what-is-erp.html [Accessed 4.6.2019]

Rouse, M. (2017). *Manufacturing execution system (MES)* [Online]. Available at https://searcherp.techtarget.com/definition/manufacturing-execution-system-MES [Accessed 29.6.2019]

SAP (2019). *Enterprise Data Management Software* [Online]. Available at https://www.sap.com/products/data-services.html [Accessed 5.9.2019]

Schubert (2019). *MES (Manufacturing Execution System)* [Online]. Available at https://www.gerhard-schubert.fr/fr/mes_manufacturing_execution_system/29 [Accessed [3.7.2019]

SE80 (2019a). *MARC SAP Plant Data for Material Table and Data* [Online]. Available at https://www.se80.co.uk/saptables/m/marc/marc.htm [Accessed 20.9.2019]

SE80 (2019b). *RESB SAP Reservation/dependent requirements Table and data* [Online]. Available at https://www.se80.co.uk/saptables/r/resb/resb.htm [Accessed 20.9.2019]

SE80 (2019c). *AUSP SAP Characteristic Values Table and data* [Online]. Available at https://www.se80.co.uk/saptables/a/ausp/ausp.htm [Accessed 20.9.2019]

Sikander, A. (2017) *Influence of Strategic Technology Management on Smart Manufacturing: The Concept of 'Smart Manufacturing Management'* [Online]. Available at https://www.intechopen.com/books/digital-transformation-in-smart-manufacturing/influence-of-strategic-technology-management-on-smart-manufacturing-the-concept-of-smart-manufacturi [Accessed 1.8.2019]

Skew Infotech (2019). *Enterprise resource software planning method* [Online] Available at https://www.skewinfotech.com/blog/2019/02/20/enterprise-resource-software-planning-method/ [Accessed 15.6.2019]

Tao, F., Zhang, M. & Nee, A.Y.C. (2019) *Digital Twin and Big Data – Chapter 9* [Online]. Available at https://www.sciencedirect.com/science/article/pii/B9780128176306000096 [Accessed 10.7.2019]

Wang, R.Y. & Strong, D.M. (2013) *Beyond Accuracy: What Data Quality Means to Data Consumers* [Online]. Available at https://courses.washington.edu/geog482/resource/14_Beyond_Accuracy.pdf [Accessed 3.8.2019]

Waycott, A. (2016). *What is Smart Manufacturing?* [Online]. Available at https://www.industryweek.com/systems-integration/what-smart-manufacturing [Accessed 2.8.2019]

Wärtsilä (2019a). *About Wärtsilä* [Online]. Available at https://www.wartsila.com/about [Accessed 4.5.2019]

Wärtsilä (2019b). *Wärtsilä Search* [Online]. Available at https://search.forms.wartsila.com/ [Accessed 5.9.2019]

APPENDIX 1. View of the full report.

Date	Order	Product hierarchy name	Material number	Material name	MRP profile	Material group	Serial number profile	Traceability	Production supervisor	Bulk indicator	Produ
5.11.2019	7714906	Standard	PAAC001430	HEXAGON THIN NUT M16X1.5	BM00	117320				X	
5.11.2019	7714906	Wärtsilä Vasa 32	0024L000001	(SEE SPEC)VIBRAT.DAMPER	PA01	99999				X	
5.11.2019	7714906	Unknown Material	0024C000000	(SEE SPEC.) TUNING MASS	PA01	99999				X	
5.11.2019	7714906	Unknown Material	0024C000001	(SEE SPEC.)DISTANCE RING	PA01	99999				X	
5.11.2019	7714913	Unknown Material	XAAB395554	20CA001 CHARGE AIR COOLER MODULE	ACTP	TOBECHECK				X	
5.11.2019	7714914	Unknown Material	XAAB372189	20CS001 CRANKSHAFT MODULE	ACTP	TOBECHECK					
5.11.2019	7714583	Unknown Material	XAAB639761	20PI001 PISTON ASSEMBLY	ACTP	TOBECHECK					
5.11.2019	7712505	Unknown Material	XAAB710702	31GV001 GAS VENTING VALVE	ACTP	TOBECHECK					
5.11.2019	7713450	Unknown Material	XAAB607425	31LO001 LUBE OIL MODULE	ACTP	TOBECHECK					
5.11.2019	7713450	Unknown Material	XAAB363425	31LO002 THERMOSTATIC VALVE	ACTP	TOBECHECK					
5.11.2019	7713450	Unknown Material	XAAB607427	31LO003 HEAT EXCHANGER	ACTP	TOBECHECK					
5.11.2019	7713450	Unknown Material	XAAB363427	31LO005 AUTOMATIC FILTER	ACTP	TOBECHECK					
5.11.2019	7713450	Unknown Material	XAAB667070	31LO006 MAIN CABINET & CABLE RAILS	ACTP	TOBECHECK					
5.11.2019	7713450	Unknown Material	XAAB667071	31LO010 COVER FOR LOM	ACTP	TOBECHECK					
5.11.2019	7713450	Unknown Material	XAAB600296	31LO011 FASTENING EQUIPMENT	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB390784	31MA001 BEARINGS	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB667074	31MA002 ENGINE BLOCK	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB565646	31MA004 UPPER FLANGES FOR EB INCL.	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAA625361	31MA006 INTERMEDIATE GEAR BEARINGS	PA01	TOBECHECK				X	
5.11.2019	7713963	Unknown Material	XAAB214189	31MA007 BALANCING SYSTEM BEARINGS	PA01	TOBECHECK				X	
5.11.2019	7713963	Unknown Material	XAAA625556	31MA008 CAMSHAFT BEARINGS	PA01	TOBECHECK				X	
5.11.2019	7713963	Unknown Material	XAAB565606	31MA009 MAIN BEARING SCREWS	PA01	TOBECHECK				X	
5.11.2019	7713963	Wärtsilä/Vasa Common	PAAE325182	31MA010 MAIN BEARING SHELLS UPPER	ACTP	NONPURCH					
5.11.2019	7713963	Wärtsilä/Vasa Common	PAAE325183	31MA011 MAIN BEARING CAPS (INCL. LO	ACTP	NONPURCH					
5.11.2019	7713963	Unknown Material	XAAB580953	31MA014 ENGINE BLOCK EQUIPMENT	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB667079	31MA016 ELECTRICAL EQUIPMENT	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB363439	31MA017 INTERMEDIATE GEARS	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB390793	31MA018 BALANCING SYSTEM	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB363441	31MA019 OIL PIPES BEHIND CAMSHAFT	ACTP	TOBECHECK					
5.11.2019	7713963	Equipments	XAAB449814	31MA020 A-BANK CAMSHAFT	ACTP	TOBECHECK					
5.11.2019	7713963	Equipments	XAAB449815	31MA021 B-BANK CAMSHAFT	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB723715	31MA022 TURNING DEVICE SUBASSEMBLY	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB711506	31MA026 CRANKSHAFT ASSEMBLY	ACTP	TOBECHECK					
5.11.2019	7713963	Unknown Material	XAAB139942	31MA028 TEMPERATURE SENSORS FOR MB	PA01	TOBECHECK				X	

Issues
< Missing material group
< Missing MRP profile
< Missing production schedulin...
< Missing production supervisor
< Missing sort string
< Missing traceability parameters
No issues

Product hierarchy name
All