The Digital Transition of Manufacturing

The gathering and visualization of real-time data and historical data to monitor, control, and improve manufacturing performance

G.D. Schutte

12-04-2021

Master Thesis

Industrial Engineering and Management

UNIVERSITY OF TWENTE.

DYKA

General Information

The Digital Transition of Manufacturing

The gathering and visualization of real-time data and historical data to monitor, control, and improve manufacturing performance

Zwolle, 12-04-2021

Author:	G.D. Schutte guus.schutte@dyka.nl g.schutte@schutteconsultancy.nl

Document:	Master Thesis Industrial Engineering and Management (essay.utwente.nl version)
Educational institution:	University of Twente, Enschede *Faculty: Behavioral Management and Social Sciences* *Department: Industrial Engineering & Business Information Systems (IEBIS)*
Educational program:	Master Industrial Engineering and Management *Specialization: Production and Logistics Management (PLM)* *Orientation: Manufacturing Logistics* *Additional specialization: Maintenance*
Thesis company:	DYKA, Steenwijk Productieweg 7, 8331 LJ, Steenwijk *Extrusion department*
Supervisors:	University of Twente: Prof. Dr. J. van Hillegersberg, *Full Professor* M. Koot, *PhD Candidate* DYKA Steenwijk: Y. de Boer, *Process Innovator* J. Pleijsier, *Manager General Engineering*

Preface

This document forms the final stage of my study Industrial Engineering and Management. It is time to start with the normal working life after 6.5 years of studying, social life, and adventure. Studying will be replaced with working, and social life and adventure will remain part of my life hopefully.

The thesis is written at the extrusion department of the plastics pipes manufacturer DYKA Steenwijk. Where I researched the future possibilities of manufacturing data gathering to monitor, control, and improve operations. This thesis consists of a literature review, which is generalizable for production companies, and a case study, which transforms the theory into practical advice for DYKA. Data utilization was/is an interesting topic to research and will remain interesting in the coming years. This because the MES development and the BI-tool development will become future projects within the company. I am looking forward to remaining part of this digital transition in my new job as a business information developer at DYKA.

I want to thank everyone who has supported me during this thesis. Especially, I want to thank my supervisors Jos van Hillegersberg and Martijn Koot from the university and Ytsen de Boer and Jacco Pleijsier from Dyka. As well as all other colleagues at DYKA who supported me with interesting information and drinking coffee.

For now, I hope that you enjoy reading this thesis. Hopefully, you will learn more about the promising future of data utilization in manufacturing environments.

Guus Schutte
Zwolle, 12-04-2021

Management summary

Monitoring and controlling manufacturing environments is a difficult task. Multiple activities and behaviors have to be managed to control operations. Process definitions and standards are crucial to create a controlled working environment. New technologies can also support an increase in control of manufacturing facilities. For instance, by adopting advanced IT-systems, automate processes, or executing simulations to optimize machine utilization. The key to control and improve manufacturing is information. This information can be created from gathered data. Manufacturing companies can design their IT-systems into data-gathering platforms that support the company with information for decision-making. This data-supported decision-making must lead to an improvement in manufacturing performance. This case will be researched in this thesis by answering the following research question:

> **Research question:** How should a production company/DYKA design its information system architecture to monitor, control, and improve operational performance and be ready for future technologies?

The research consists of two parts, the theoretical answer to the research question in the literature part and the practical answer to the research question in the case study part. The combination creates a generalizable theoretical body of knowledge for production companies and specific advice with future implementations for the case study company DYKA.

Literature

Modern technologies are interesting for improving operational performance. These modern technologies can be grouped under the name of Industry 4.0, which indicated the fourth industrial revolution. Industry 4.0 focuses on the digitalization of manufacturing. The foundation for the modern Industry 4.0 technologies is data. A company can effectively integrate new technologies if there is a data foundation because the technologies consume data and create data. This sequence of data usage and creation transforms companies into industrial internet of things. This means that there is real-time data communication of physical objects. Integration is the key to the digital transition of manufacturing. The internet of things will be created by the horizontal integration and vertical integration of information systems. Vertical integration is the software integration from shop floor systems to top-level management systems in a company and horizontal integration is the integration of systems between companies within the value chain. The advancement of integrated IT-systems and processes at companies can be measured by its digital maturity. This research explains four stages of digital maturity: a company can be a digital novice, a vertical integrator, a horizontal integrator, or a digital champion. The research focuses on the step-change from a digital novice into a vertical integrator.

The vertical integrator maturity level can be obtained by integrating the information system within the company according to the ISA-95 standard, which forms the basis for most of the manufacturing integration standards. This standard explains the interfacing and functions of systems that connect the operational shopfloor with strategic top floor management. This is realized by the connection of machines (SCADA-network), manufacturing operations monitoring and control systems (MES-system), and business planning and logistics systems (ERP-system). The literature explains three steps

to gather and utilize data for improving manufacturing performance. The three steps to becoming a vertical integrator are:

Step 1: The overall implementation of a data acquisition platform
Step 2: Real-time data gathering and monitoring
Step 3: Getting an alarm list that monitors the production parameters

This digital shop floor management results in effective monitoring, diagnosing, and prognosticating of activities on the shop floor, which will decrease downtime of machinery and increases the quality of manufactured products. The vertically integrated information systems form the basis for manufacturing intelligence and the adoption of modern industry 4.0 technologies.

Connected data will support manufacturing environments with real-time operational parameters and performance data to monitor and control operations. In addition to this, the created information will support operational staff with stored data for real-time and historical analytics. The created big data must be accessible and understandable to be supportive and effective for operational, tactical, and strategical decision-making. Vertical integration will create a manufacturing facility that can be monitored and controlled. This data-supported operation will be created by connecting the MES-system and ERP-system into data-gathering platforms. In this case, the systems will provide information that supports decision-making processes. In addition to this, the data foundation forms the basis for adopting and connecting modern industry 4.0 technologies and other analytics to create a smarter manufacturing environment.

Advice for the company
DYKA is currently in the digital novice state of digital maturity. This means that it has a partial integration of IT-systems, first digital solutions, isolated applications, no manufacturing digitalization focus, and the analytical capabilities are mainly based on semi-manual data extracts. Currently, manufacturing data is hardly utilized for controlling operations and improvement projects. Machinery data is gathered in most extruder machines but is not gathered in a central data gathering platform. The step-change into the vertical integrator maturity level is attained when the company transforms IT-systems to create digital manufacturing coordination control, a homogeneous IT architecture, a connection between different data cubes, a Machine-to-Machine network, and data as the key differentiator for the business. This digital factory can be created by connecting the ERP system, the MES-system, and the shop floor to create an industrial internet of things. Information sharing between those systems will result in real-time data supervision of operations. The gathered big data can be visualized in a BI-tool, this tool functions as a big data platform that stores and presents the data by providing reports and dashboards.

Those dashboards and reports can provide constructive information for controlling KPIs. The research proposes four examples that can be created with gathered manufacturing data. These examples provide insights that should motivate the company to embed data-gathering in systems and visualization in a BI-tool into their strategy. Example one explains the importance of the alarm function of the MES-system. This function can prevent downtime and quality issues. The second example shows the monitoring and control dashboard for the mass per meter KPI of the extrusion department, which is important because monitoring this parameter will control the material usage costs. The third example presents a quality dashboard for operators and staff, this dashboard is created to show the ease of use of a BI-tool in comparison with the current complex to use reporting tool. The last example contains an overall equipment effectiveness (OEE) dashboard, which will be used to control production performance. Currently, each different information system has its independent reporting tool, which makes the ease of use and data accessibility difficult. One BI-tool that can be connected to all different data sources has a major advantage in the ability to execute and the completeness of vision. This adoption of a BI-tool creates opportunities for data visualization for both manufacturing data and business performance data. The MES development and BI development will support the company in obtaining the vertical integrator digital maturity level. The development of the MES-system contributes to an improvement of operational performance and the utilization of a BI-tool leads to an increase of accessible information of both business and manufacturing data.

In summary, the information system architecture must be vertically integrated to support the company in monitoring, controlling, and improving its manufacturing performance. The MES-system should utilize real-time data to monitor and control operations. A big data platform with a BI-tool as a front should be developed to perform historical analytics on the gathered data. Modern technologies create and rely on data, so gathering and utilizing data is the key to make new industry 4.0 technologies accessible. The manufacturing environment should be transformed into an industrial internet of things (IIoT) to create smart manufacturing.

Table of contents

List of abbreviations

APICS	American Production and Inventory Control Society
BCG	Boston Consulting Group
BI	Business Intelligence
BPNM	Business Process Model and Notation
CMMI	Capability Maturity Model Integration
DSM	Dutch State Mines
ERP	Enterprise Resource Planning
ETL	Extract, Transform, Load
IIoT	Industrial Internet of Things
IoT	Internet of Things
IP	Internet Protocol
ISA-95	International Society of Automation 95
IT	Information Technology
JDP	John Davidson Pipes Ltd
KPI	Key Performance Indicator
MES	Manufacturing Execution System
Mesa-11	Manufacturing Enterprise Solutions Association 11
MFI	Meld Flow Index
MTTF	Mean Time to Failure
OEE	Overall Equipment Effectiveness
OPC	Open Platform Communications
OPC UA	Open Platform Communications United Architecture
PE	Polyethylene
PLC	Programmable Logic Controller
PP	Polypropylene
PVC	Polyvinyl chloride
PWC	PricewaterhouseCoopers
QC	Quality Control
RPM	Revolutions per minute
SCADA	Supervisory Control and Data Acquisition
SMEs	Small and Medium-sized Enterprises
SPICE	Simple Protocol for Independent Computing Environments
STIS	Specific Tangential Initial Stiffness
TPM	Total Productive Maintenance
USB	Universal Serial Bus

List of definitions

Big Data	Large volume of gathered data
Data accessibility	The ease of access and the convenience of creating information
Data gathering	The collection/gathering of data
Digital maturity	The maturity level of the digital advancement of a company
Digital transition state	Transition/transformation of IT-systems into the required future
ERP-system	The cross-organizational system that controls the information flow through various business and functional units in the organization
Horizontal integration	Covers the IT integration of companies within the supply/value chain
Industrial internet of things	Real-time communication of physical objects. Integrates also the value chain stakeholders across companies and thus demands simultaneously both vertical and horizontal integration
Industry 4.0	A collective noun for uprising new technologies, with the vision to connect physical systems such as production environments with virtual models, in short, the computerization of manufacturing
Information system architecture,	IT-systems structure, interfaces between systems, and technology within a company
ISA-95	International standard for developing interfaces between information systems and control systems
MESA-11	ISA-95 based standard indicating the 11 core functions of MES
MES-system	An manufacturing system for monitoring and controlling operations
Real-time data	Data that is passed to the receiver in a short time, aiming to create information about the process as fast as possible
SCADA-network	Interconnected data transferring network that gathers the data from the shop floor and sends this data to the MES-system, this can be a network with or without a control system
Vertical integration	Aligning processes and data within the company and connecting information from Product Development to Manufacturing, Logistics, and Sales for cross-functional collaboration, resulting in a smart manufacturing environment

List of figures

List of tables

1 Introduction

Operational information is important for monitoring, controlling, and improving production processes. Improved performance information from advanced software systems and the experience of process users can lead to better operational decision-making. The key to process improvement is having accurate data to identify the root cause of problems (Industry Directions, 2004). Currently, most operational decision-making is based on the experience of the worker. The development of real-time process monitoring and big data analytics will extend and transform this decision-making into data-supported decision-making. Subsequently, production companies can control their processes which are partly a black box currently. This research focuses on the development of a data gathering system for a case study at a plastic pipeline manufacturer, DYKA Steenwijk. The information system architecture must be upgraded and extended with data gathering to create information that supports decision-making for operators, operational staff, and management. This data gathering must activate operators to react to changing trends of process parameters and will form the foundation for future innovations and process improvements.

The research aims to provide information about the opportunities that evolve when real-time data is used to monitor the processes and gathered big data is used for historical and real-time analyses. The research consists of two parts. The first part is a systematic literature review about the development of a supportive information system architecture and the development of a big data platform. The literature review is generalizable for production companies. The second part is a case study where the information from the literature review is connected to a practical case, this case study provides advice about the development of big data gathering in a manufacturing execution system at the plastic pipeline manufacturer.

This document consists of five chapters. Subsequently, the introduction, the literature review, the context analysis, the problem solution, and the conclusion and discussion.

1.1 Background information

This paragraph contains a description of the company and the research motivation. The research focuses on the extrusion department of DYKA Steenwijk. However, the research can be generalized and used as a theoretical base for other departments or companies.

1.1.1 Company description

The research will be executed at DYKA, a plastic pipe systems manufacturer based in Steenwijk, the Netherlands. The company produces piping systems for residential, and utility buildings and land, road, and water construction. DYKA serves the market with a wide range of products, with applications as inner drainage systems, ventilation duct systems, drainage systems, rainwater systems, systems for filtration of rainwater, water pipeline transport, gas transportation, and electrical installations.

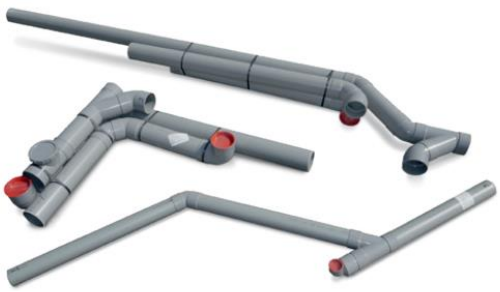


Figure 1: Products of DYKA

These products are produced from plastic materials like polyvinyl chloride (PVC), polypropylene (PP), and polyethylene (PE). For instance, the products stated in figure 1. PVC is the main compound used in production. The raw materials are thermoplastic materials, which means that the material can be recycled to manufacture new products.

DYKA is founded in 1957 by a plumber called Mr. van Dijk and a plastic technologist called Mr. Katers, these two names combined formed the company name DYKA. The company is founded to provide building materials for post-war rebuilding and the building in the newly created Flevo polders in the Netherlands. Most building materials were scarce in the post-war economy. However, plastic was widely available and also lighter in weight than traditional building materials as metal. As a result of this, the company has grown in a plastic manufacturer with multiple manufacturing facilities and over 70 branches in Europe in the past decades, with production locations in the Netherlands, Belgium, France, Germany, Poland, and Hungary. DYKA was sold in 1987 and became part of a joint venture of Tessenderlo Group and D.S.M. (the former Dutch state mines). D.S.M. sold its 50% share to Tessenderlo Group in 1989, which is the current shareholder of DYKA. Tessenderlo Group is a diversified industrial group that focuses on agriculture, valorizing bio-residuals, power plants, and industrial solutions. The parent company of DYKA is based in Brussels and is listed on the Euronext stock exchange in Brussels. DYKA is part of the industrial solutions segment of Tessenderlo under the name DYKA Group, which also contains the companies Nyloplast and JDP.

Figure 2: Extrusion department DYKA Steenwijk

This research focuses on the production site of DYKA in Steenwijk. This manufacturing facility has three departments: the injection molding of fittings, the extrusion of pipes, and the special products and prefab department. The combination of products from these departments creates a complete product portfolio for piping solutions. The case study focuses on the development of data-supported operations at the extrusion department. A selection of machinery of the extrusion department is visualized in figure 2.

1.1.2 Motivation of the research

The motivation of the research can be separated into three parts. The first part is the lack of current availability of production history data. There is not enough data available to create an overview of the production activities; in consequence, operations are partly a black box. Secondly, the company lacks an information system architecture that is designed to monitor the real-time performance of machinery and the quality of products. Thirdly, the data foundation is not developed for connecting and adopting modern technologies. Modern technologies generate data and rely on available data. Information systems should be designed to gather data and connect different data sources. Data gathering and utilization are the gateways toward more innovation and improvements in the future. The following initial problem statement is compiled in collaboration with the manager extrusion and the process innovator of DYKA.

> **Initial problem statement:** How to improve the data gathering strategy and information creation with the view on Industry 4.0 and ISA-95 principles, to make a step-change in process reliability and overall extrusion costs?

The company wants to make a step-change toward a higher industrial level. The process performance has to be improved to obtain a higher performance in reliability and costs. The information about process availability, quality, and costs isn't reliable currently. Industry 4.0 technologies may be interesting but not accessible currently. Interfacing information systems according to ISA-95 standards may provide a solution. Data-driven improvement approaches are the trend in the industry currently. This trend of improving information and communication technologies is utilized by many manufacturers, this tendency presents an opportunity for the creation of manufacturing intelligence systems (Unver, 2013). The research aims to create advice about how the company can monitor and control its operations to obtain a higher performance on production performance and quality performance. Besides this contribution to practice, the contribution to theory is that the research provides information that forms the bridge between the current state of companies and the current research focus of academics. Leaping this gap will make modern research accessible for production companies that are at the starting point of the digital transition.

1.2 Problem identification

The assignment provides an idea of the problem. This is the lack of complete information about processes. Manufacturing knowledge is mostly based on the workers' experience and is not logged in data-gathering systems. This paragraph explains what is known about the problem and which resulting core problem has to be solved.

An overview of problems related to the current data gathering policy and the software systems is obtained by interviewing various stakeholders from the company. The main contributors are the production workers, production management, the IT-department, and the engineering department. The obtained problems are structured in a problem cluster (Heerkens & Van Winden, 2012). This cluster presents the problems and the relation between the problems. The relations are found using the five whys method of Taiichi Ohno, father of the Toyota Production System, and former vice president of Toyota. This method leads to the root cause of problems (Alukal, 2007). Figure 3 illustrates the problem cluster, with at the right side the problem that needs to be solved according to the project assignment and on the left the root problems that cause this problem.

Figure 3: Problem cluster

There are three root problems in this cluster. Each of them is a possible core problem for the research. (confidential)

The core problem

The three problems from the previous paragraph have to be solved to make a step change toward process reliability and overall costs. The problems of the previous paragraph are handling problems. These problems are perceived discrepancies between norm and reality (Heerkens & Van Winden, 2012). A more developed data supervision of operations is the norm and the reality is that the current system does not provide the required information about the process performance and costs. The problem with the biggest impact on the organization is chosen to be solved in the research. The core problem is formulated in the following problem definition:

> **Core problem:** The data gathering in the Manufacturing Execution System is not developed to meet contemporary information requirements for monitoring and controlling production parameters and storing data for historical analyses.

The focus of the research will be on the data gathering of shopfloor data in the MES-layer. The other two root causes are going to be solved by other stakeholders. The manufacturing industry has a growing need for tools that supports decision-making, a major challenge is to access and extract useful data from the factory floor (Farooqui, Bengtsson, Falkman, & Fabian, 2019). The research focuses on that gap, which data is required and what to do with that data.

1.3 Research questions

The research can be separated into two parts, the literature review, and the case study. The literature review is generalizable for production companies and the case study transforms this literature into practical information for the development of information systems and big data usage at DYKA. Both the literature review and the case study has the same research question. The difference is the focus on production companies for the literature review and the focus on DYKA for the case study. The research questions in the literature review and the case study are structured in sub-questions. The research is structured according to the following research questions:

Literature review:

1. How should a production company design its information system architecture to monitor, control, and improve operational performance and be ready for future technologies?
    - Why is data gathering the foundation for executing operational analyses and adopting modern Industry 4.0 technologies?
    - How should an information system architecture be designed to obtain a sufficient digital maturity level for production companies?
    - What are the technical and managerial challenges for the design and development of a real-time big data platform for operational and analytical support?

Case study:

2. How should DYKA Steenwijk design its information system architecture to monitor, control, and improve operational performance and be ready for future technologies?
    - What is the current situation of the extrusion production process and the information system architecture?
    - What are the evolving opportunities and the design requirements for the creation of a supportive manufacturing execution system and a big data platform?

The answers to the research questions form the basis for the start of an information system architecture development and implementation project for a production company, which is in this case DYKA. The relation between the research questions and the problem definition is that the research questions forms are theoretical and practical answers that solve the problem. The research questions form the structure of the research, which is explained in the next sub-chapter.

1.4 Research method and deliverables

The research is structured according to the following research elements/chapters: introduction, literature review, problem context, problem solution, and conclusion and recommendations. This general research method forms the structure of the research. Other research methods are considered, but most methods are focusing on specific problems. For instance, the Design Science Research Methodology, that focuses specifically on the design of products or systems (Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007). This research contains a manufacturing improvement part and a software system design part. A software design method or general design method does not cover all elements of the research. Other methods are used within the general research method, for instance, the managerial problem-solving method (Heerkens & Van Winden, 2012) in the problem identification, the general procedure for conducting a literature review (Templier & Paré, 2015) in the literature review, and the open group architecture framework (TOGAF) to visualize the current and required IT-architecture of manufacturing (The Open Group, 2019). The combination of methods used in the research can be described as a mixed method. This approach is used for the development of information systems that combine both qualitative and quantitative techniques (Wu, 2011). Different approaches are required because the project needs input from the users (qualitative information) and it needs information from data handling (quantitative information).

The literature review is generalizable for production companies. It focuses on the development of information system architectures to create the data foundation for analytics and future technologies. This information is made practical in the case study. The case study consists of the problem context and the problem solution. The problem context provides all relevant information about the extrusion process and the information system architecture of the company. This information is required to construct the problem solution of the case study. The problem solution provides insights into the potential of developing and extending the information system architecture. This future state of data gathering and connecting is visualized in a big data platform prototype made in the BI-tool Tableau.

The insights from the big data platform and the technical, and managerial challenges for developing data gathering in a manufacturing execution system are the deliverables of the research. The research does not focus on creating a specific design for an information system and does not implement an information system. The research aims to provide a body of knowledge about gathering big data and connecting data from different systems in a BI-tool, this will form the theoretical foundation for the initialization of a manufacturing execution system development project at the case study company DYKA.

2. Literature Review

This chapter presents a general literature review on the development of information systems at production companies. It aims to provide information about possibilities for operational performance control and adopting modern technologies. This review is separated from the case study to create information that can be generalized within the manufacturing industry. The literature review aims to answer the following research question.

> How should a production company design its information system architecture to monitor, control, and improve operational performance and be ready for future technologies?

This research question is partitioned into three research questions, which are stated and answered in subchapters 2.2 to 2.4. This body of knowledge forms the basis for the case study and can be used as a theoretical foundation for case studies at other production companies.

The manufacturing industry has a growing need for tools that supports decision-making, a major challenge is to access and extract useful data from the factory floor (Farooqui, Bengtsson, Falkman, & Fabian, 2019). The research focuses on that gap, how and why data gathering in information systems will support decision-making processes for supporting and improving operations. The literature review provides information about technologies that will be accessible if information systems have a particular maturity level. In addition to this, the research will provide information about important design decisions and considerations that will lead to a useful and valuable information system. The research will start broadly, in which future technologies are elaborated. After this, the research is narrowed to the first steps and design decisions that companies should execute before they can implement modern technologies.

The literature review is structured according to the sub-research questions of the literature review, stated in chapter 1.3. Firstly, the literature search method (Ch. 2.1) is described, this subchapter contains the literature search strategy. The second to the fourth subchapters (Ch. 2.2-2.4) answer the research questions of the literature review. Lastly, the conclusion and discussion (Ch. 2.5) are provided. Subsequently, a second discussion paragraph connects the literature review to the case study (Ch. 2.5.2).

2.1 Literature search method

The literature review is conducted to create an in-depth overview of the information that is required to make decisions in information system architecture design. The objectives of a literature review are: helping a researcher acquire an understanding of the topic, providing an overview about what research is already done, providing insight into how the topic has been researched, and presenting what the key issues of the topic are (Hart, 1999).

Those objectives can be achieved if most of the relevant literature is included in the research. A literature review method can guide this process. The literature review is conducted according to the general procedure for conducting a literature review; the guidelines of this method specified for information systems research are used in this research (Templier & Paré, 2015). This method consists of six research steps, visualized in figure 4.

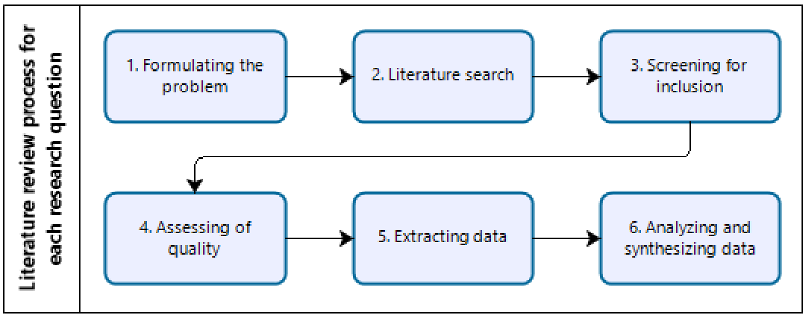


Figure 4: The literature review process for each research question

This process will be repeated/executed for each research question in the literature search. The research questions are stated in chapter 1.3 and at the beginning of each subchapter in this literature review. A description of the activities in each phase is stated in table 1.

Step	Title	Activity
1	Formulating the problem	Creating research questions to structure the literature review
2	Literature search	A literature search in the literature database (Scopus) using keywords. Keywords are selected by combining keywords from the research questions, checking these keywords on synonyms, and extending the keywords with other keywords found in the searched literature.
3	Screening for inclusion	Screening on title, year, and abstract. Focus on recent articles and improving production performance
4	Assessing quality	A quick scan of the article, focus on information quality, especially focusing on the relevance of the information to the topics of the literature review
5	Extracting data	Reading/extraction of information from selected articles. Extraction of information from articles that answers a specific research question of the literature review
6	Analyzing and synthesizing data	Structuring and reporting the information

Table 1: Actions within the systematic literature search

The systematic literature search is stated in table 2. This table contains the used keywords and the number of articles found after each step. The remaining articles after step four are part of the literature research. The search strategy and explanation of keyword choice are stated in appendix A. Appendix B contains the list of articles and summary data of the selected articles.

Sub-chapter	Keywords search terms	2. Literature search	3. Screening for inclusion	4. Assessing quality
2.2	TITLE-ABS-KEY ("data gathering" AND "manufacturing")	194	28	6 (used in all chapters)
2.2	TITLE ("data collection" AND "manufacturing")	39	7	1
2.3	TITLE-ABS-KEY ("industry 4.0" AND "maturity model")	124	5	3
2.3	TITLE-ABS-KEY("isa-95" and "industry 4.0")	22	4	1
2.4	TITLE-ABS-KEY ("real-time data" AND monitor* AND manufacturing)	262	9	1
2.4	TITLE-ABS-KEY("real-time data" and visuali* and manufacturing)	44	16	2
2.5	TITLE ("overall equipment effectiveness" AND performance)	39	3	1

Table 2: Number of articles selected in each process step

The number of articles included in the research is extended with forward and backward tracing from the articles selected in the systematic literature search. Besides the selected articles and forward, and backward tracing, other articles are added to the literature review if the information from those articles contributes to a more improved answer to the research question.

2.2 Manufacturing data gathering and Industry 4.0

Production companies have the objective to improve their production processes continuously. The input of operational decision-making is the experience of the users and information about the processes. An improvement in either experience of the operator or provided process information can optimize the consequence of the decision. New technologies can support the creation of information, manufacturing systems can be made more intelligent by the use of new technologies.

This is not a recent trend in the manufacturing industry. Three decades ago, Meieran (1993) indicated that new technologies and applications may directly improve resource utilization within the factory and the new technologies can be used to capture, preserve, and apply knowledge to facilitate improvement in the decision-making processes. This gathering of information is still trending in the industry and is becoming more trending because of new progressive technologies aggregated under the name Industry 4.0. In particular, the collection and processing of data from different sources into information is an important feature of industry 4.0 visions, these sources will make the decision-making process more efficient (Obitko & Jirkovský, 2015). This subchapter explains why data gathering is important for the creation of information and the adoption of new technologies, the following question will be answered:

> Why is data gathering the foundation for executing operational analyses and adopting modern Industry 4.0 technologies?

A company can automate the gathering of data in order to use the data for analyses and to monitor, detect, correct, and fix problems; because deviating trends in the data might become costly if they continue to be undetected for very long (Meieran, 1993). The implementation of supporting data gathering systems is a major challenge for companies, this subchapter focuses on why a company should implement a data gathering system. The following paragraphs focus on the importance of gathering data, uprising new technologies, and the combination of these two subjects.

2.2.1 Manufacturing data gathering

Providing information about the condition of processes can support the working staff in improving the performance. This statement is easily stated and is a possible valid statement. Still, an overall system that supports operators and engineers better understand their processes is rare in the manufacturing industry; it is difficult to debug problems and improve performance in a reliable and verifiable manner because of the lack of an IT-system (Farooqui, Bengtsson, Falkman, & Fabian, 2019). The potential of gathering real-time data is to significantly reduce the time spent identifying and correcting operational issues, for instance, preventing the occurrence of unplanned downtime and predicting the time of the optimal maintenance interval (Saabye, Kristensen, & Wæhrens, 2020). The creation of such a system is promising but the creation is a challenge. This is because factory floors consist of a diverse mix of technologies, connecting diverse technology to one system is a complex problem to solve (Farooqui, Bengtsson, Falkman, & Fabian, 2019). The data gathering system must be designed to fulfill the information requirements composed by the company before the start of the software development project. For instance, one issue that has to be solved is the handling of the increasing amount of data that needs to be processed and analyzed for rapid decision-making that leads to more improved productivity (Obitko & Jirkovský, 2015). A data gathering system may improve productivity but one important factor is the gap in skills of users in contrast with the creators of the new technologies (Saturno, Pertel, Deschamps, & de Freitas Rocha Loures, 2017). The system must be supportive and understandable for its users. Promising features of such a system are, real-time statistics, real-time reporting, clustering machines, benchmarking of key performance indicators (KPI's), predictive maintenance, and other pattern recognizing (Obitko & Jirkovský, 2015).

A logical option, when embarking on a transformation towards the digitalization of manufacturing systems, is to acquire new industry 4.0 technologies; Industry 4.0 technologies support operations by providing analytics that monitor and improve the performance of machines by the use of real-time data (Saabye, Kristensen, & Wæhrens, 2020). Industry 4.0 is an effort led by German engineers with the vision to connect physical systems such as production environments with virtual models, in short, the computerization of manufacturing (Obitko & Jirkovský, 2015). A key aspect of effective implementation of new industry 4.0 technologies is the requirement of a supportive learning environment and supportive leadership; this is required because new technologies do not automatically result in the desired changes to operators' behaviors, they must learn to utilize these technologies (Saabye, Kristensen, & Wæhrens, 2020). Thus, the users are important stakeholders in both the creation and the implementation of new technologies. The connection of more physical systems will be a major challenge for production companies but will bring many advantages including real-time data, real-time analysis, historical analysis, and Industry 4.0 technologies.

2.2.2 Fundamentals of industry 4.0

Industry 4.0 is the fourth technological improvement since the industrial revolution. The first technological innovation was the creation of the steam engine in the nineteenth century; the second was electrification which led to mass production in the early part of the twentieth century; the third advancement was the automation of industry in the seventies of the twentieth century (Rüßmann, et al., 2015). The key feature before implementing Industry 4.0 technologies is the creation of a big data acquisition platform containing data from horizontal integration through value networks and vertical integration through networked manufacturing systems (Oesterreich & Teuteberg, 2016).

Industry 4.0 technologies can be adopted when information systems are developed to a particular maturity level, more about maturity is stated in the next subchapter. Industry 4.0 is a collective noun for the uprising new technologies and has many interpretations in the literature. Most sources refer to the nine fundamental pillars of 4.0 Industry 4.0 composed by the Boston Consulting Group (BCG), specifically by researchers based in Germany and Austria under the lead of M. Rüßmann and published in 2015. The articles of (Erboz, 2017), (Moeuf, Pellerin, Lamouri, Tamayo-Giraldo, & Barbaray, 2017), and (Schumacher, Nemeth, & Shin, 2018) from this literature review refer to BCG.

Nine fundamental pillars of industry 4.0

The fundamental technologies are related to each other. The connecting element between the technologies is the exchange of data. Sensors, machines, workpieces, and IT-systems will be connected within the value chain of a company; these connected systems, also referred to as cyber-physical systems, can interact using Internet-based protocols and analyses data to predict failure, configure themselves, and adapt to changes (Rüßmann, et al., 2015). The different state-of-the-art technologies can be grouped into nine fundamental pillars of industry 4.0, which are visualized in figure 5 and explained below the figure.

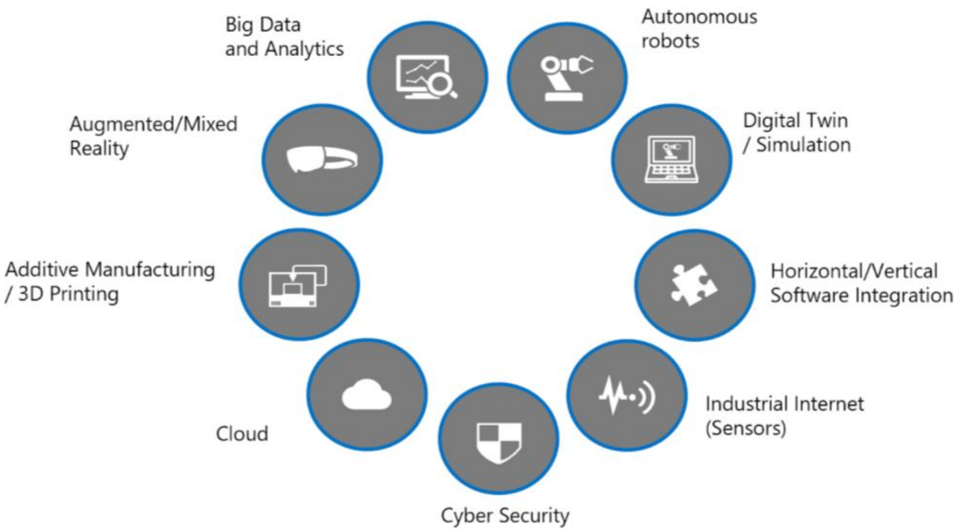


Figure 5: The nine Industry 4.0 technology pillar (Foster, et al., 2018)

1. *Big data and analytics*, analytics based on large data set are trending in the manufacturing industry, where it is used to optimize production quality, save energy, improve equipment availability, and support real-time decision-making (Rüßmann, et al., 2015).

2. *Autonomous robots*, robots with embedded sensor technologies are becoming more flexible, communicative, and cooperative (Michniewicz & Reinhart, 2014).
3. *Simulation*, simulation is used to generate operation schedules online and for analyzing and modifying current production systems (Moeuf, Pellerin, Lamouri, Tamayo-Giraldo, & Barbaray, 2017).
4. *Horizontal and vertical system integration*, vertical integration refers to flexible and reconfigurable systems and machinery inside the factory and the extent of fully integrative systems to achieve agility (Erboz, 2017). Automation and integration of equipment are essential to optimize and improve production processes (Saturno, Pertel, Deschamps, & de Freitas Rocha Loures, 2017). Horizontal integration covers the integration of partners and other companies within the supply chain (Erboz, 2017). An enterprise transforms into an industrial network by vertical and horizontal integration. This industrial network collects big data to optimize system performance, creating a smart factory (Erboz, 2017).
5. *Industrial internet of things* (IoT) (figure 6), real-time communication of physical objects can result in the monitoring of various products and system states in real-time and facilitates the decentralization of decision-making (Moeuf, Pellerin, Lamouri, Tamayo-Giraldo, & Barbaray, 2017). More devices, products, and IT-systems will be connected using standard technologies, this allows devices to communicate with one another and enables analytics and real-time responses (Rüßmann, et al., 2015). The objective of IoT is to connect the Internet by collecting data from physical systems (Erboz, 2017). IoT integrates also the value chain stakeholders across companies and thus demands simultaneously both vertical and horizontal integration (Lesjak, Druml, Matischek, & Ruprechter, 2016).

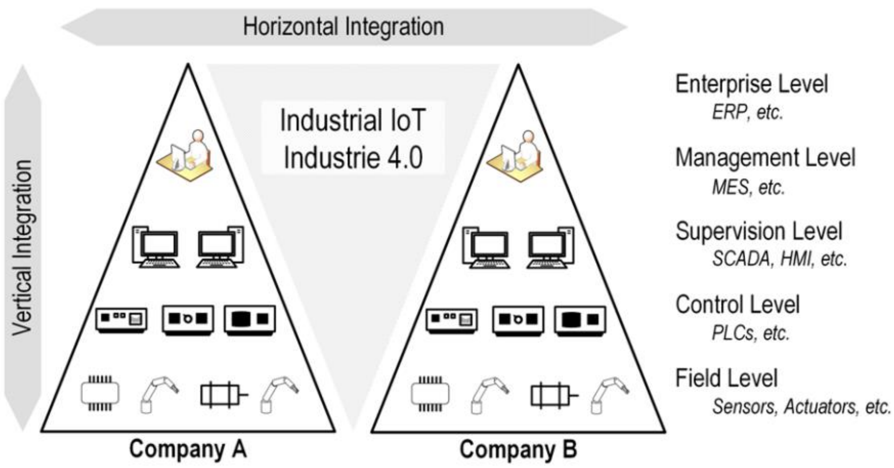


Figure 6: Vertical integration, horizontal integration, and industrial internet of things (Lesjak, Druml, Matischek, & Ruprechter, 2016)

6. *Cybersecurity*, a new problem that arises with the creation of increasingly connected systems and the use of standard communications protocols, the internet of things must be protected. Cybersecurity needs to protect critical industrial systems and manufacturing lines from increasing dramatically cyber threats (Rüßmann, et al., 2015).

7. *Cloud*, cloud communication, and exchange of information result in a network of connected networks in real-time to ensure that data and applications are available/accessible everywhere (Moeuf, Pellerin, Lamouri, Tamayo-Giraldo, & Barbaray, 2017).
8. *Additives manufacturing/3D printing*, additive manufacturing is regarded as the process of making products from 3D models; this technology completes products by layer upon layer, making process activities as milling and machining redundant and enables the production of customized products designed and customized by customers (Erboz, 2017).
9. *Augmented reality*, this futuristic technology uses data to simulate an environment containing real and simulated objects that can be used to visualize and enhance designs and manufacturing processes (Moeuf, Pellerin, Lamouri, Tamayo-Giraldo, & Barbaray, 2017).

Those groups of modern technologies are promising for the manufacturing industry. However, the realization is a difficult process. The next paragraph outlines the potential and applicability of new technologies.

2.2.3 Potential of data gathering and Industry 4.0

Data gathering and Industry 4.0 technologies have a strong relationship. Most industry 4.0 technologies rely on gathered and processed data. On the other hand, industry 4.0 enables the possibility to expand gathering and analyzing data across machines because of the new technologies in those machines. In short, modern technologies create data, and data can be used to improve those technologies and operational performance.

A cyber-physical production system has a major impact on industrial systems. A developed production IT-system can support projects throughout the entire company using the gathered data, for instance (Lesjak, Druml, Matischek, & Ruprechter, 2016):

- Better equipment can be engineered by leveraging operational performance data.
- Equipment operations can be optimized.
- Remote control and management of equipment are made possible.
- Service activities can be predicted and triggered.
- Remote diagnostics replace field service activities.
- Field service can be optimized.
- Information and data-driven services can be provided, e.g. to customers or suppliers.

The operational data can be used to create real-time monitoring platforms that support operations. Current Industry 4.0 initiatives in small and medium-sized enterprises (SMEs) focus on monitoring industrial processes, the data is only in a few cases used to determine warning thresholds, support decision-making, and exploit gathered data to optimize operations in real-time (Moeuf, Pellerin, Lamouri, Tamayo-Giraldo, & Barbaray, 2017). This potential will emerge if a company develops a monitoring system that is developed for these purposes.

Data exchange between machines and IT-systems will create an environment/platform that has the possibility of gathering, processing, and visualizing data. The interaction between IT-systems and machinery is possible because most software of both machinery and IT-systems interacts with each other using standard Internet-based protocols, which enables the opportunities in, for instance, analyzing data to predict failure, configure themselves, and adapt to changes (Rüßmann, et al., 2015). The network of machines and IT-systems requires flexible and open protocols of communication, integrating all the components of architecture enables a company to access all relevant data in real-time (Saturno, Pertel, Deschamps, & de Freitas Rocha Loures, 2017).

Conclusion

Data gathering is the key to improve processes and adopting modern technologies because analysis and new technologies rely on performance information. Architectures must be developed to apply an intelligent manufacturing concept, functions/applications within the architecture must be developed to support the technologies (Saturno, Pertel, Deschamps, & de Freitas Rocha Loures, 2017). Vertical and horizontal integration forms the basis to create an industrial internet of things (IIoT), which is needed to create a smart manufacturing environment. New technologies can be connected to this IIoT-network to create an interconnected manufacturing facility. In conclusion, horizontal integration and more importantly vertical integration forms the foundation for manufacturing improvement and industry 4.0 technologies.

2.3 Digital maturity for manufacturing companies

Vertical and horizontal integration is required to open the way for advanced analyses and new technologies. But the question hereby is, which digital maturity level is required to gain access to these promising business improvement approaches? This subchapter provides information about digital maturity levels of production companies and the information system architecture to reach this maturity. The following research question is answered in this subchapter:

> How should an information system architecture be designed to obtain a sufficient digital maturity level for production companies?

Maturity models are important to help manufacturing companies assessing their processes and to figure out when they are ready for the digital transformation, it also helps in developing their transformation roadmap (de Carolis, Macchi, Negri, & Terzi, 2017). Specifically, digital maturity can be seen as the advancement in Industry 4.0’s basic concepts such as vertical and horizontal integration in manufacturing systems and value chains (Schumacher, Erol, & Sihn, 2016). This subchapter is separated into three parts, which narrows the research field from the broad general information from the previous subchapter to specific information about the design of an IT-system architecture. The first topic is the relation between digital maturity and industry 4.0. The second subject provides information about how this maturity can be reached by providing information about information system architectures, and at last, specific information about the implementation of manufacturing execution systems using the ISA-95 protocol.

2.3.1 Digital maturity toward industry 4.0

The digital maturity of a company can be assessed to provide an overview of the current state of advancement of the company’s IT-environment (Schumacher, Erol, & Sihn, 2016). There are scientifically grounded approaches and practical approaches to assessing this maturity (Schumacher, Nemeth, & Shin, 2018). Scientific grounded maturity models focus on quantifying the maturity level of a company. Most scientific maturity models are inspired by the Capability Maturity Model

Integration (CMMI) framework or the ISO/IEC 15504 Simple Protocol for Independent Computing Environments (SPICE) (von Wangenheim, Hauck, Salviano, & von Wangenheim, 2010). Information about the objectives and characteristics of these two frameworks is stated in appendix C.

Maturity models focusing on industry 4.0 objectives are less detailed than the detailed quantifiable scientific assessment tools (Schumacher, Nemeth, & Shin, 2018). Industry 4.0 maturity models have more the format of a practical roadmap. A company has to run through a couple of states to obtain the required maturity level. An article by Schumacher, Erol, and Sihn (2016) has provided a general literature review on industry 4.0 maturity models. The model of the PricewaterhouseCoopers consultancy (PWC) company is selected from that review to be elaborated in this research. This model is chosen because of its practical usability and focus on digital operations development. Other models weren't chosen because those were more difficult to apply and understand because of extensive calculations, the objective of this literature review is to inform the production company with applicable information. The model provides four states in which a company can be situated. The PWC maturity model contains the following four stages, stated in figure 7 (PWC, Geissbauer, Vedso, & Schrauf, 2016). These four stages are explained in seven dimensions that explain the advancement of IT-systems in a stage. Figures about the seven maturity dimensions, the efficiency gains, and the emerging opportunities are stated in appendix D.

Figure 7: Stages of digital maturity (PWC, Geissbauer, Vedso, & Schrauf, 2016)

Stage four of this model is difficult to reach, a company must first go through the other stages before this level has been reached. The implementation of IT capabilities and driving the cultural shift will take years, but companies should move the focus beyond technical details and should consider what impact new applications could have on the value chain and relationships with customers and other stakeholders of the company (PWC, Geissbauer, Vedso, & Schrauf, 2016).

2.3.2 Vertical Integration and Information system architecture

The first step of moving toward the digital champion stage of the maturity model is to move from a digital novice to a vertical integrator. Vertical integration of IT-systems is about aligning processes and data within the company and connecting information from product development to manufacturing, logistics, and sales for cross-functional collaboration, resulting in a smart manufacturing environment (Oesterreich & Teuteberg, 2016). The smart factory concept can be created by the integration of the Manufacturing Execution System (MES) with the shop-floor layer on the one side and the Enterprise Resource Planning (ERP) layer on the other side; this results in insight into the manufacturing execution in real-time (Mladineo, Veza, Gjeldum, Crnjac, Aljinovic, & Basic, 2019). For instance, the vertical integration from sensors to obtain real-time production planning for better machine utilization and faster throughput times (PWC, Geissbauer, Vedso, & Schrauf, 2016). This shopfloor-layer, MES, and ERP alignment create an IT-architecture that communicates with each other intending to increase the amount of information at each layer.

Information system architecture

The traditional theory about controlling operations is about decision-making in three different levels of control, strategical, tactical, and operational (Borisovich & Htun, 2019). Each level has different problems and requires different software-support/information to solve those problems. All levels of control are strongly related to each other, information from one level is needed to solve problems at another level. Automation of systems supports companies in aligning these vertical levels. This automation and connection of applications result in the information system architecture of a company. An information system architecture for a manufacturing company can be divided into systems that are designed to solve specific problems. The system for strategic decisions is the enterprise resource planning (ERP) system, the application for tactical decisions is the manufacturing execution system (MES), and the platform for operational decisions is the supervisory control and data acquisition (SCADA) system (Borisovich & Htun, 2019). Those three information system layers are required to connect machinery information (operational data) to the corporate information (strategical data), the intersection of features of the three levels is stated in figure 8.

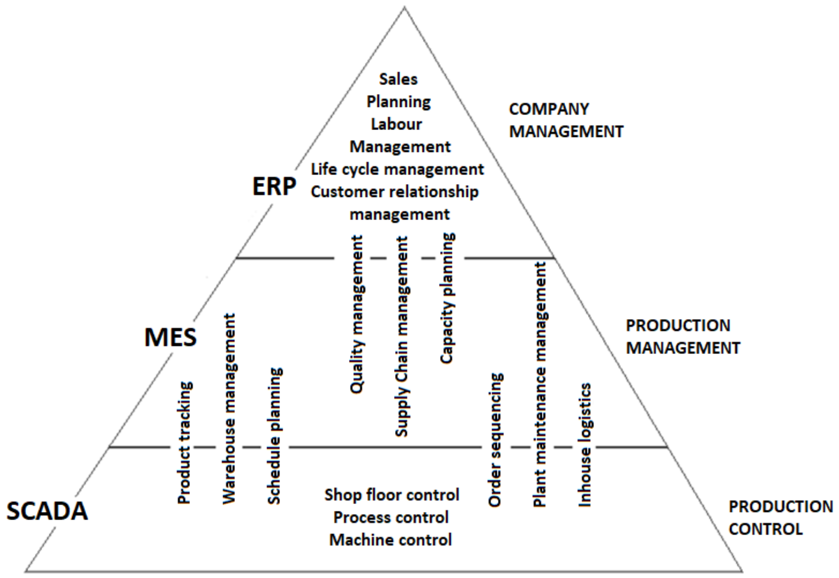


Figure 8: Intersections between ERP, MES, and SCADA (Modrák & Mandul'ák, 2009)

The ERP-system is the cross-organizational system that controls the information flow through various business and functional units in the organization (Beric, Stefanovic, Lalic, & Cosic, 2018), briefly the business' planning and logistics system. All information about the supply chain of the organization is described in this information system, the corporate control level. The manufacturing process is one part of this supply chain. The MES-system controls that part of the supply chain. The MES-system receives production order and bill of material information from the ERP-system and sends material use, order status, quality data, and process data back to the ERP (Beric, Stefanovic, Lalic, & Cosic, 2018). The MES-system monitors and collects information about the entire production cycle and provides real-time information to support decision-making (Beric, Stefanovic, Lalic, & Cosic, 2018). The SCADA-network is the machine layer, which consists of the human-machine interfaces (HMI) and process control systems on the shop floor. The SCADA-network is a process control system that collects the data in real-time and transmits the data to the MES-system (Borisovich & Htun, 2019).

2.3.3 Manufacturing execution systems and ISA-95

The previous paragraph illustrates the general information system architecture for production companies. This paragraph focuses on specific elements of that architecture. This paragraph aims to provide specific information about the capabilities and advantages of an integrated MES-system and how this system layer can be integrated using an interface protocol.

MES has two purposes, the identification and execution of the optimal operational planning and managing the bottom-up data flow (D'Antonio, Macheda, Bedolla, & Chiaber, 2017). This production management IT-system consists of multiple modules that gather, process, and visualize data. An MES-system consists of the following key areas: production management, maintenance management, inventory management, quality management, operational performance management (Telukdarie, 2016). The integration of these elements into one system results in a platform that contains a various range of information. This platform is in charge of gathering data from the shopfloor, analyzing this through proper mathematical techniques, and extracting the information in real-time to provide an exhaustive picture of the current state of the process (D'Antonio, Macheda, Bedolla, & Chiaber, 2017); the advantage of this is that analysis can be performed in real-time, this results in the ability to improve real-time decision making to control the process. The MES-system connects information between multiple software applications and machinery, extending from a company’s Enterprise Resource Planning (ERP) to shopfloor quality applications. (PWC, Geissbauer, Vedso, & Schrauf, 2016). Developing the MES-system and other systems into a sufficient maturity level can play a strategic role in supporting Industry 4.0 technologies; focusing on the future, the data gathering platform transforms data into information, which can feed simulation models and can be used in data mining models to optimize the business’s performance (D'Antonio, Macheda, Bedolla, & Chiaber, 2017). Challenges in the process of aligning IT-systems and developing smart factory environments are the connection and interfacing of IT-systems and machinery.

The standard for interfacing information system architectures is the ISA-95 standard. Other standards describes also the functionalities of an MES-system, but the interfacing focusing ISA-95 is mostly the basis for these models (Fucheng, Haibo, & Bin, 2015). For instance, another model from the literature is more focused on the functionalities of an MES-system, the MESA-11 model. This model explains MES in eleven functionalities of MES: operational/detailed sequencing, resource allocation and status, dispatching production unit, performance analysis, maintenance management, process management, quality management, data collection/acquisition, product tracking and genealogy, labor management, document control (Fucheng, Haibo, & Bin, 2015). An extensive explanation of the MESA-11 functions is added to appendix E. The ISA-95 standard has a broader view, it describes the capabilities of MES and the interfacing with the entire information system architecture.

ISA-95

The international society of automation initiated the ISA-95 standard in 1995. This standard provides definitions about the functionalities of MES and how data can be exchanged between the enterprise resource planning system (ERP) and control systems (MES or other control systems) (Kannan, Suri, Cadavid, Barosan, van den Brand, & Alferez, 2017). The ISA-95 standard aims to reduce the risks, costs, and errors during the implementation of ERP and MES (Modrák & Mandul'ák, 2009). The model describes the interface between five levels of application layers: the physical process (level 0);

the intelligent devices and sensors (level 1); the process control, human-machine-interface (HMI), and supervisory control and data acquisition (SCADA) functions (level 2); the operations management functions (MES) (level 3); and enterprise planning and logistics functions (ERP) (level 4) (Gifford & Daff, 2020). These layers are visualized in figure 9, level zero and one are presented as one layer in that figure. The figure is in line with the information system structure of figure 8 in chapter 2.3.2. Different parts of the ISA-95 are describing the interface decisions between different layers, for example, ISA-95.01 is describing the interface between MES and ERP. These parts provide information structured in nine subjects, which are visualized in figure 10.

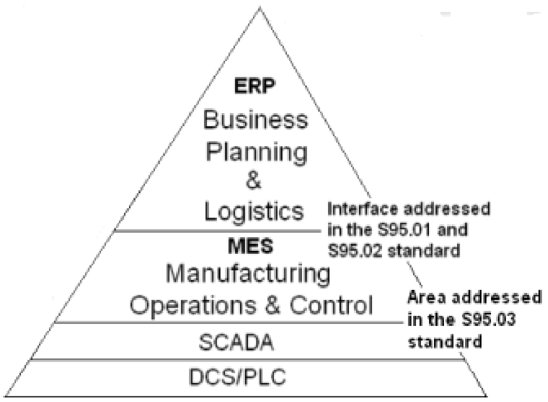


Figure 9: Interface levels of the ISA-95 standard (Modrák & Mandul'ák, 2009)

Figure 10: Subjects of ISA-95 (Gifford & Daff, 2020)

The integration between levels three and four, described in ISA-95.03, is the connection between the networked machinery and the MES-system. The interface between machines and information systems can be realized using standards as Object Linking and Embedding for Process Control Unified Architecture (OPC UA), this machine-to-machine communication protocol makes industrial automation possible (Prades, Romero, Estruch, García-Dominguez, & Serrano, 2013). OPC UA is the standard for the connection of machinery into a SCADA-network; the standard was released in 2008 to replace the old OPC standard, the upgraded standard focuses on accessing automation and control of systems by creating internet protocol (IP) networks (Akerman, 2018). The objective of this new standard is to create an industrial internet of things factory environment. Companies should align their information system architecture with an industrial standard such as the ISA-95 or the older ISA-88 because industry 4.0 technologies are built on the ideas of these standards (Kannan, Suri, Cadavid, Barosan, van den Brand, & Alferez, 2017). Designing the MES-system and the information system architecture according to industrial standards as the ISA-95 will prepare companies to acquire modern technologies and to gather information about their performance.

Conclusion

The digital maturity of a company can be assessed and assigned to one of the following four industry 4.0 digital maturity levels: digital novice, vertical integrator, horizontal collaborator, and digital champion. A company should design its information system architecture according to the ISA-95 standard to obtain the maturity level of a vertical integrator. This architecture is a structure of connected systems that connects the operational shopfloor with strategic top floor management by the connection of machines (SCADA-network), manufacturing operations monitoring and control systems (MES-system), and business planning and logistics systems (ERP-system). Information sharing between those layers results in real-time data supervision of operations and the creation of constructive information for decision-making through the entire organization.

2.4 Developing a real-time big data platform

The previous subchapter explained why vertical integration the first step is toward digital maturity. This first step creates data and information about processes and manufacturing performance. Factories with connected systems are more efficient, productive, and smarter than their non-connected competitors; because they can improve decision-making due to the use of more specific process information (He & Wang, 2018). This can be done in a business intelligence tool (BI-tool), which can visualize the gathered and stored data from the IT-systems. This subchapter analyses literature about the potential of the use of real-time data and the development of a big data platform. The following research question will be answered in this subchapter.

> What are the technical and managerial challenges for the design and development of a real-time big data platform for operational and analytical support?

The integration from shop floor data (SCADA) to top floor data (ERP) result in high visibility of a company's performance. Such an advanced real-time data system is beneficial because it generates accurate automated data flows, reproducible KPI's and performance reports (Telukdarie, 2016). The gathered data can be used for improvement projects and operational decision-making. For instance, sending event notifications when the process is not performing as planned, a very powerful way to communicate event information (Gifford & Daff, 2020).

This subchapter is structured into six paragraphs. Firstly, the potential advantages and challenges of real-time data are elaborated. The second paragraph examines the subject in-depth, providing relevant decisions for designing the real-time big data platform. The third paragraph provides information about the involvement of the human factor in a software implementation project. The fourth paragraph focuses on how manufacturing performance can be improved and measured. The fifth paragraph focuses on one performance indicator specifically, the overall equipment effectiveness (OEE) performance indicator. The last paragraph explains how a company should develop and maintain its reporting portfolio in an information system.

2.4.1 Real-time data potential

Acquiring industry 4.0 technologies and the digital transformation is the future objective of many businesses, this will result in smart manufacturing, a higher performance level, and thereby new competitive advantages (Clausen, Mathiasen, & Nielsen, 2020); but the question hereby is: to what extent companies has adopted this type of digital transition? Only a few companies have already explored the benefits of working with such digital systems that connect machines to systems (Clausen, Mathiasen, & Nielsen, 2020). The potential of such systems must be revealed to facilitate motivation to start a project for gathering and using manufacturing data. One focus of Industry 4.0 smart manufacturing is to create manufacturing intelligence using real-time data to support accurate and timely decision-making (He & Wang, 2018). Gathering manufacturing data can be seen as shopfloor management to achieve high manufacturing efficiency, low manufacturing costs, high product quality, and high employee satisfaction (Clausen, Mathiasen, & Nielsen, 2020). In addition to this, high-level optimization and information extracted from massive data can also be used for the optimization of planning and scheduling (Qin, 2014). Thus, an improvement in information systems can result in manufacturing intelligence and an increase in performance, but why have not all manufacturing companies implemented such a system?

The realization of a real-time data acquisition platform is a difficult process, it has influencing forces for and against this change in manufacturing approach, from experience-driven to experience and data-driven production (Clausen, Mathiasen, & Nielsen, 2020). Examples of these forces for and against digital transformation are stated in table 3.

Influencing forces for the digital transition of the shop floor layer	**Influencing forces against the digital transition of the shop floor layer**
Real-time and reliable data (no “hidden factory” syndrome)	High investment
Improved data foundation	Habitual mindset/procedures
Enhanced data accessibility	Data blindness
Enhanced data transparency	Time-consuming
Early problem detection	Unsuitable IT architecture
Real-time/big data enabling efficient decision making	Digital immature technologies
Intelligent technologies for decision-making	Higher vulnerability if IT-systems fail
Digitization is a prerequisite for competitiveness	Poor data quality in the company
	Low commitment for changes at SFM layer
	Managers deprioritizing the digital transition
	Low know-how of the opportunities

Table 3: Forces for and against the digital transition (Clausen, Mathiasen, & Nielsen, 2020)

Both forces for and against are important to notice, those challenges must be solved to come up with an effective information system. Real-time data can be used for real-time production monitoring and industry 4.0 technologies such as big data analytics and the industrial internet of things. Most of these technologies are not accessible for companies that are in the novice state of digital transformation. Therefore, the conclusion of chapter 2.2 was to start with vertical and horizontal integration. Resulting in the conclusion of chapter 2.3 that vertical integration is the first step of improving digital maturity. But what to do first if such a system is not developed to meet the contemporary data requirements? Zhang and Wang (2020) proposed the first three steps of creating an industrial internet of things data acquisition platform:

Step 1: The overall implementation of a data acquisition platform
Step 2: Real-time data gathering and monitoring
Step 3: Getting an alarm list that monitors the production parameters

This subchapter (Ch. 2.4) focuses on the implementation of a data gathering system, where chapters 2.1 to 2.3 more focusses on the theoretical body for the architecture. So the previous knowledge is required to understand why these implementation steps are important. Firstly, the implementation of the data acquisition system contains the connection of different information systems layers, explained in chapter 2.3.2 and according to standards explained in chapter 2.3.3. In the second place the real-time data gathering, the industrial big data acquisition platform must provide real-time data to facilitate monitoring parameters in real-time and storing the historical data to make data available for analysis. Thirdly, the alarm list, the platform must recognize abnormal data and machine failures. Using massive data with monitoring and controlling error tolerances is a major improvement step, the goal is to turn data into useful knowledge and to support effective decision-making and optimization (Qin, 2014). In conclusion, digital shop floor management results in effective monitoring, diagnosing, and prognosticating of activities on the shop floor (Clausen, Mathiasen, & Nielsen, 2020).

2.4.2 Technical design challenges and decisions MES/big data platform

The realization of a big data platform that monitors production is a challenging project. This data gathering in the MES-system is a difficult process for companies that traditionally rely on the workers' experience rather than data-supported operations. This change toward smart manufacturing is a paradigm that requires strategic innovation of existing manufacturing techniques by integrating people, technology, and information; it is the embodiment of a fusion between advanced IT and manufacturing techniques (Lee, Kim, & Kim, 2017). Gathered and stored big data can be used for smart analytics. An indication of smart statistical analytics is provided in appendix F. Those analyses are not part of this research. The focus of this research is to provide advice about the development of the big data platform, which is the foundation for analytics in the future. As stated in the literature, data gathering must be performed first to create big data (Lee, Kim, & Kim, 2017).

This paragraph provides some practical design decisions that will be encountered during an implementation project. Starting with a project management method. Literature provides multiple approaches, most methods contain general elements of project management and some practical elements that focus on decision-making for a big data platform. The model of (Abdel-Fattah, Helmy, & Hassan, 2019) is selected to form the guide for an implementation project. This is because of its clear project structure, important decisions that are relevant for creating a big data platform, and the application for specific production companies. The method is visualized in figure 11.

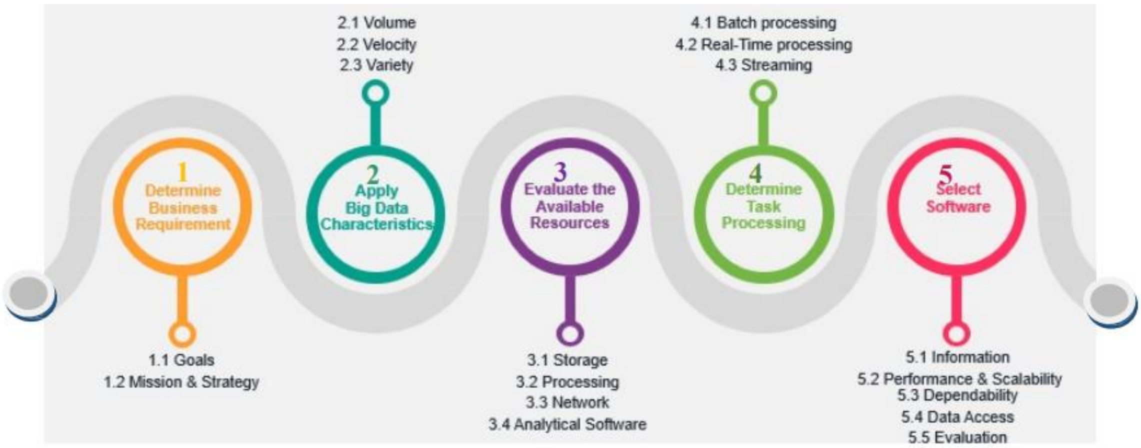


Figure 11: Project management method big data decisions (Abdel-Fattah, Helmy, & Hassan, 2019)

Step two of this method is about making decisions about the data characteristics. These characteristics must be in line with the business requirement goals of the company. There are many variations on these three V's, several literature sources extent these three decisions with more V's or other important decisions. This literature research presents 4V's to characterize big data, these 4V's are: volume (the size/scale of the number of data points), variety (the form/format of the data), velocity (the rate/interval of the data being produced), and veracity (the uncertainty/reliability of the data) (He & Wang, 2018). More information and citations from the literature about these characteristics of big data can be found in appendix G.

The implementation team will encounter many challenges during the project. Some are already stated in figure 11. The article of Chen, Mao, and Yunhao (2014) provides a list of challenges that has to be solved during a project. The practical decisions, challenges, and considerations for creating a value-added big data platform are (Chen, Mao, & Yunhao, 2014):

- Data representation: many datasets have certain levels of heterogeneity in type, structure, semantics, organization, granularity, and accessibility. Data representation aims to make data more meaningful for computer analysis and user interpretation.
- Redundancy reduction and data compression: generally, there is a high level of redundancy in datasets. Redundancy, reduction, and data compression are effective to reduce the cost of the entire system on the premise that the potential values of the data are not affected.
- Analytical mechanism: the analytical system of big data shall process masses of heterogeneous data within a limited time
- Data confidentiality: most big data service providers or owners at present could not effectively maintain and analyze such huge datasets because of their limited capacity. They must rely on professionals or tools to analyze such data.
- Energy management: the energy consumption of mainframe computing systems has drawn much attention from both economic and environmental perspectives
- Expendability and scalability: the analytical system of big data must support present and future datasets. Algorithms must be able to process increasingly expanding datasets.
- Cooperation: analysis of big data is an interdisciplinary research, which requires experts in different fields to cooperate to harvest the potential of big data

In conclusion, designing a big data gathering platform for improving operational performance is a challenging process. The project faces a lot of risks for failure. This paragraph provided some challenges that will emerge during the process of designing, implementing, and usage of the big data platform. These technical guidelines can be used to form a roadmap for solving problems. The next paragraph elaborates information about challenges regarding the people involved in the project.

2.4.3 Management of change and Employee's acceptance

Interesting statistics and technologies are promising for the manufacturing industry. Implementing such technologies is a difficult and time-consuming process. Besides the technological innovation, the organizational change must be managed as well during the implementation of an information system (Ray, Wang, Chang, & Hubona, 2011). The usage of the implemented systems and technologies needs to be explained continuously to preserve the proper use of the system or technology. This paragraph provides information about aspects to consider when changes in information systems are implemented within an operational organization.

Companies must not ignore the importance of management of change strategies that ensures the communication of the change, collaborates with the users, demonstrates management support, supplies technical availability, and conducts training to increase perceived ease of use (Ray, Wang, Chang, & Hubona, 2011). The article of Ray, Wang, Chang, and Hubona (2011) investigates the management of change strategies to ensure the success of information systems. The article proposes to consider and control four change management approaches: management of change effectiveness, readiness to change, resistance to change, and end-user computing satisfaction.

The management of change construct focuses on effectively sharing information with the employees and the handling of the employees' concerns, involving communication, support, fairness, technical availability, and training (Ray, Wang, Chang, & Hubona, 2011). Involvement and the opportunity for sharing concerns are important to create acceptance. This requires dedicated leadership and governance of the project. This includes leadership and social involvement when defining the objectives, aiming to create a cooperative behavior of future users (Erlicher & Massone, 2005). Secondly, the readiness (acceptance) to change reflects the extent to which the users are cognitively and emotionally inclined to accept, embrace, and adopt an improvement/change (Ray, Wang, Chang, & Hubona, 2011). The user's opinion about the change must be monitored to react properly and to create a readiness for the change. The third approach of change management is managing the resistance to change. Resistance must not be seen as a disruptive element, but the complaints must be managed. The thoughts and way of thinking of the users affect the conflicts and actions during the implementation (Erlicher & Massone, 2005). Therefore, leadership is needed to manage this resistance, for instance, solving factors as lack of involvement in the change process, lack of management support, poor quality of the future system, and the lack of designer-user interaction (Ray, Wang, Chang, & Hubona, 2011). Lastly, end-user computing satisfaction must be guaranteed. Satisfaction in terms of how the system supports the users' work, which information is generated, and the ease of use (Ray, Wang, Chang, & Hubona, 2011). Another piece of advice for obtaining employees' acceptance is the magnitude of the change. There is a distinction between radical and incremental changes for gaining support for changes that are required to improve the desired use of the innovation by employees (Erlicher & Massone, 2005).

2.4.4 Designing a performance measurement system

The previous paragraphs focus on the design and management challenges of embedding the manufacturing execution system with big data functionalities in the organization. This paragraph focuses on the daily use of the system, on which performance indicators are presented to the users.

The performance measurement should create a supportive work environment; performance measurement can be defined as "the process of quantifying the efficiency and effectiveness of actions" (Kamble, Gunasekaran, Ghadge, & Raut, 2020). Constructing a supportive performance measurement system can be challenging when a company does not have clear objectives. Possible fallbacks and challenges of a performance measurement system are (Hon, 2008):

- Lack of training on how to feed the performance measurement system.
- Poor understanding of provided information. Simplicity is therefore a useful attribute for a measurement system.
- Poor division of tasks, vagueness on what tasks must be executed by management and by operators. These groups are key parties to obtain efficient data gathering for analysis.
- Lack of awareness is felt most strongly by operators (the result of lack of training).
- Lack of resources for the efficient and continuous operation of the system/tasks.

These challenges need to be solved for the proper use of the performance measuring system. The portfolio of important performance indicators, the key performance indicators (KPI's), must be composed to support the decision-making processes in operations. The management should create a set of performance indicators that supervises their performance and this set of indicators should be improved continuously to focus on an increase of performance. The performance measurement system can contain different indicators from different categories. The following overview of categories and indicators can be used for composing the set of performance indicators.

Examples of performance indicators grouped by different categories (Hon, 2008):
- Time measures: average batch processing time, average lead time, changeover time, cycle time, machine downtime, mean flow time, on-time delivery, setup time, Takt time, throughput time.
- Cost measures: overhead cost, scrap cost, setup cost, tooling cost, total quality cost, unit labor cost, unit manufacturing costs, unit material cost, work in progress.
- Quality measures: average outgoing quality limit, incoming quality, MTTF, not right first time, process capability index, return rate, rework %, scrap %, vendor quality rate, warranty claim.
- Flexibility measures: component reusability, delivery flexibility, machine flexibility, number of different parts, process flexibility, process similarity, routing flexibility, supply chain flexibility, total system flexibility, volume flexibility.
- Productivity measures: assembly line effectiveness, direct labor productivity, machine effectiveness, network effectiveness, overall equipment effectiveness, return on assets, stock turn, throughput efficiency, total productive maintenance, value-added per employee.

The smart manufacturing performance measurement system (SMPMS) is a method that structures the performance indicator creation process. The SMPMS method for improving operational performance and designing the portfolio of performance indicators is stated in appendix H. The productivity measure of overall equipment effectiveness (OEE) is widely accepted as a tool to monitor the actual performance of equipment (Corrales, Lambán, Korner, & Royo, 2020). This performance indicator is used at the case study company. This measurement will be explained more in-depth in the next paragraph.

2.4.5 Overall Equipment Efficiency

Overall equipment effectiveness (OEE) is a key performance indicator (KPI) of the total productive maintenance (TPM) approach (Corrales, Lambán, Korner, & Royo, 2020). This method is developed by Toyota and has the objective of improving production performance. The OEE is the productivity ratio between real manufacturing performance and planned manufacturing performance (Corrales, Lambán, Korner, & Royo, 2020). The framework of the OEE is stated in figure 12. An explanation of the six big losses is stated in appendix I.

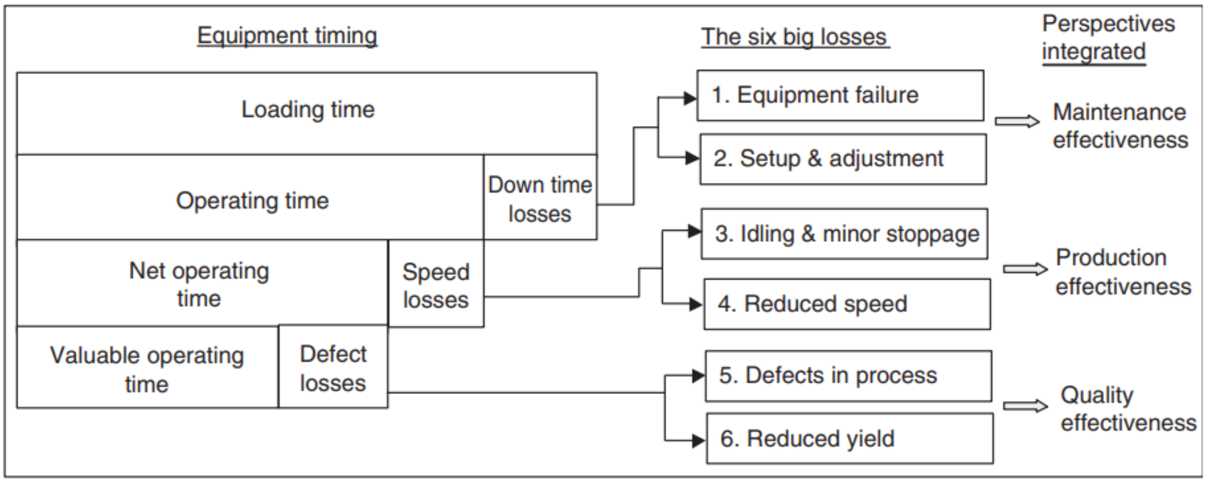


Figure 12: OEE, six big losses, and the perspectives integrated (Muchiri & Pintelon, 2008)

The OEE is a multiplication of three separate KPI's: the availability of machinery, the performance of the machinery, and the quality of the products. These three KPI's are presented as a percentage. The KPI's divides the production time into different production loss classes: downtime losses, speed losses, and defect losses, which can be divided into six big losses. The result of the OEE calculation is the valuable operating time. The corresponding formulas are stated in formulas one to five. These formulas are the basic formulas of OEE from the article of Muchiri and Pintelon (2008)

$$OEE = A \times P \times Q \ (1)$$

$$Availability\ rate\ (A) = \frac{Operating\ time\ (h)}{Loading\ Time\ (h)} \times 100\% \ (2)$$

$$Operating\ time = Loading\ time - Down\ time\ (3)$$

$$Performance\ efficiency\ (P) = \frac{Theoretical\ cycle\ time\ (h) \times Actual\ output\ (units)}{Operating\ time\ (h)} \times 100\% \ (4)$$

$$Quality\ rate\ (Q) = \frac{Total\ production\ (units) - defect\ amount(units)}{Total\ producion\ (units)} \times 100\% \ (5)$$

In short, availability—'Is the machine running or not?'; performance—'How fast is the machine running?'; and quality—'How many products satisfied the requirements?' (Corrales, Lambán, Korner, & Royo, 2020).

2.4.6 Development of reporting

Designing and developing a manufacturing execution system that gathers real-time data for real-time decision-making and for analyzing historical data is the first step. In addition to this, performance measurements must be designed to support operational, tactical, and strategical decision-making. After executing this, a company must also update and improve the performance monitoring system. This can be realized by the continuous development of a system of reports which informs all layers of the company, from shift leader to higher management. The management analyzes problems and reports should be created to monitor and control these problems. A real-time operational data gathering system should be designed to report performance feedback to its users, the information from these reports should be used to improve this performance (APICS, 2017). The American Production and Inventory Control Society (APICS) created a guide for the development of a performance measurement system, the supply chain operations reference model. This model has the purpose to define process architectures in a way to align processes with key business functions and goals (APICS, 2017). One part of the book elaborated on the creation and updating of reports. The report updating cycle consists of six steps: step 1, initiate reporting on performance measurements; step 2, analyze reports, receive and plan maintenance requests; step 3, find root causes; step 4, maintenance request, IT request, or prioritizing of root causes; step 5, develop corrective actions; step 6, approve and launch corrective action; step 1, initiate reporting (APICS, 2017). This reporting creation cycle can be used to update the report portfolio of a production company, the process is explained in appendix J. This continuous updating is required to achieve incremental performance improvement.

Conclusion

Big data analytics offers many opportunities to evaluate data into information. The information can be utilized to support operational, tactical, and strategical decision-making. The MES-system and ERP-system gather data and provide information that can support these decision-making processes. The MES-system can be developed to function as a big data platform that gathers shop floor data, this will be the foundation for the operations performance measurements and big data analytics. Developing a big data platform is a challenging process. Technical and managerial challenges have to be solved to create a supportive IT-system for the organization. The main technological challenges are decisions about how to design the system. This includes making decisions about the characteristics of big data, which can be summarized in the four V's: Volume (the size/scale of the data), Variety (the form/format of the data), Velocity (the rate of the data being produced), and Veracity (the uncertainty/reliability of the data). The managerial challenges are also important factors because this has a major influence on a successful IT-system implementation. A company must create a management of change strategy. The important factor is the involvement of users/employees, this is important to create acceptance and satisfaction.

Performance measurements are important to support decision-making if the IT-system is operational. The OEE indicator is a widely accepted measurement to monitor the performance on availability, production speed, and quality. Performance information can be presented by generating reports. A company should improve the set of performance measurements and the reporting portfolio continuously. Both technical and managerial solutions will contribute to the objective, creating a supportive and effective real-time big data platform.

2.5 Conclusion and discussion literature research

The conclusion subchapter of the literature review consists of two parts. The first part is the general conclusion and discussion of the literature review. The second part is the discussion that forms the connection between the literature review and the case study.

2.5.1 Conclusion and discussion

The literature review presented information about information system architectures and the development of a big data platform. The review aims to answer the following research question: How should a production company design its information system architecture to monitor, control, and improve operational performance and be ready for future technologies?

Production companies should develop their information system architectures to form a basis for creating manufacturing intelligence and adopting modern industry 4.0 technologies. The information systems should be developed into a sufficient maturity level to reach this objective. The first step of growing to a higher digital maturity level is the implementation of the industry 4.0 technology vertical integration. The connection of three different manufacturing application layers is required to obtain this vertical integration. The ERP-layer, MES-layer, and SCADA-layer must be connected to monitor and control operational performance. The connections can be realized according to the ISA-95 standard, which describes the interfaces between those layers. The development in vertical integration and subsequently horizontal integration will result in an industrial internet of things platform that will convert the factory floor into a smart manufacturing environment.

The first three steps of the vertical integration are the overall implementation of a data acquisition platform, real-time data collection and monitoring, and generating an alarm list that monitors the production parameters. The digital maturity of a company will transform into a vertical integrator after developing these features in the information system architecture. The connected data will support the operations with real-time operational performance data and will support supporting staff with stored data for real-time and historical analytics. The big data must be accessible and reported understandably to be supportive and effective for all users throughout the company.

This literature review is focused on providing practical information on the development of the information system architecture of production companies. The review provides guidelines for decision-making during the implementation process. This research supports production companies in upgrading their information system architecture toward the vertical integrator maturity level. Available literature focuses mostly on implementing new industry 4.0 technologies, but literature focuses less on forming the basis/foundation for the implementation of these technologies. The information from this literature study forms a bridge between the current state of most production companies and futuristic industry 4.0 production environments. Future research can focus on providing more information about vertical integration implementation. More initial information will decrease the risks during an information system architecture implementation. This will contribute to the aim of the project, making new analytical possibilities and new technologies assessable and implementable for production companies.

2.5.2 Discussion, connection to DYKA-case

Companies should start with the creation of information by vertical integration. The objective is to extended/develop MES-software systems into a supportive MES-layer, which contains all information that is required to produce products efficiently and with the required quality. This MES-layer requires both real-time production data, job status information, and product, and machine specifications.

Companies should create a strategy to develop the MES-system according to the three steps, these steps are the overall implementation of a data acquisition platform, real-time data collection and monitoring, and getting an alarm list that monitors the production parameters (Zhang & Wang, 2020). Step one is the creation of the SCADA-layer that connects the machinery to a system, step two is the real-time data gathering, storing the real-time data, and analyzing the historical data. Step three is the implementation of the use of the data in real-time.

In contrast to the literature study, which can be generalized across the manufacturing industry, the case study focuses on a specific industry. The case study will provide a prototype of a big data platform and will create information that is required for the execution of an implementation project for an extrusion department in the plastics industry. The case study company already has MES-software in place, but this is not developed into the contemporary available data presentation and analysis technology. The objective of the case study is to provide insights into how the company can improve its data usage and visualization in the future. This case study should form a body of knowledge and insights to start IT development projects throughout the company in the future. This explorative in-depth case study will be guided by information from the literature research.

3. Problem Context

This chapter is the first part of the case study. This case study part describes all relevant information to understand the extrusion production process and the use of data within this process. The information from this chapter and literature review will form the foundation for the ideas about big data gathering and information system architecture improvements of chapter four. The following research question is answered in this chapter:

> What is the current situation of the extrusion production process and the information system architecture?

This chapter consists of a process description, the current data processes in different software systems, the current calculations of KPIs in the MES-software, and the opportunities for the future presented by the MES-software supplier.

3.1 Process description/ System description

The information about the production process of the extrusion department is elaborated within this subchapter. This subchapter contains the general description of the department, the general description of an extrusion machine, and an explanation of the quality checks performed by operators and quality engineers.

3.1.1 General department description

This paragraph provides information about the extrusion department which is required to create a data-gathering strategy. This paragraph consists of a description of the departments that are involved in the extrusion production process, the overview of machines within the department, the department's staff structure, and the operational organization on the work floor is explained.

Departments involved in the extrusion process

The extrusion department includes the production of pipes, belling machines, and packaging machines. A belling machine enlarges the end of a pipe to make it possible to slide pipes into each other to connect the pipes. This research focuses on the processes within this department. The other departments are stakeholders who, in the future, can benefit from the improved information system that will be developed at the extrusion department. These stakeholders are departments that form the internal supply chain of DYKA, starting with the supply chain department and ending at the storage and transport department. These processes are visualized in figure 13, using the business process modeling notation (White, 2004). The explanation of the symbols is stated in appendix K. This figure, which is validated by stakeholders, provides an overview of the processes within the departments that are involved in the production of pipes.

The input information of the process is about material availability, planning, and process settings. The filling of the blended material in buffer silos is the responsibility of the blending department, the extrusion department starts when the blended material is sent from the buffer silo to the extrusion machine. The extrusion department produces the pipes, post-processes the pipes with the belling machines, packages the pipes, and executes the first quality check on the work floor. The quality department is doing the other checks and will send feedback if products are not within specifications. The products will be released if they pass the quality checks of both departments. The well-produced products, packages of pipes, are in that case handed over to the storage and transport department. The rejected products will be sent to the plastic crusher at the blending department and will be recycled. This material will be used as raw material for the core layer of three-layer pipes.

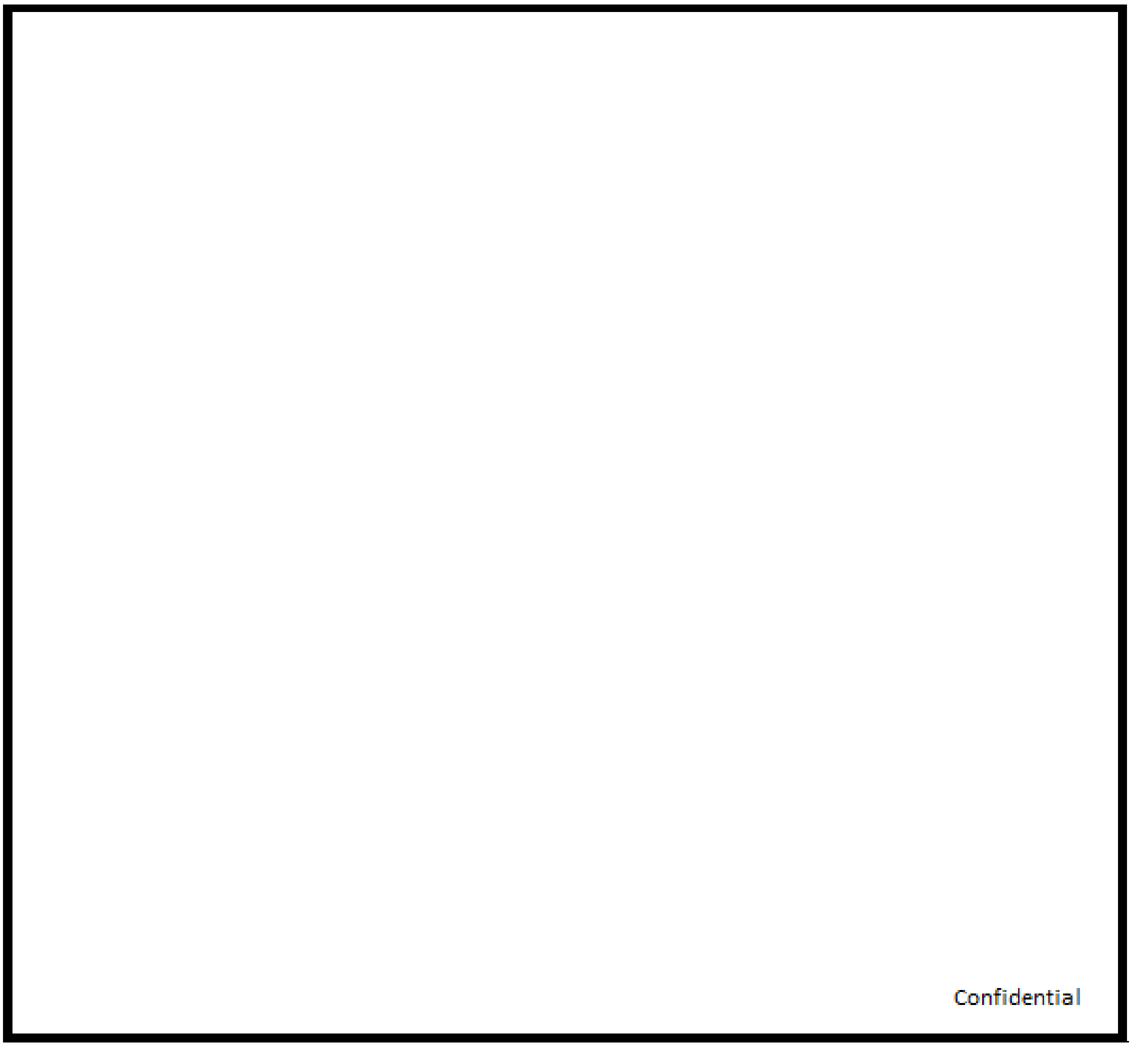


Figure 13: Processes involved in the process of extruding pipes

Extrusion department

The extrusion department consists of (confidential) production lines that produce plastic pipes. An overview of the department is provided in figure 14, this is the floor plan of the extrusion department. The machine numbers are stated at the bottom in the figure, the extruder machine is situated above these numbers, and the product direction is from the bottom to the top.

The plastic pipes are produced for various applications as described in the introduction, chapter 1.1.1. The department has a product portfolio of (confidential) products, in various colors, diameters, wall-thickness, compounds, and other product characteristics. (confidential) machines are producing PVC pipes varying from wide diameters and small diameters. (confidential) of these machines can produce three-layer pipes with two compounds. These pipes consist of A-product PVC at the inner, and the outer layers of the pipe and recycled PVC in the core layer. Besides the production of PVC pipes, Polypropylene (PP) pipes are produced, these pipes can resist higher temperatures. The department has (confidential) machines for producing PP pipes. The last machine to make the (confidential) lines complete is the PVC granulate machine. The output from this granulate machine is used as raw material for the other machines.

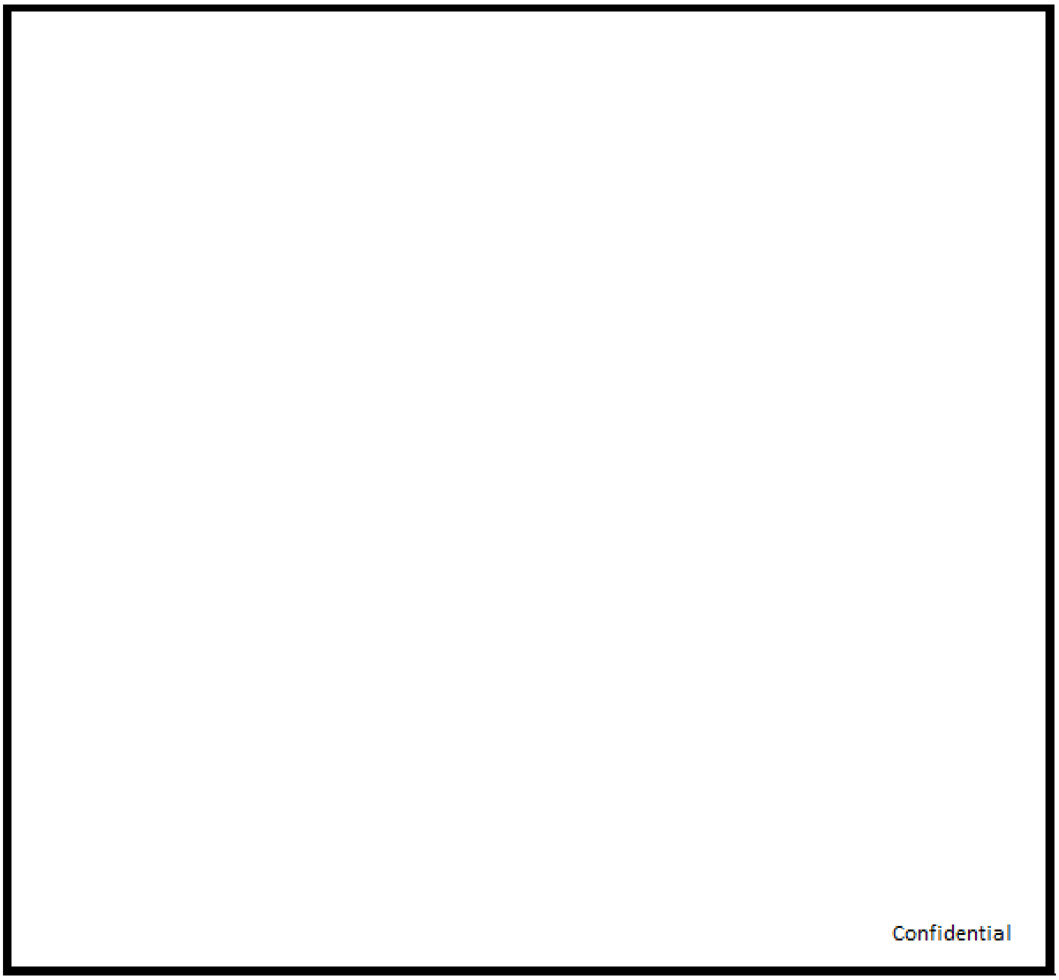


Figure 14: The floorplan of the extrusion department

The department has also off-line machines that finish the products. Three separate belling machines are used to make pipe to pipe fittings at the end of the pipe. In addition to this, three other belling machines are placed in line with the extrusion production lines. The department also has two plastic packaging machines to package the pipes or package the pipes in a seal. Table 4 provides an overview of the production machines and the type of machine.

These production lines are machines from five different manufacturers, Krauss Maffei, Theyson, Battenfeld-Cincinnati, Rollepaal, and the new machine is from Weber. An overview of the production lines and their manufacturer can be seen in appendix L. The machine control version is also added to provide information that can be used for the plan to connect machines to software later in the research.

Confidential

Table 4: Machine types

Organizational structure
(confidential)

Operational structure
(confidential)

3.1.2 General description Extrusion machine

This paragraph provides information about the extrusion line, the basic material science, the functioning of the extruder, and the problems that cause downtime.

Extrusion line
The extrusion line is the machine that produces the plastic pipes. The extrusion process is the process from the material dosing in the machine until the packaging station. This paragraph takes a closer look at the 'Extrusion of pipes' event illustrated in figure 13 in paragraph 3.1.1. Figure 15 illustrates a basic extrusion line.

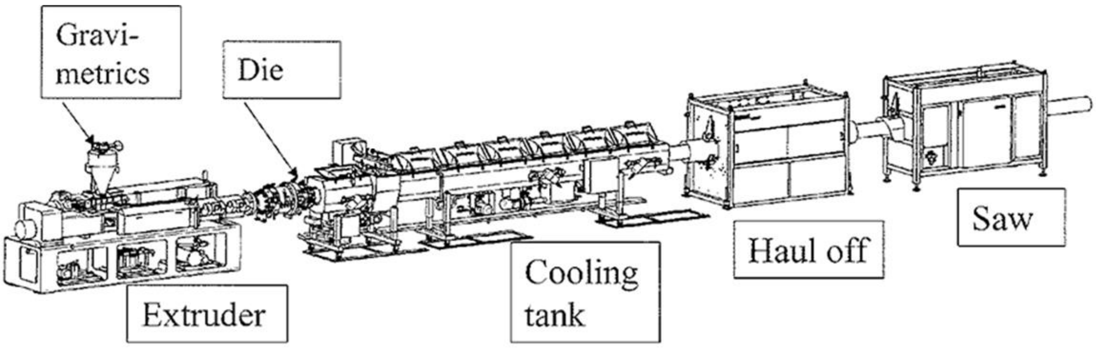


Figure 15: Visualization of an extrusion production line (Rollepaal, 2019)

The production of the pipe starts with the dosing of materials, more about materials will be explained in the material science paragraph 3.1.2. This dosing can be volumetric dosing or gravimetric dosing. Volumetric dosing doses continuous volumes with a screw into the machine. The disadvantages of this method are that it is inaccurate because the system cannot react to changing circumstances and there is no registration of consumption. The gravimetric dosing is a smarter dosing solution because it is a self-regulating system that regulates the dosing on the result of the ratio mass per meter and logs the consumption data. The extrusion department has already improved (confidential) machines with gravimetric dosing modules. The other machines are supplied by the older technology of volumetric dosing. Volumetric dosing will be gradually replaced by gravimetric dosing in the future. The second basic step of the extrusion line is the extruder. This is the main part of the machine, more about this process step will be explained in paragraph 3.1.2. The third step is the extruder die-head. This part has different configurations so it produces pipes with different diameters. The exchanges of die-head cause a large part of the machine downtime. The next part is the vacuum tank with a calibrator. This part ensures the roundness of the pipe and starts cooling the pipe. A wall-thickness measurement can be inserted after this first vacuum tank, but the place of the wall-thickness measurement device can be varied. However, it is better to install it as close as possible to the extruder machine, because it is preferred to gather product quality data as soon as possible after a machine parameter adjustment. Wall-thickness measurement is promising for the future with the development of self-calibrating terahertz wall-thickness measurements because this data will become interesting if the company has developed its MES-system. The number of stations after the vacuum tank with calibrator can vary, it can be extended with more vacuum tanks and cooling tanks depending on the product that is produced. The printing station is positioned after this variable part. This device prints the product type and other information on the pipe. Fifthly, the haul-off machine pulls the pipe through the machine. After this pulling machine, the pipes are sawn into a particular length. An extrusion line can be extended with belling machines and a rubber ring positioning machine after this step. In the end, the pipe is packaged at the package station. After this, the packages are sent to the special products department or storage and transport department.

Material science extrusion process

Basic material science is needed to understand the occurrence of quality issues in PVC products. The PP production is disregarded because this application is more simple and not the core business of the department. This paragraph is written with the use of the reader of the company's education program (Beerda, et al., 2020). Polyvinyl chloride (PVC) is a thermoplastic material, which means that it can be melted and reshaped after usage. The company uses this recycled PVC as the core layer of three-layer pipes, this recycled material is called recyclate. The A-product PVC is called dry blend. Table 5 shows the material, additives, and its percentage of weight used to produce the PVC pipe.

Confidential

Table 5: Material and additives of dryblend and recyclate PVC

The recipe for PVC products has a fixed part and a variable part. This fixed part is a proportion of a specific material that must be added to create a stable compound. The variable part, adding lubricants, is standard 0.5% of the total weight of the material and can be added to create proper compound characteristics for producing products to gain the required quality. This composition can vary because the material characteristics can be different due to different suppliers. The main process characteristic which results in quality characteristics is mass temperature, this process parameter can be increased and decreased by adding or lowering the amount of lubricant.

There are two important indicators for the production of high-quality products. These indicators are mass temperature and mass pressure. Controlling these parameters will ensure that the material is a stable compound. PVC is an amorphous thermoplastic, which means that the molecule structure is not crystalized. The material PVC can be in three different transition phases, the glass phase, the rubber phase, and the liquid phase. The material must be in the rubber phase to have in the best circumstances to produce pipes. The material is in this phase if it has a mass temperature between approximately 190 to 210 degrees Celsius. This temperature is important because this is the main indicator of the degree of gelation, this occurs in the rubber phase. Gelation is the transformation of the primary chain structure, PVC molecule, toward its secondary chain structure, the crystal structure (Piszczek, Tomaszewska, & Sterzynski, 2010). The gelation degree is the degree of links between polymer chains. The unlinked polymers are mixed, this combination of links and mixing results in better mechanical properties of the material. The gelation degree is optimal at a level of 75% (Beerda, et al., 2020). The mechanical properties can be measured in impact strength, shrinkage, and burst pressure, the gelation degree can be tested with dichloromethane-test. The mass temperature is directly linked to the mass pressure, the next paragraph explains more about this indicator and its relation to temperature. The mass temperature is a predefined process parameter. The specific temperature depends on the produced product and has mostly a tolerance of plus-minus two degrees Celcius. Controlling the two process parameters ensures a large part of the product's quality.

Extruder machine

This paragraph focuses on process variability that controls the mass temperature and pressure, which influence the quality of the product consequently. The values of these two process parameters are the result of the input material and the machine settings of the extruder machine. An example of process/machine settings can be found in appendix N. The five important process parameters according to the department's process engineer are the process parameters mass temperature, and mass pressure and the machine parameters, the screw torque, material dosing values, and line speed. The machine parameters control the process parameters and the process parameters control the quality. An example of the design of an extruder can be seen in figure 16.

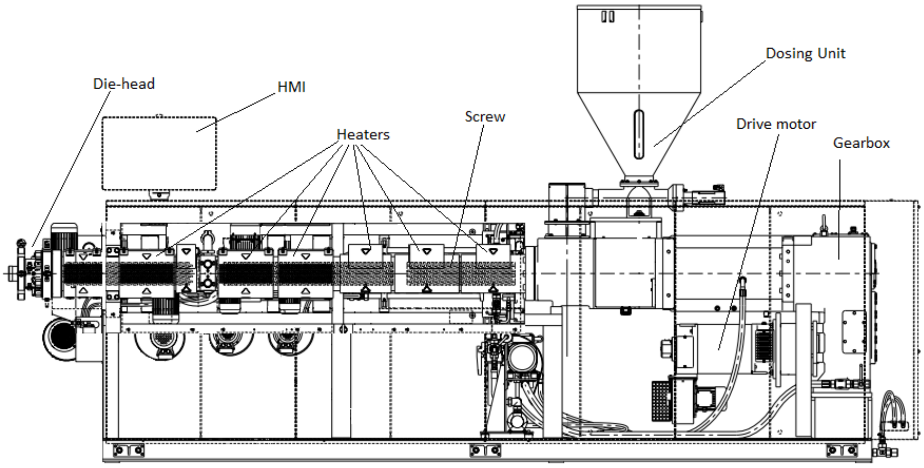


Figure 16: Extruder machine

The important machine parameters can be recognized in figure 16. The dosing control unit is the unit at the top of the machine. Focusing on data gathering, these values can be obtained from the gravimetric system. The screw rotational speed (rpm) is the speed of the two extrusion screws, which are supplied by the electrical drive motor and controlled by the gearbox. The line speed is controlled by the haul-off at the end of the extrusion line (not in the figure). The mass pressure and temperature are measured at the end of the screw. The other settings from appendix N are influencing these parameters, but most machine settings has often the same values.

The screws are designed to create friction because this friction results in mass pressure and an increase in mass temperature. The proper machine settings and material dosing result in controlling these parameters. Controlling these main process parameters ensures the control of quality parameters partly. An overview of the influencing machine settings and the quality effects of the process parameters are stated in figure 17.

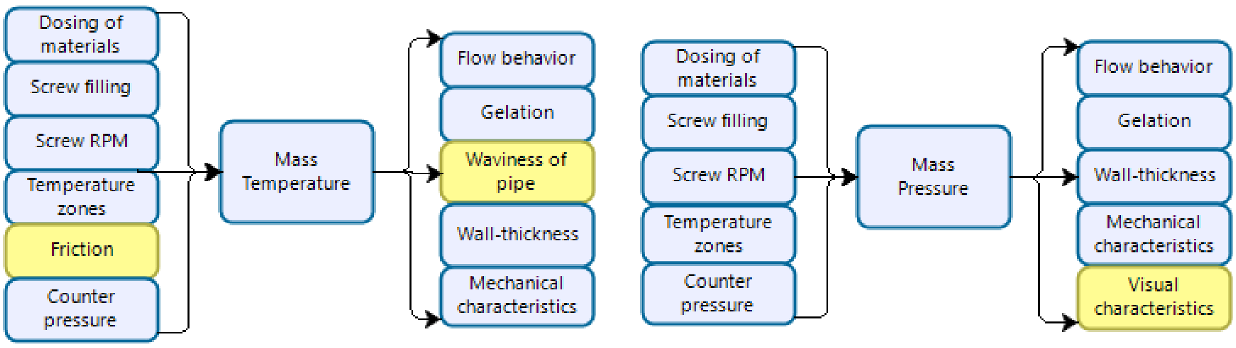


Figure 17: Influencing process settings and quality effects of the two important process parameters

The mass temperature and pressure are strongly related to each other. The machine settings input and quality output are almost the same. The differences between the influencing machine settings input and the quality output of the process indicators are marked in yellow in figure 17.

Problems causing downtime

This paragraph provides a list of problems that can cause downtime in production. This is done by interviewing stakeholders and analyzing machine alarms from extruder machine nineteen with the process engineer. The knowledge of previous paragraphs is required to explain these occurrences. Table 6 shows an overview of the causes of failures that result in machine downtime and quality issues. These possible failure reasons are extracted from the machine alarms of production line nineteen. Those alarms are already present at the HMI of most machines, but the design and utilization can be improved. Some alarms are also presented during the setup time of the machine because the system does not take into account the production status. The alarms are evaluated with the process engineer of the department and grouped into the groups of table 6. The analyzed alarm list is stated in appendix O. The analyzed machine data is the data of line nineteen from 18-8-2020 to 10-11-2020. The numbers and totals are the summary statistics of that period, this is added to indicate the distribution of the alarms. In other words, the numbers cannot be used as conclusions and be generalized for the performance of the entire department.

Confidential

Table 6: Possible failure reasons

An incorrect mass temperature and a failing gravimetric system result in quality issues and the machine fails when it is too long in that state. For instance, a production line will have a burning failure in the die-head if the machine is longer than 15 minutes in one of these states, therefore these problems result in downtime which results in lower performance. This production performance is measured by the OEE, which will be explained in paragraph 3.3.2. The controlling of the machine parameters ensures the input for the extrusion process, and the quality is the output of the process. Utilizing data that describes the state of the production and quality parameters will increase the ability to react to this discrepancy and eventually may prevent failures or downtime.

3.1.3 General description quality check

Controlling machine parameters and process parameters is important to obtain products within the required quality. Post-processing quality checks are needed to check the quality of the product before it will be delivered to customers. The quality checks can be divided into the quality checks at the extrusion department and the quality checks at the quality department.

The quality checks on the department are executed continuously because the operators (A, B, C) are responsible for visual quality checks. Operator B checks the quality parameters every four hours and at the start of a newly produced product. These parameters have target values with tolerances. These quality parameters are saved in the quality module of the MES-system. The failure products are sent to the scrap crusher and the material will be reused in multi-layer pipes. The scrap is measured in the quality performance indicator of the OEE, these quantities are not entered accurately into the MES-system because of the lack of declaring the scrap by workers. This is the reason for the low data quality of the data gathering in MES. The quality checks of operator B and the quality department are stated in table 7. The table functions as an overview, there is no relation between columns.

Confidential

Table 7: Overview of the quality checks on the department and at the quality department

(confidential)

In conclusion, gathering and monitoring the process and quality parameters will provide more insights into the process. In addition to this, access to more information can support both the extrusion and quality departments in improving their performance. Therefore, it is important to control and monitor these parameters during production to prevent downtime or quality issues.

3.2 Information system architecture and current information

The general processes are explained previously, this subchapter focuses on the information and information systems used during those processes. Thus how the IT-systems are aligned and which information is used from these and other systems during the production of a pipe.

3.2.1 Information systems architecture DYKA

This paragraph describes the information system architecture of DYKA according to the literature of chapter 2.3.3, where the ISA-95 system layers and the MESA-11 MES features are explained. The company has fewer connections between the different architecture layers in contrast with that literature. The information flows between different modules within the three information systems software layers are visualized in figure 18. The lack of connections between systems results in little available information currently. This figure is constructed using the Archimate modeling language, which is a visual language for describing, analyzing, and communicating enterprise architectures (The Open Group, 2019); an enterprise architecture is the combination of business processes (processes modeled in yellow), supporting applications (blue), and the physical technological process (green). An explanation of the Archimate symbols is added to appendix P. This appendix contains also an enlarged figure of figure 18. This chapter consists of three figures that represent the current situation of operations and IT. Figure 13 explains the process steps that are required to produce a pipe, figure 18 explains the available software applications used during production, and in the next paragraph, figure 19 combines these two figures into the information/software used during the production of pipes.

The ERP-system contains the production order planning and finished product flow. The department receives the production order and article information from the ERP. This can be seen in the interface layer between the ERP and MES of figure 18. This information is utilized to produce the pipes and in the end, the final packages with pipes are scanned and added to the warehouse management module of the ERP-system. This is a manual action, there is no feedback interface from the MES to the ERP, this action is stated at the right bottom of figure 18. (confidential). There is no connection between the HMI’s and the MES-system, and there is no SCADA-network integrated at the department. A SCADA-network can be defined as an interconnected machine-to-machine network that gathers manufacturing data. The process parameters from the extruder are saved in the extruder, but not connected and stored in an interconnected system. The extruder machines are stand-alone systems.

Confidential

Figure 18: Current information system architecture, information, and connections per layer

Currently, DYKA is in the digital novice state according to the maturity model of PWC. This means that DYKA has a partial integration of IT-systems, first digital solutions, isolated applications, no manufacturing digitalization focus, and analytical capabilities are mainly based on semi-manual data extracts (PWC, Geissbauer, Vedso, & Schrauf, 2016). DYKA should invest more in the digital transition to grow into the next stage of the PWC maturity model, the vertical integrator stage. Where vertical integration of IT-systems is implemented, the connection between different data cubes is realized, a machine to machine network is in place, data is utilized as a key differentiator and the analytical capabilities are supported by a central BI-tool (PWC, Geissbauer, Vedso, & Schrauf, 2016).

3.2.2 Information usage in the extrusion process

The current production is mainly executed with the use of the workers' experience. Information systems are static systems that provide information. A more dynamic/smarter system uses different information sources to advise the operator, for example with alarms if process parameters are not in between tolerances. Most information sources are stand-alone systems/databases that are not connected to other systems. The current information creation and usage during the production of one production order is visualized in figure 19. The information system architecture of figure 18 should cover the information used during the process, but that architecture is not developed to fulfill these IT requirements. Therefore, figure 19 is added to provide a complete overview of the utilized information during the production of pipes. The events/tasks that are executed in software systems contain a human or a system symbol. The human symbol refers to a manual action by a user and the system symbol refers to an automatic action executed by a software system. This is visualization is inspired by the modeling approach of the business process model and notation (BPNM) based on the International Society of Automation 95 (ISA-95) (Prades, Romero, Estruch, García-Dominguez, & Serrano, 2013).

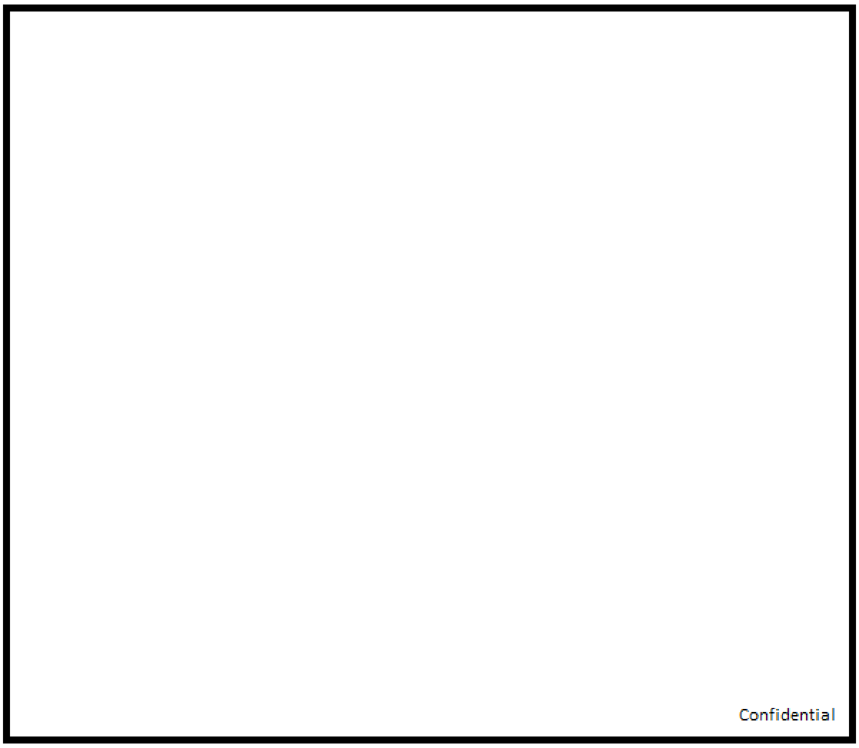


Figure 19: Information used during the production of one production order

The BPNM diagram from figure 19 illustrates the usage of information and software systems. The figure can be compared with the events in figure 13. This overview is a simplified version of that process because the focus is on information and software usage and less on the actual production process. This figure does not contain the SCADA-layer of information system architecture because this layer is scarcely connected. The saw movements, line speed, and production status are monitored in the production control of the MES-system. Other SCADA elements as the HMI's are used during 'the preparation, production, and packaging of pipes' event in the process steps pool of this figure, these HMI's are standalone systems.

3.3 Current MES-software use and capabilities

The objective of the research is to improve the data-gathering strategy of the department. This data connection will be created in the SCADA-layer and the data will be processed and visualized in the MES-layer. Therefore, the MES-system requires a more in-depth explanation than the ERP-system. This subchapter describes the current use of the MES-system. Secondly, the current key performance indicators are investigated and the pros and cons of these calculations are determined.

Current use of MES-software

The current MES-layer of the information system architecture contains a small part of the currently available applications and technologies for this layer. This layer is the foundation for analytics and other comprehensive data processing algorithms and technologies. As described earlier, although there are few connections between the architecture layers, these connections are needed to strengthen the data processing capabilities. (confidential)

Current calculation of OEE KPI's in MES

The company's major Key Performance Indicator (KPI) is the Overall Equipment Effectiveness (OEE). This KPI is a valuable measurement that provides information about the source of lost time and lost production, so this tool supports the organization in optimizing the performance of the existing capacity (Muchiri & Pintelon, 2008). The theory of OEE is described in the literature chapter 2.4.5. This paragraph focuses more on the practical side of the OEE calculation in MES in contrast with the theoretical approach of the literature. Figure 20 contains an overview of losses that are nonvalue adding time to production. The OEE focuses on the three losses that decrease the loading time. This figure is made with the information of MES. This figure is the same as figure 12 in the literature chapter 2.4.5, one difference is the addition of the available time layer.

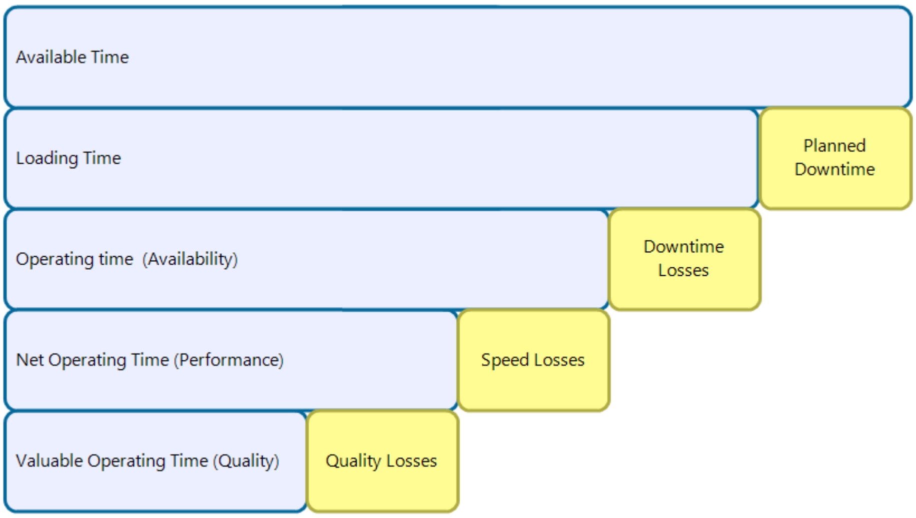


Figure 20: OEE, time losses during production

The OEE calculations at the extrusion department are calculated by the use of the manually added declaring statuses, the produced meters of pipes, the planned production speed, and the incorrect produced products. The specific calculations of the OEE are stated in appendix S. The basic OEE formulas in the reporting explorer of the current MES-system are described in formulas 1 to 4.

$$OEE = OEE.Availability * OEE.Performance * OEE.Quality \ (1)$$

The OEE is calculated with the use of manual declarations and production counters. Both inputs are unreliably caused by the improper usage of the system by the department. The operators are not declaring the right status and are not consistent with the declaration of scrap. This results in a performance measurement that is not a reflection of reality. Availability is the most reliable indicator, the performance and quality indicators are more unreliable because the input for the calculation is not entered correctly. Availability is mostly calculated with automatic counters, therefore, this performance indicator is more reliable than the others.

$$Availability = \frac{Operating\ time}{Loadingtime} \ (2)$$

Availability has the operating time and loading time as input. The operating time is the loading time minus the production time declared in stop statuses, which are grouped in the downtime OEE group. The operator does not have a precise definition of the declaration statuses, it is not clear when which status should be declared. The sum of stop declarations is reliable, but the division into the statuses is not reliable. The different stop statuses are stated in appendix R in the column 'OEE stop group'. The planned downtime and the weekend status of machines declared by the supply chain department in the planboard are not part of the downtime group of the OEE calculation. Overall, this OEE calculation part is the most reliable of the three calculations because the calculation is mostly automated. The machine is either running or in downtime. The only manual influence is the declaration state of a planned stop, which is not counted as downtime.

$$Performance = \frac{Net\ Operating\ time}{Operating\ time} (3)$$

This KPI calculates the performance on achieved production speed. This calculation uses the forecasted production time (net operating time) and the achieved operating time. This net operating time is calculated by multiplying the planned production time per meter and the produced meters. The first part is unreliable because it is not continuously updated by the department's staff, this value is mostly inserted in the system after product development. This planned production speed has to be updated and the line speed measures should be calibrated to make this KPI valuable.

$$Quality = \frac{Good\ production}{Total\ production} (4)$$

The quality calculation of the OEE is the most unreliable. The input of this calculation is the total produced pipes and the declared scrap. The problem with this OEE calculation is the input of scrap declarations. This must be done manually by the operator, but the amount of scrap production is mostly not declared in the MES-system. The OEE performance on quality is high, but in reality, this number is much lower than the attained value.

In conclusion, the inaccuracy can be improved by training the personnel. The data quality is satisfactory if the operator declares the right production status and declares the produced scrap. The gathered data is a reflection of reality in that case. The OEE calculations at DYKA become even more reliable if the input is automated and if the calculation is not dependent on manually adding information. Data gathering is an important part of performance measurement. The use of OEE as a production performance measurement necessitates accuracy in the collected performance data, otherwise, the measure can easily lead to a lack of credibility (Muchiri & Pintelon, 2008). The measurement must be a reflection of reality. Otherwise, the value of the results is not substantiated. An automated data collection and a validated calculation can result in an improvement in performance measurement.

3.4 MES-software opportunities with current software

This research aims to provide advice about data gathering and the creation of information from data. This knowledge and the knowledge about the capabilities of the MES-software enable the idea of how the extrusion department will evolve in the future. This subchapter provides the steps to be taken to gather data and to make use of that gathered data. In addition to this, the ideas are elaborated on the capabilities of the current software vendor. This information is obtained during a consultancy meeting with a consultant. The following steps will be elaborated in this subchapter, the gathering of process data, the gathering of quality data, the real-time alarms, ERP connection, and the BI connection. Those subjects are explained and connected to the software capabilities to provides a reasonable overview of the future capabilities of the system. Figures about these capabilities can be found in appendix T.

Gathering process data

Data gathering is the first step. Analytics can be done if there is a sufficient amount of data. Machine behavior data is important for production performance. Mastering these machine and process parameters results in proper machine performance and quality. The input of the production can be monitored by logging, alarming, and analyzing these parameters.

(confidential)

Gathering Quality data

Quality parameters must be collected in addition to machine and process parameters. These three different types of parameters must be gathered in one single system to compare the obtained values with each other.

(confidential)

Real-time alarms

The next step after gathering the data is to make use of the data. The data can be used for supporting the operations and for improving operations by analytics. The first step is to control the operations. This is done by creating alarms if the machine, process, or quality parameters are not in between tolerances. An operator must serve multiple machines, therefore, it is difficult to monitor all machines continuously.

(confidential)

ERP-connect

Controlling the operations is the main goal of MES-system development. The second goal is to connect the information system architecture layers. SCADA- and MES-layer connection is the first connection to make for the creation of information for events and alarms. The next architecture connection is the MES- and ERP-layer connection. This connection aligns the operations' performance with the enterprise's performance. The ERP-system contains cross-organizational information. Specifically, it contains all information to control the supply chain. Feedback from operations can improve the performance of the decision in the supply chain, as sourcing, planning, and quality controlling improvements. More specifically, the MES- and ERP-system can communicate with each other by transferring material use from and to the ERP system. Another possible connection is the sending of quality data, and the check of produced products in the MES-system and the input of the warehouse. So controlling the input and output of the department.

BI-connect

The gathering of data enables the organization to develop analytics that supports the decision-making at the department and the overall company. Controlling operations is the foundation of the department's performance. Analytics is the next step after the creation of the foundation of data gathering and operations control. The prerequisite before starting with analytics is proper data quality. The data must be reliable to reflect on reality. Automatically generated data ensures this data quality, manually inserted data is more unreliable than automatically generated data. The MES-database can be connected to a business intelligence tool. The data can be visualized in this tool. This enables the organization to support the creation of performance information and information for improvement projects.

(confidential)

3.5 Conclusion

The extrusion process is complex and requires the involvement of multiple departments. Obtaining qualitative products is difficult because of the complexity of the process, but are obtained because of the experience and know-how of the operators. Monitoring an extrusion line is difficult because an operator has to serve more than one machine. More information about the process could support the operators in monitoring and controlling their processes. The monitoring of important process parameters as mass temperature, mass pressure, screw torque, material dosing, line speed, and the mass per meter will result in proper product quality and in controlling KPIs as OEE.

The current information system architecture does not contain all elements that are required to monitor and control the production process. Currently, the process and quality parameters are available but not always accessible directly. Because information of process parameters is only presented locally at the HMI of the machine and the manual quality checks have a delay because of the execution interval of four hours. Information systems should communicate with each other and features should be added to the system. The extended manufacturing execution system can be supportive for the operators if this system presents understandable information about the process and quality in real-time. This future situation of data gathering and utilizing is researched in the next chapter.

4. Problem solution

The previous chapters provided information to form the foundation for advising the company. The research aims to provide information about the opportunities that evolve when real-time data is used to monitor the processes and gathered big data is used for historical and real-time analyses. This chapter provides the requirements for fulfilling this aim. In which the manufacturing execution system must be developed and the gathered data will be connected in a big data platform. The following research question is answered in this chapter:

> What are the evolving opportunities and the design requirements for the creation of a supportive manufacturing execution system and a big data platform?

This chapter consists of sub-chapters that provide information about the objective of the development of MES, the insights from the big data platform, the data science method for the creation of the BI platform, resulting in important big data challenges and decisions, and the continuation of the project.

4.1 The objective of the development of MES

This subchapter recaps the information from the literature review and transforms this theoretical information into practical information. Aiming to provide an overview that connects the literature and the information from the problem context, this to create a summary of information that forms the foundation of an MES implementation project.

4.1.1 Objective of data gathering in MES

The digital transition of DYKA should become a strategy instead of just an IT-project. Information Technology (IT) has evolved from administrative support toward strategic roles within organizations (Henderson & Venkatraman, 1993). This is also the case at DYKA currently. The potential of data usage will support the company in monitoring, controlling, and improving operations in the future. Currently, the company is in the digital novice stage of the digital maturity model of PWC (Ch. 2.3.1). the company's information system architecture has to be transformed into the vertical integrator maturity level. This vertical integration of IT-systems is about aligning processes and data within the company and connecting information from product development to manufacturing, logistics, and sales for cross-functional collaboration, resulting in a smart manufacturing environment (Ch. 2.3.2). Data gathering and data analysis must visualize the processes. This level of maturity can be created by the integration of the MES-system with the shop-floor layer on the one side and the ERP-layer on the other side. This results in insight into the manufacturing execution in real-time (Ch. 2.3.2). Important considerations of initiating the development of this data-driven production monitoring are stated in table 3 of chapter 2.4.1, which contains the forces for and against the digitalization of a production environment. In summary, the objective of manufacturing data gathering can be seen as improving the shopfloor management by utilizing data to achieve high manufacturing efficiency, low manufacturing costs, high product quality, and high employee satisfaction (Ch. 2.4.1).

Shop floor data is not available in the MES-system currently. DYKA should develop and extend the functionalities in the MES-system to gather process and quality data. The following steps should be executed to acquire a supportive MES-system for operations (Ch. 2.4.1):

Step 1: The overall implementation of a machinery data acquisition platform (SCADA)

Step 2: Real-time data gathering to monitor operations and analyze historical data

Step 3: Getting an alarm list that monitors the production parameters

4.1.2 Required information system architecture

The advice for DYKA is to transform its information system architecture into a vertical integration environment. According to the PWC maturity model, this is attained when the company transforms IT-systems to create digital manufacturing coordination control, homogeneous IT architecture, the connection between different data cubes, a Machine-to-Machine network, and data as the key differentiator for the business. This digital factory can be created by connecting information system layers. Connecting the ERP-system, MES-system, and SCADA-network will create an industrial internet of things on the production site. Information sharing between those layers results in real-time data supervision of operations and the creation of constructive information for decision-making through the entire organization. The future situation of the information system architecture is visualized in figure 21. An enlarged model and a simplified figure of the information flow in the system are added to appendix U.

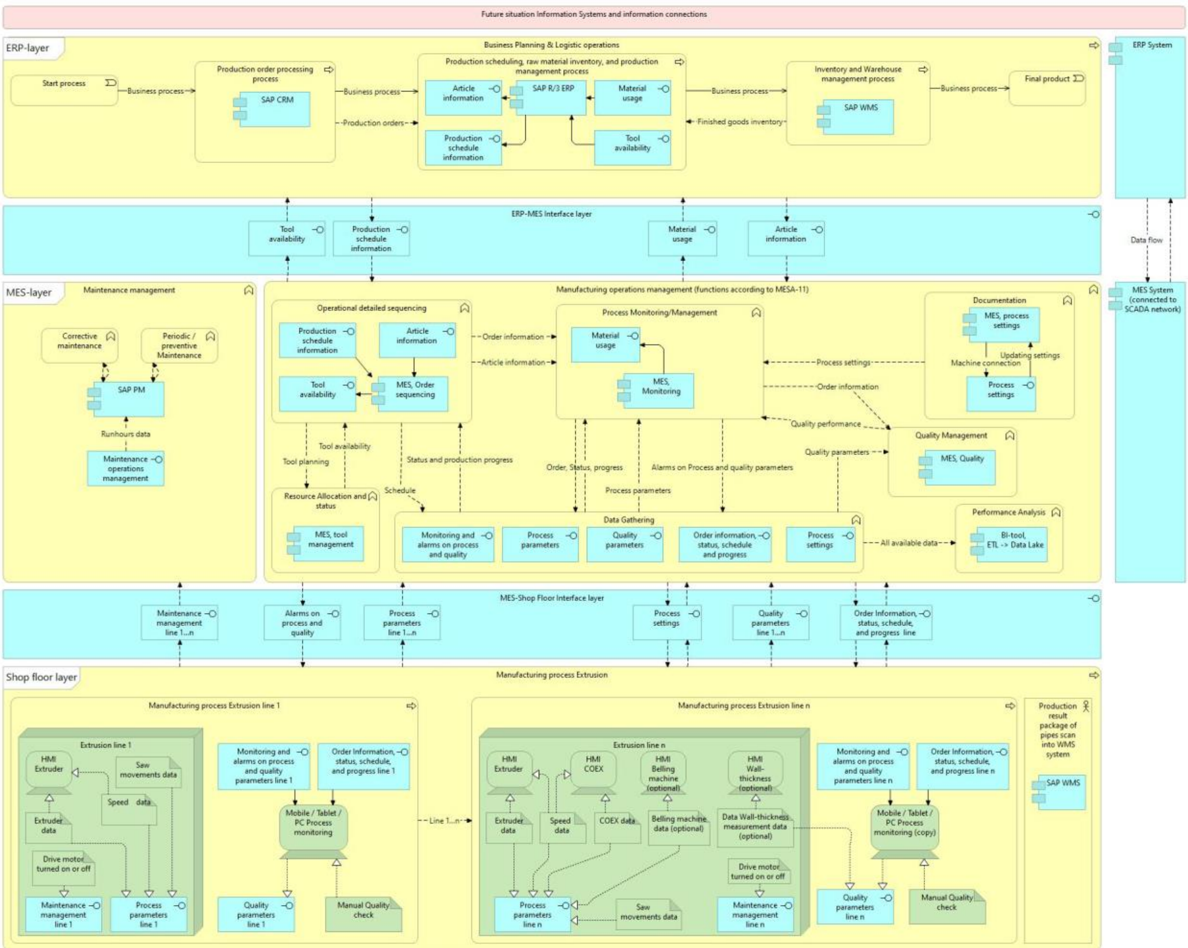


Figure 21: Future situation Information System Architecture and data interfacing

The functions of MES are the required MESA-11 functions, which will be explained in chapter 4.1.3. This is the future situation of the current architecture, which is stated in figure 18. The functions of the MES-system in figure 21 are extended with the information created or used during production, which is stated in figure 19. The information from the top lane of this BPNM-model is integrated into the MES-system functionalities of figure 21. Information from different systems is barely connected and combined at DYKA currently. The connection of the ERP-system and MES-system with data fed from the SCADA-network will increase the amount of available information at each ISA-95 layer. For instance, the administration of the raw material usage can be optimized with real-time data. This improvement will provide a material usage overview, which is a black box currently. According to the literature (Ch. 2.4), factories with connected systems are more efficient, productive, and smarter than their non-connected competitors. They can improve decision-making due to the use of more specific process information. This information creation can be obtained because of the performance analysis function of MES, which sends the gathered data to the database of the BI-tool.

Besides the MES development project, the BI-tool development project will create the ability to process the gathered data. Connecting different information from different information systems/databases with a BI-tool will create the ability to combine information and present this information in an overview or dashboard. A possible future situation of the information utilization process in a BI-tool is visualized in the BPNM figure 22. An enlarged figure from this model is stated in appendix U. This figure is inspired by the data warehouse framework of (Turban, Sharda, Delen, & King, 2010).

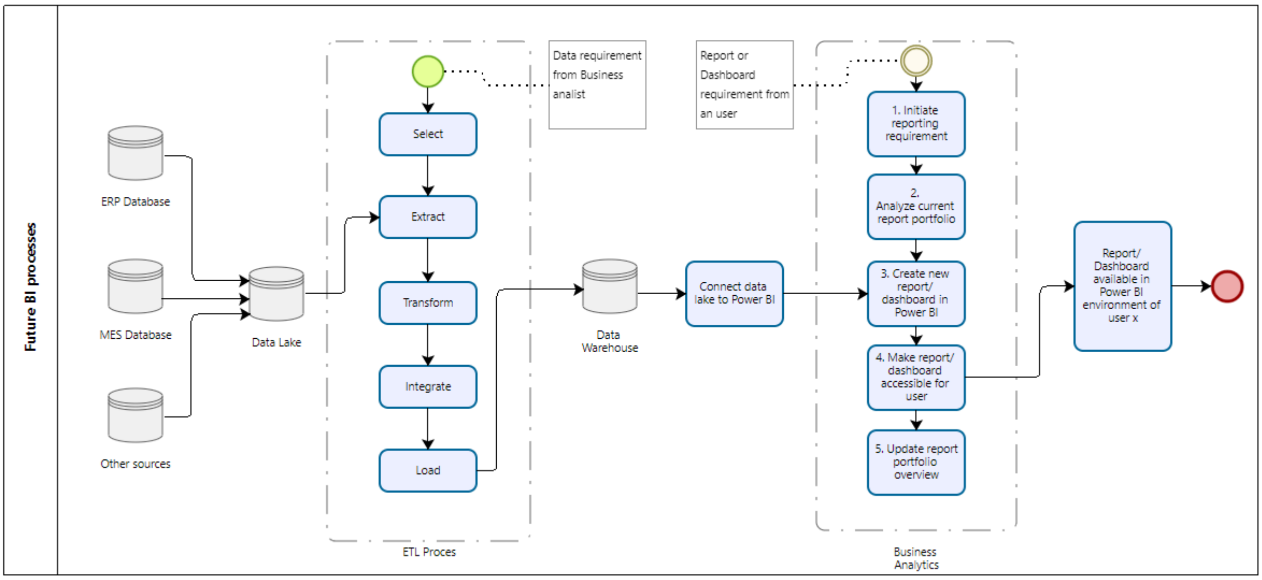


Figure 22: Future data process of information creation in the BI-tool

This process transforms raw data into structured information. Firstly, the data is gathered in a data lake. A data lake is a repository for large quantities and varieties of data, both structured and unstructured (Llave, 2018); this term is introduced by James Dixon, who is the chief technology officer of Pentaho. The gathered data is processed and structured by an ETL-process to create cleaned data, this ETL-process is illustrated in chapter 4.3. The prepared data is stored in the data warehouse. This data warehouse is connected to the BI-tool, more about the BI-tool choice is stated in chapter 4.2.1. Business analysts can create reports and dashboards with data from this data warehouse.

In summary, DYKA should update and extend its information systems in a way to gather the required data, this overview is provided in chapter 4.1.4. Each system layer must contain specific information that supports decision-making. According to the literature (Ch. 2.3.3), The ERP-system should contain information to support various business and functional units in the organization. The MES-system monitors and gathers information about the entire production cycle and provides real-time information to support decision-making. The SCADA-network is a process control system that collects the data in real-time and transmits the data to the MES-system. In addition to this, a BI-tool should be utilized to visualize the gathered information in an understandable overview or dashboard.

4.1.3 Required features in MES

The MES-system requires a major development to attain the vertical integration maturity level. The MES-system should be interconnected with the ERP-layer and the machine layer of the ISA-95 standard. In that situation at the ERP interface layer, the MES-system receives production order and bill of material information from the ERP-system, and material use, order status, quality data, and process data will be sent back to the ERP (Ch. 2.3.2). At the shop floor layer, the MES-system monitors and gathers data about the entire production cycle and provides real-time information to support decision-making (Ch. 2.3.2).

Focusing on the case at the extrusion department of DYKA. The MES-system should be developed according to the three steps provided at the end of chapter 4.1.1, these steps are the overall implementation of a data acquisition platform, real-time data monitoring and analysis of historical data, and getting alarms on production parameters. The current MES-system functions should be upgraded and extended or replaced by a system that can fulfill the requirements presented in this paragraph. The ISA-95 based MESA-11 model from the literature review explains the different functions of MES. The MESA-11 function model is stated in figure 23 and is explained in appendix E.

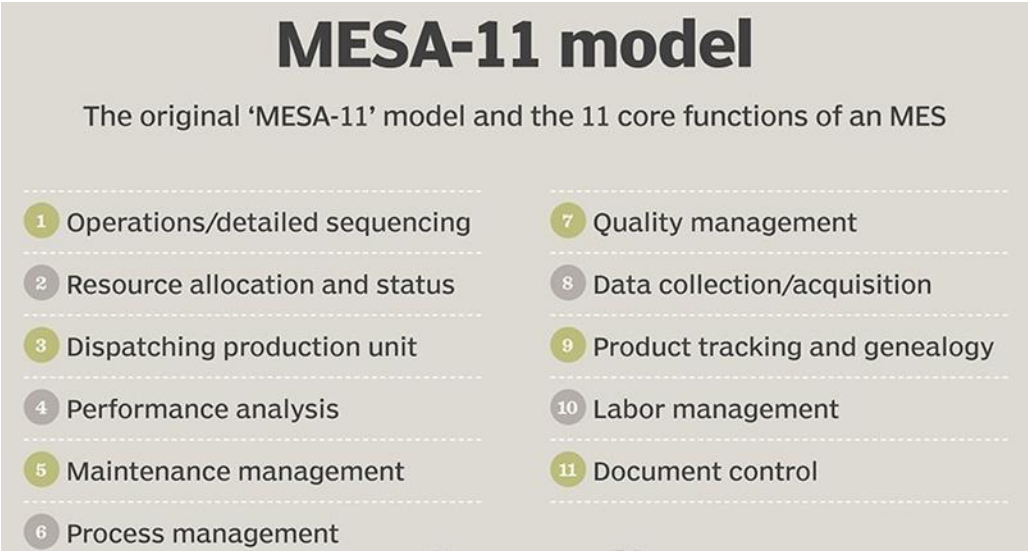


Figure 23: MES functionalities according to the MESA-11 model (TechTarget, 2020)

The functions stated in figure 23 should be developed in the MES-system to obtain the required functions to monitor and control operations. The detailed specification of each function must be developed with stakeholders. This brainstorm and decision-making of this specifications description are not part of this research. This research focuses on advice and a body of knowledge that will support the initiation phase of the MES project.

Developing the MES-system and other systems into that sufficient maturity level can play a strategic role in the need of supporting Industry 4.0 technologies to optimize the business's performance (Ch. 2.3.3). The gathered real-time data in the data collection/acquisition function of MES must be extracted, transformed, and loaded into a data lake. This data lake is connected to the BI-tool, which will operate the performance analysis functionality of the MES-system. This centralized BI-tool can easily compute the data of MES and combine it with data from other sources.

4.1.4 Which operational decisions have to be improved with the gathered data

Improving operational performance is eventually the objective of data gathering and utilization. This operational performance is measured in the OEE performance indicator. The Development and extension of the MES-system must result in prevented downtimes and fewer quality issues. The future state of the IT-systems will monitor and control operations and will provide alarms and advice when production and quality parameters are not in between tolerances. The IT development project will change/influence the operator's work and habits. Therefore, the involvement of personnel is a major influencing factor for the success of the project. The management of change strategy for the digital transition should be created to control the involvement of the employees. In addition to this, the project team should focus on effectively sharing information with the employees and handling the employees' concerns, involving communication, support, fairness, technical availability, and training (Ch. 2.3.3).

Production is partly a black box currently (Ch. 1.1.2). The process information will be captured and utilized to improve the performance of the process. Table 8 provides an overview of situations that can be controlled by using data to support operational decision-making.

OEE group	Improvement on OEE by
OEE A	Preventing downtime, controlling failures
OEE A	Controlling material flow
OEE A	Controlling start-up and changeovers
OEE A	Controlling mass temperature, prevent burning
OEE A	Controlling process parameters
OEE A	(next phase) improving maintenance plan
OEE P	Updating production speed
OEE P	Controlling production speed
OEE Q	Controlling mass temperature, gelation degree
OEE Q	Controlling mass pressure
OEE Q	Controlling other quality parameters
Other KPI	Controlling mass weight

Table 8: OEE improvements by utilizing data

Controlling process parameters is the key to improving the performance of OEE. Deviations from the standard will be recognized directly and can be solved to produce according to the standard. Controlling process parameters will also improve the performance of quality parameters (Ch. 3.1.2). Utilizing extruder process data is the first step. The required process parameters for monitoring operations are mass temperature, mass pressure, screw torque, material dosing, line speed, and the mass per meter. An overview of these and the other available process parameters from the extruder machine are stated in appendix V. The data gathering of process parameters can be extended in the future, for instance, extending with the gathering of more process parameters or automatic quality measurement data. More information about the quality measurements and the potential for transforming manual tasks into automatic data gathering tasks as stated in appendix V, this table is compiled and validated with the process engineer.

Automatic monitoring of process and quality parameters will increase the amount of information about the process and will save time that is currently used for measuring. The operator can use this time to monitor and control its processes. This continuously controlling of operations will prevent disturbances in the process. Consequently, the production performance will increase because of fewer disturbances in the process and controlled quality of the products.

4.2 Insights from the BI platform

Business Intelligence (BI) is the process of making “intelligent” business decisions based on the analysis of available data (Jensen, Pedersen, & Thomsen, 2010). Extending and developing the information system architecture will support the company by increasing the amount of data and the resulting information. This subchapter provides information about the future use of the BI-tool and provides insights from examples of the connected data in the BI-tool created in this research.

4.2.1 Future BI-tool usage DYKA

The development of machine data gathering in de MES-system will contribute to the expansion of the available data. This future situation of connected data from ERP, MES, and machine data can be analyzed in a BI-tool. This future situation is created in this research, where a prototype of the future big data platform in the BI-tool Tableau. This prototype is created with data from 01-09-2020 until 01-12-2020. The aim of creating this model is to provide insights into what information can be created with connected data and provide insights into the ease of use of a BI-tool. This aim is also stated in the literature review, where is stated that gathered data must be accessible and understandable. Information systems have their separate reporting tool currently, for example, the reporting function in ERP, the reporting explorer in MES, this is explained in chapter 3.3. Searching for information is a time-costly and rather difficult activity. Combining data from different sources is a more time-costly and not possible is one tool. Currently, excel is used for presenting and combining information from different sources. The computation capacity of excel rises to its limits because excel cannot compute large amounts of data and data updating is not automated. A BI-tool connected to the database or another data lake will solve these computing problems, chapter 4.5 proposes a process for creating the data lake and updating the reporting portfolio. Chapter 4.3 explains the process of the prototype creation by structuring this in an ETL-process.

Modern analytics and business intelligence platforms are characterized by easy-to-use functionality that supports a full analytic workflow (Gartner, 2020). Tableau offers an intuitive, interactive, visual-based exploration experience that enables business users to access, prepare, analyze and present findings in their data without technical skills or coding (Howson, Richardson, Sallam, & Kronz, 2019). This statement is tested in this research by practicing in the BI-tool Tableau. DYKA does not utilize a BI-tool currently. The parent organization of DYKA, Tessenderlo group, is testing several BI-tools. The result of those pilots will show which tool fulfills the requirements of the organization, these requirements are known currently. (confidential). This overview of BI-tools is made by Gartner, which focuses on two variables: completeness of vision and ability to execute (Gartner, 2020). Gartner is a global research and advisory company that provides advice about information tools. (confidential). The overview of BI-tools is called the magic quadrant of Gartner, a figure with this overview is added to appendix W. The model of the research is made in Tableau because of the available license, DYKA will going to work with the comparable BI-tool (confidential) in the future.

4.2.2 Examples of insights from the data connections

Connecting information from the ERP-system, MES-system, and machine data opens the possibility to combine information into data visualizations. This paragraph shows four examples, which show the flexibility, accessibility, understandability, and ease of computability of data in a BI-tool. Four examples are composed, the figures are presented in this paragraph, and screenshots of the creation process in Tableau are stated in appendix X. This is added to the appendix to show the ease of computing, this dynamic presentation is difficult to present in a static report.

Example 1: Potential of real-time alarms

The first example is stated in figure 24. This figure shows the potential of gathering manufacturing data and preventing downtime with alarms. This example shows that real-time alarms in the MES-system can prevent downtime. A BI-tool can be used to analyze this event if it is not prevented. For instance, the process engineer can analyze this situation and can train the operators using a data visualization or can request a change in the MES-system if problems have to be monitored in real-time in the future. For the case in figure 24, the MES-system must be developed to monitor the process parameters in real-time and the system should provide alarms if these parameters are not in between tolerances.

The frequency of this burning event is difficult to determine because of the inaccurate data in the current MES-system. But, an indication of the number of events is stated in appendix Y. Which indicates the number of ‘Reiniging’ (cleaning) production status declarations per machine during the three months of available data of the model. According to the data, there were fifty ‘Reiniging’ events during that three months, which are distributed over five production lines. Most cases could have been prevented if a data gathering system with alarms generation was in place at the department. This statement is validated by the process engineer of the extrusion.

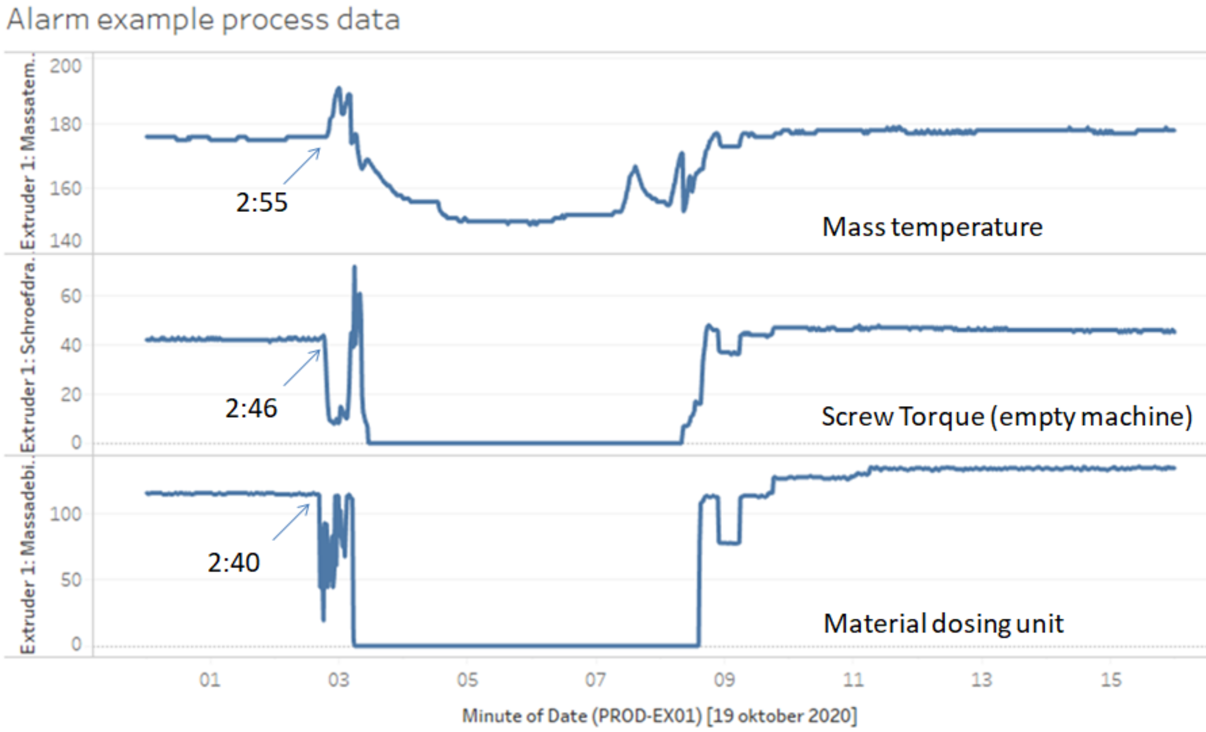


Figure 24: Example 1, process parameters, a situation that could be prevented with alarms

The die-head of the extrude (Ch. 3.1.2) is burned in this situation because the mass temperature is more than 10 minutes 5 degrees higher. But the problem started at the material dosing unit, where the material flow stopped. This resulted in an empty extruder machine (low screw torque). Subsequently, it resulted in the burning of the remaining material in the die-head. This equipment must be cleaned after this occurrence ('Reiniging' status in MES). In the case of this example, it resulted in six hours of downtime. This occurred at production line one, the downtime increases if this happens at a production line with a higher diameter, with up to approximately 24 hours. The burning of the material can be prevented in most cases. Also, in this case, the material dosing unit failed at 2:40, the operator could have solved probably the material dosing problem in the first five minutes, after that he/she had ten minutes to fill the machine with anti-burning material. Anti-burning material fills the extruder with material to prevent the burning of the remaining material. This PVC material with much-stabilizing material is used for changeovers and machine stops. This burning was probably prevented if an operator received an alarm message at the time that the material dosing unit was failing. There is a high chance of approximately 80% that a die-head burning event could have been prevented if an operator was informed in time. This statement is validated by the process engineer of the extrusion department. In conclusion, the operator had 15 minutes to solve the material dosing problem before the extruder die-head burned, which should have prevented six hours of downtime in this case. An alarming function in the MES-system will prevent these downtimes. If not, the gathered data can be analyzed in a BI-tool when this occurrence is not prevented.

Example 2: Combination of data from different sources

Example two shows information about the production result on diameter for a specific production line. This overview is created by combining quality data (MES-QC-EX), job data (MES-EX), and article data (ART-EX). These tables of the model are stated in figure 28 in chapter 4.3. All tables consist of the dimension Job, which is the connecting link between these tables. Example two is stated in figure 25, which contains the dashboard of quality parameters.

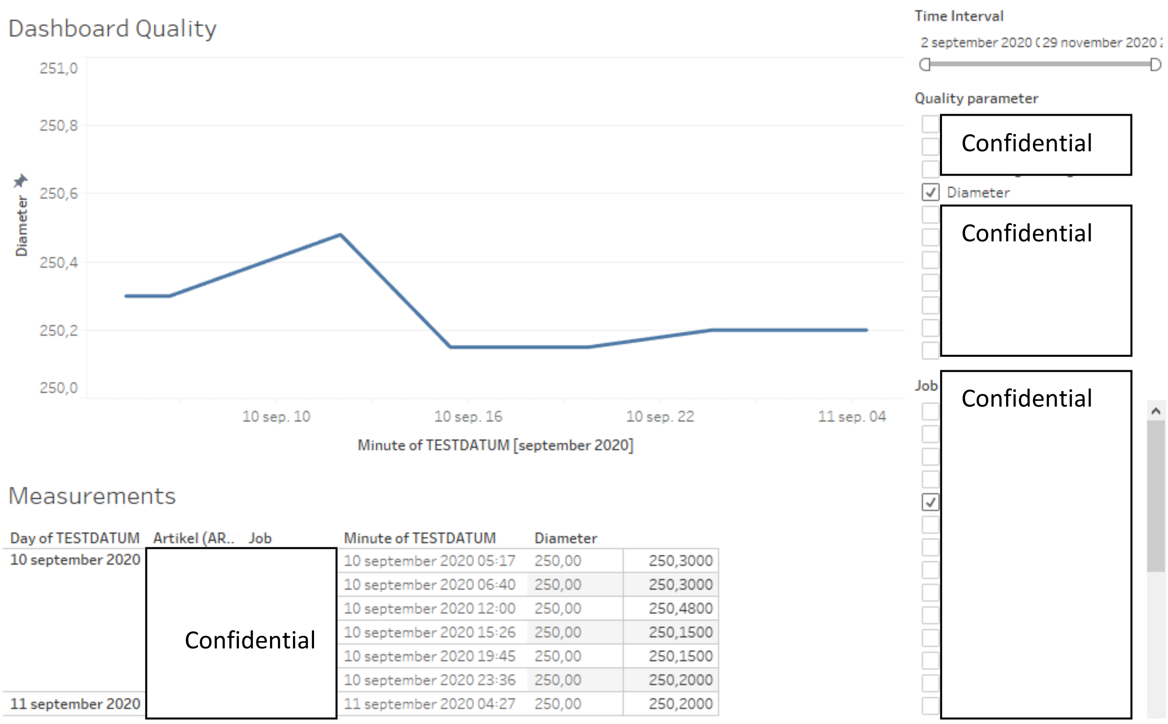


Figure 25: Example 2, the dashboard on the attained quality parameters per order

This example shows an overview of the attained diameter per production job at production line 14. The process engineer of the extrusion department could compile overviews about the quality result of production. The diameter is one example but the quality information of other quality parameters can be presented as well in this dashboard by changing the selected parameter in the filter. Combining data from different sources is difficult to compute currently. The data must be extracted from the different reporting tools of the different systems and then combined into one overview, which is computed mostly in excel. Standard dashboards as stated in figure 25 or other overviews on the quality result can automatically update the data, which enables the process engineer to monitor the quality results without information computing time.

Example 3, controlling the mass per meter performance indicator:
Example three presents the check on the attained mass per meter. This dashboard is stated in figure 26. This mass per meter is a performance indicator for the material costs of a pipe. This dashboard is created with the combination of process parameter data (SCADA-PROD-EX01), job data (MES- EX), quality data (MES-QC-EX), and process list data (MES-PROCESLIJST-EX).

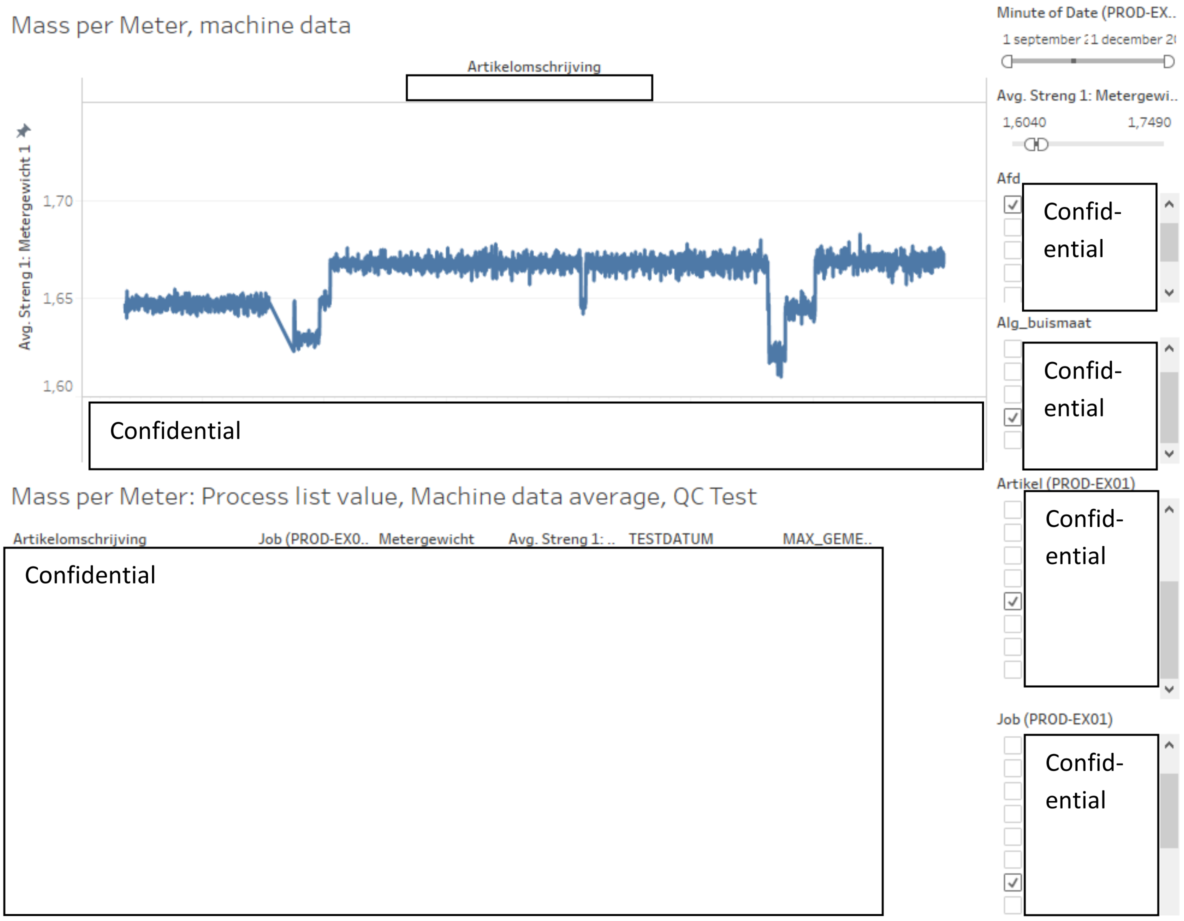


Figure 26: Example 3, the dashboard of the production result on the mass per meter

The initial product settings are calculated on 1,667 kilograms per meter. This initial setting can be monitored and updated. This KPI is important because the material counts up to 80% of the products, the other 20% are for example labor, production, and development costs. Gathering manufacturing data is required to control this performance indicator. Currently, the process engineer calculates this result from the manually added data on the mass per meter from the quality module of the MES-system and the process engineer validates this with sample checks. This manually added data is not always accurate, so the creation of this overview is a time-costly activity. Automatic process data gathering and standard overviews/dashboards in a BI-tool can substitute this time-consuming activity of creating information and will save time that can be invested in other projects.

Example 4, OEE dashboard

Example four presents an OEE dashboard that is made with production data (MES-EX) and production result data (ERP-EX). This dashboard is the result of a BI-tool test day with the data analyst operations. This day was organized in collaboration with the site director to prove the ease of modeling in the BI-tool and to enthusiasm the business controller for working with this tool. The business analysts of DYKA are working in excel currently. A BI-tool connected to databases will extend the available data processing capabilities and standard dashboards can be compiled to inform different stakeholders. An example of a dashboard is stated in figure 27. This OEE dashboard is created for shift leaders and managers to monitor the KPI OEE.

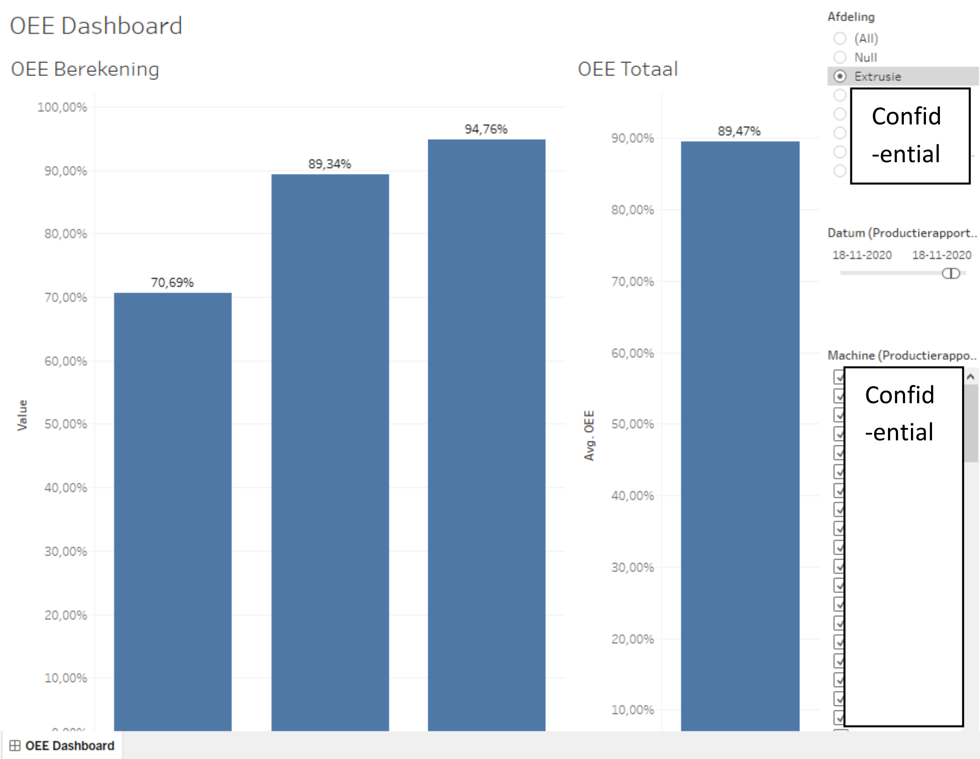


Figure 27: Example 4, OEE dashboard

This dashboard has three filters, the department, the time interval, and the particular production line. The users of this dashboard are not required to have the skill of analyzing data. They just have to understand and interpret the presented information in the dashboard. The dashboards and data overviews can be created by the business developer and the result of this creation can be made accessible for the users that require that information to fulfill its job. More about this process of creating data and creating data overviews is stated in figure 22 of chapter 4.1.2. For instance, this OEE information can help a shift leader in assigning operators to a specific production line that is performing under the standard.

In conclusion, different databases or a ‘data lake’ should be connected to a BI-tool to make data accessible and processable. Dashboards and production overviews can support the production managers, shift leaders, and staff to support decision-making. Besides the historical analyses, the MES-system should be designed to perform real-time monitoring of production, which is mainly used by the operators and shift leaders. This MES-system should be able to generate alarms if the process and quality parameters are not in between tolerances.

4.3 Data science method BI platform

Extending and automating data gathering will provide a large amount of available data. This data can be used for real-time and historical analysis. A Business intelligence (BI) platform can transform this data into information. Utilizing a BI-tool as a front dashboard of different databases results in the ability to combine data from different sources. This combination of data from different sources is created in this research with the use of the BI-tool Tableau. Data from machines (SCADA), production results (MES), and finished products (ERP) are connected in this tool. Aiming to provide insight into operations and showing what information can be created by connecting data. This is done according to the Extract-Transform-Load (ETL) process. The ETL process consists of extracting data from data sources, transforming and cleaning to prepare the data, and loading the data into a data warehouse (Jensen, Pedersen, & Thomsen, 2010).

4.3.1 Data extraction

Data extraction is about selecting and gathering data from the required sources. The research aims to connect data from the three ISA-95 layers, ERP, MES, and SCADA. The data is gathered from different database sources of information systems of DYKA and the machine parameters are gathered by extracting the data from the extruder machine with USB. There is not yet a SCADA-network in place at DYKA. Table 9 presents the ISA-95 layer of the data source, the gathered data, and the source of the data.

ISA-95 layer	Data	Source
ERP	Article information	Confidential
ERP	Produced packages (production history)	
MES	Production report per shift	
MES	Eventlog job changes	
MES	QC Quality information	
MES	Processlists / machine settings	
SCADA	Machine parameter data Extrusion line x	
SCADA	Machine parameter data Extrusion line x	

Table 9: Data gathered for research

This data is not complete, the data from the quality department is missing. This department gathers data in excel. This source is excluded from this research because of time barriers. The information in table 9 is used to provide advice about utilizing the BI-tool, this list of sources will be extended with more information in the future. The objective is to create a platform that can visualize most of the available information of the company. The gathered data from table 9 forms the data lake of the model in Tableau. The multiple variables of the different tables in this data lake are stated in appendix Z.

4.3.2 Data transformation

The transformation phase of the ETL process is the execution of a series of rules to transform the extracted data into standard formats (Chen, Mao, & Yunhao, 2014). This consist of the integration of data, cleaning of data, and the redundancy elimination of data (Chen, Mao, & Yunhao, 2014); the integration of data involves the combination of data from different sources and provides users a uniform view of data; the cleaning of data is the process of identifying inaccuracies, incomplete information, or unreasonable data and subsequently modifying or deleting data to improve the quality; redundancy elimination is the redundancy detection, data filtering, and data compression to decrease the unnecessary data transmission expense and defects in storage systems. An extensive explanation of the data preparation process is added to appendix AA.

The data must be prepared to create a cube structure. A cube is a multidimensional data structure that can be created for capturing and analyzing data, a collection of related cubes is referred to as a multidimensional database (Jensen, Pedersen, & Thomsen, 2010). The data from DYKA is connected by integrating the following dimensions: date-time, production job, Machine, and article. The remaining columns of de different data sources contain facts and measures. A fact is an object that represents the filter on the dimensions and the measures are the numerical facts of that selection (Jensen, Pedersen, & Thomsen, 2010). The transformation of the data presented in table 9 is executed in Rstudio. R is an open-source programming language and software environment, which is designed for data mining and visualization (Chen, Mao, & Yunhao, 2014)

The data from different sources is transformed to a situation that all connecting dimensions have the same column header, data format, and additional information as a production line and date-time. The extracted sources are CSV-files and the transformed tables are saved as Excel-files after the data processing in RStudio. These excel files are manually loaded into the BI-tool after the data transformation. The programming code of the data transformation in Rstudio states in appendix AB. This data preparation focused on the integration and cleaning data, less on the redundancy issue. This is chosen because Rstudio can handle the amount of data and the filters could be added in the BI-tool, which also can handle this amount of data.

An important activity in the data preparation process is data validation. The validation of ERP data is performed by combining data from the extrusion department (ERP) and data from the warehouse management system (WMS) of the storage and transport department. This validation step ensured that all data is captured in the gathering process. The validation of MES data was difficult because of the inaccuracy of MES data. The personnel of DYKA is not well trained to guarantee the accuracy of the manual filled data. The stop or running status of the production line is sometimes wrong, the job change event is sometimes not on the time added to the system, and the declaration of scrap is inaccurate. The job change problem is tested, but the others are not tested. This because of the time limits of the research and the focus of the research, which is not checking the declaration behavior of the personnel. The stop status declaration behavior is difficult to validate, but the scrap declarations can be checked by comparing the finished products data from ERP with the production output of MES. But this research only tests the job change moment, this problem is checked by comparing the job change moment in MES with the scan moment of the first package of a job in ERP minus the time to produce one package. The result of this comparison is presented in table 10. This is an estimation based on the available data of three months of production. This validation step is added to the model and is called MES-EX-Eventlog. The job changes in MES are too late in 13% of the change events.

Production line	EX-x1	EX-x2	EX-x3	EX-x4	EXx5	Total
Number of incorrect job declaration	7	5	9	1	4	26
Total job declarations 1-9-2020/1-12-2020	32	58	36	34	37	197
Percentage correct declarations	78%	91%	75%	97%	89%	87%

Table 10: Estimation of incorrect declarations of jobs

Next, The SCADA data from the machine is extended with the job, status, and article from MES. The date-time of both sources is connected to create a database with production and job data. The time on the machine is adjusted to real-time because the machines are not connected to the internet and are not updated manually. The job data is not always accurate as stated in table 10, but sufficient enough to provide insights from data, which is the aim of the data analysis. At the end of the data transformation, approximately 18,1 million data points are gathered, this is 85 MB of data. This is data form from three months, 94 percent of the data is from the production data of production lines 01 and 23. Gathering data from multiple systems and multiple machines is possible if the company provides enough data storage to increase the gathering of valuable production data. Consequently, this data can be utilized to solve problems to improve operational performance.

4.3.3 Data loading and visualization

The data warehouse is fed with new data in the loading step. The data is gathered cleaned and integrated before this step. The data sources from table 9 are inserted in Tableau and the different cubes are connected to create a multidimensional warehouse. The data source is connected by connecting the following indices: date-time, production line, job number, and article number dimensions. The data warehouse is visualized in figure 28. This is comparable with the functionalities in the information system architecture of figure 21. Figure 29 shows the resulting tables with the table of machine data EX-x unfolded. The connecting dimensions are recognizable in this figure, Date (date-time), Afd (production line), Job (job number), and Artikel (article). The process monitoring function is the center of the information structure (figure 21), similarly, MES-EX is the center of the model in the BI platform. The advice for DYKA is to implement a BI platform that can present data from different databases.

Confidential

Figure 28: Star diagram BI platform

Confid-ential

Figure 29: Example table structure BI

Multiple overviews and dashboards can be created with the use of data from different sources. All data is connected, so it is possible to combine information. The next subchapter provides information about utilizing a BI platform to present the available data.

4.4 Important big data decisions

Gathering data, using data for decision-making, and utilizing data to automate processes in IT-systems will create new challenges. The data quality must be guaranteed to validate the outcome of analyses and to ensure that information is the reflection of reality. This subchapter provides problems and challenges that have to be solved to create accurate and useful data. This information is created by combining the theory with practical examples, which are obtained during the data transformation and from interviews with stakeholders. These challenges and problems can be grouped into the four V's of big data decisions: volume, variety, velocity, and veracity. This theory is stated in chapter 2.4.2, where the article of He and Wang (2018) is the main article. This article is a literature review on big data challenges and decisions.

The characteristic volume is about making decisions in the number of variables and observations. These decisions influence the required storage space, increasing the number of variables of the number of observations per time interval will increase the required storage. DYKA should enlarge this capacity and gather all available data in the first place. Looking to the future, as much as possible data can be utilized in future correlation studies or can be utilized for condition-based maintenance. Later on, data variables can be deleted if variables are not valuable for analysis.

The characteristic variety deals with the format of the data. The structure of the gathered data can be different, this can be different formats or different units. For instance, process data, quality data, and metrology data can be presented in various data formats and units. The advice for DYKA is to process data into standard formats and units. For instance, date-time always in the same date-time format for all cubes and pipe length and diameter in meter or millimeter. Standardizing units and format will preventing wrong calculations and wrong conclusions. Additionally, it will prepare the data for future correlation studies between different gathered data.

The characteristic velocity is related to the speed of data generation. This can vary between real-time and data summarized in a batch of data. There are three types of velocity, batch mode, streaming mode, and mixed-mode. The batch data is analyzed parallel to operations, the data can be reported for historical analysis to optimize products and processes. Steaming data is used for analysis in real-time, this requires a higher data gathering speed. For instance, process monitoring and control by diagnostics and prognostics. A mixed-mode combines both data types into continuous data. It utilizes historical data that is summarized in longer intervals with real-time data. This mode can provide overviews, trends of parameters, and for example, execute predictive maintenance. The advice for DYKA is to gather the data in real-time and storage this real-time data per minute. This is in line with the data-gathering policy in the extruder machines. This data per minute must be stored to gather data for historical analysis in a BI-tool. The advice is to store all available data and make redundancy decisions after a pilot period. This prevents the removal of valuable data. The company must develop a system that preprocesses the data to facilitate real-time data that is accessible as fast as possible. The real-time data must be updated at a maximum of one minute to have valuable data for real-time alarming.

The characteristic veracity involves the quality or cleanness of the data. For instance the data outliers, noisy data, delayed data, or data asynchronism. These data cleanness problems may lead to misconclusions. But perfectly accurate data is the traditional approach of cleaning and solving data errors. Modern data analytics should consider these data errors and messiness as unavoidable. Large amounts of data will smoothen the outliers and errors. Reducing the data inconsistency and thereby increasing the data quality must be executed to create close to perfection data. Feature extraction is important to create consistent and valuable data. Many problems originate at the data gathering device, so data gathering sensors or machines should periodically being inspected to validate the data gathering process. For instance, by calibrating machines or IoT-sensors periodically. This problem has to be faced during the MES implementation project and the solvable problems must be solved. The more difficult problem can remain or can be solved later in the implementation process. 100% data veracity is not needed for big data analytics but should be reduced as much as possible to create high-value data.

The problems and challenging issues of the creation of the BI platform are stated in table 11. These challenges have to be solved during the MES implementation project at DYKA.

Big data V	Problem challenge of data at DYKA
Volume	Computing time of BI-tool can increase if the volume is too high
Volume	Current data is difficult to access
Variety	Data types in different systems are not the same
Variety	Data presentation can differ in different systems, e.g. job number or article number
Variety	Column headers of the same data are different in different information systems
Variety	Data units must be converted to standard units to combine data from different sources

Variety	Different machine manufacturers process and stored data in a different way, Theyson event-based and Kraussmaffei per minute
Velocity	Data in MES is stored per 8 hours, so the production result and parameter averages over 8 hours. This data is not suitable for extensive historical analysis
Veracity	Production line time is not real-time (can be automatically proper by connecting to the internet)
Veracity	Temperature data does not have the right format
Veracity	There are data outliers in the machine data
Veracity	There is missing machine data
Veracity	Data in MES is inaccurate due to the incorrect declaration behavior of the workers
Veracity	Static information is not updated accurately (machine speed, material lists)

Table 11: Challenges for big data gathering found during the data analysis

Most variety and veracity problems and challenges are solvable with a proper ETL-process, this process is stated in the next subchapter. The velocity and volume problems and challenges can be solved by proper hardware selection. The project of data gathering will face problems and challenges. Validation of the gathered data is an important project step, problems should be solved or accepted if it is not a critical problem. In conclusion, start the data-gathering project in the MES-system and evaluate the data quality continuously.

4.5 Continuation of the project

Becoming a vertical integrator is a major challenge and the implementation will not be a straightforward project, both organizationally and technically challenges have to be solved to obtain this objective. Chapter 4 provides information that should guide this development and implementation process. The data model in the BI-tool of chapters 4.2 and 4.3 provides information about the future situation of data availability and accessibility. The MES development and BI development will support the company to obtain the vertical integrator digital maturity level. The digital transition will gain insight into the production performance and the obtained quality of production by creating overviews and dashboards in a BI-tool. In addition to historical data utilization, real-time data utilization will support operations to monitor and control production.

Evaluation of the BI-tool

This future state of data gathering and visualization is provided in the BI-tool. Data from ERP, MES, and machinery are connected to present this future digital state of data advancement. The model is presented and evaluated during individual evaluation sessions with different future users from different departments and functions. The objective of these evaluation sessions was to provide insight into possible future utilization of IT-systems, to provide insight into the compatibility of a BI-tool, and to encourage participation in the future project. Focusing on this research, the objective was to convince and motivate the company that the digital transition is required to increase the enterprise's performance. The cases stated in chapter 4.2.2 are presented during the evaluation sessions and the participant was asked which data/information would improve their daily work. An overview of evaluation participants and their future data usage is stated in appendix AC. The objective of the evaluation is achieved. The enthusiasm about data gathering and presenting in a BI-tool will be converted into real data transition projects. One is the development of MES and the other project is developing and implementing the use of a BI-tool to update and control the reported portfolio. The enthusiasm about the model in Tableau has convinced the management to prioritize the BI project. DYKA will be going to develop its BI in a BI-tool This choice is made because the parent organization, Tessenderlo Group, has chosen to use this BI-tool as standard.

In conclusion, the future situation in the BI-model has motivated DYKA to adopt a BI-tool and the evaluation has enthusiasm the future users of this BI-tool. This research has contributed to convincing the management about the importance of this part of the digital transition

Roadmap to data supported manufacturing environment

The foundation for operational improvement measurement has to be built to monitor and control the performance of the company. This foundation can be created by utilizing the current data and extending this data with process data. DYKA should search for MES-software that supports the ideas about monitoring and control operations, which is stated in chapter 4.1.3. The digital transition of the company is a comprehensive project that needs the involvement of different future users. The project will have an impact on the operators, staff, and management. The involvement of stakeholders in the digital transition of manufacturing is important. The change has to be managed by a change strategy and this strategy of people's involvement had to be evaluated continuously. The data/information will be utilized throughout the organization. Obtaining the vertical integrator maturity level will become the strategy for the next few years. Data can be used to create value from the shop floor layer to the enterprise resource planning layer. This vertical integration of the information system architecture supported with a BI-tool will support DYKA to grow to the vertical integrator maturity level in which the company can monitor and control its processes. In contrast to the current situation, where the manufacturing process is partly a black box. The implementation steps during the digital transition of DYKA are visualized in a roadmap towards data-supported manufacturing, presented in figure 30.

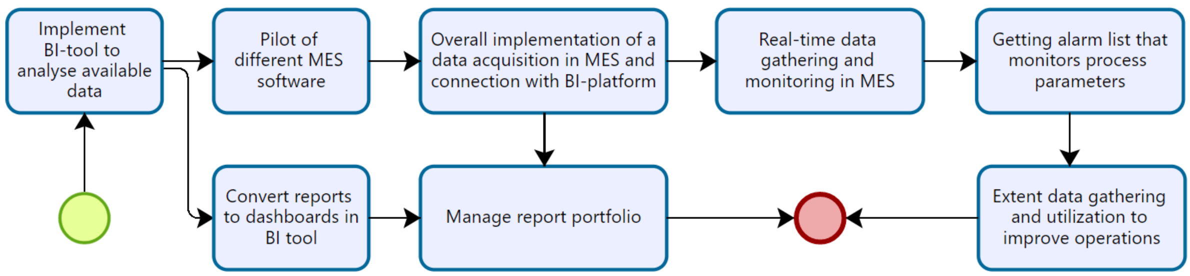


Figure 30: The roadmap to data supported manufacturing

This roadmap consists of two major projects, the implementation of MES and the implementation of a BI-tool. Both projects must contribute to the data availability and accessibility objective of the organization. The current MES and ERP data can be connected in the BI-tool already. The data available for the BI-tool will be extended with real-time process parameters and quality parameters after the MES implementation project. The reporting portfolio of the company should be updated continuously to increase the amount of performance information. The creation of dashboards and overviews in the BI-tool can support decision-making processes throughout the company. The gathered data should be transformed into valuable information to monitor, control, and improve operations.

4.6 Conclusion

This chapter answers the research question about the evolving opportunities and the design requirements for the creation of a supportive manufacturing execution system and a big data platform. The evolving opportunities are the ability to monitor and control operations in real-time and utilizing data for historical analysis. The company has to develop part of its information system architecture. Firstly, the available data must be connected and presented in a BI-tool to make data available, computable, and accessible. Secondly, the MES system must be developed with more features than currently. The future MES-system should at least contain the following functions of the MESA-11 model: operational/detailed sequencing, resource allocation and status, document control, data collection/acquisition, process management, quality management, and performance management. The MES-system should gather process and quality data and should use this data to monitor and control parameters in real-time. In addition to this, the process lists and the inspection lists should also be managed in the MES-system. The extended MES-system should be connected to the data warehouse of the BI-tool. This gathering and connection of data will transform the company into the vertical integrator maturity level.

5. Conclusion and Discussion

This chapter provides the conclusion and discussion of the research, in which the conclusions, reflections, limitations, and recommendations are presented. The research aims to provide information about the opportunities that evolve when real-time data is used to monitor and control the processes and gathered big data is used for historical and real-time analyses to improve operational performance. This aim is transformed into the following research question:

> **Research question**: How should a production company/DYKA design its information system architecture to monitor, control, and improve operational performance and be ready for future technologies?

The research question consists of four parts: designing the information systems architecture, monitoring and controlling operational performance, improving operational performance, and forming the foundation for future technologies. The literature review answers this question for general production companies and the case study answers this question specifically for the case company DYKA.

5.1 Conclusion

Firstly, the conclusion on the general development of the design of the information system architecture. This architecture must support the company in monitoring controlling and improving its performance. This can be obtained by developing and extending information systems. The company has to focus its strategy on the digital transition of operations. The current digital state can be measured in digital maturity. DYKA is in the digital novice state according to the maturity model of PWC. This means that DYKA has a partial integration of IT-systems and analytical capabilities are mainly based on semi-manual data extracts. DYKA has to develop its IT-systems into the vertical integrator maturity level. So, DYKA has to develop its IT architecture from the first stage into the second stage of the maturity model stated in figure 31. This means that DYKA has a vertical integrated IT architecture, machine-to-machine network, and analytical capabilities in a central BI-system. The vertical integration can be described as the integration of the shop floor (SCADA-layer) with the MES- and ERP-system. This integration of IT-systems is described in the ISA-95 standard, which is a standard for IT-system functions and integration between functions. The vertical integration of IT-systems is required to monitor and control operations in real-time. Subsequently, a BI-tool is connected to the data of these systems to perform historical analytics to improve operations. The vertical integration of operations can be attained by performing three steps, these are the overall implementation of a data acquisition platform, real-time data collection and monitoring, and generating an alarm list that monitors the production parameters.

Figure 31: Stages of digital maturity (PWC, Geissbauer, Vedso, & Schrauf, 2016)

In the second place, the developed information system architecture should support operations by monitoring and controlling the operational performance. The gathering of shop floor data in MES will support the operators in advising and alarming when the process or quality parameters are not in between tolerances. The MES-system should consist of the following features of the Mesa-11 model to obtain this goal: operational detailed sequencing, resource allocation and status, document control, data collection/acquisition, process management, quality management, and performance analysis. Monitoring at least the important process parameters will prevent unnecessary downtimes and will result in the required output quality. The important parameters are mass temperature, mass pressure, screw torque, material dosing, line speed, and mass per meter. These parameters can be extended by the quality parameters, which can be automated as well in the future. The operational process is partly a black box and production relies almost entirely on the experience and know-how of operators. These future users are important stakeholders for the success of the MES development project. Management of change strategy should be created, which should contain ideas about the involvement of personnel. Monitoring and controlling the process parameters and quality parameters with real-time information and alarms will enable the company to envision its processes.

Thirdly, the gathered data can be analyzed to create more information about the performance of processes. The data should be extracted from different sources, transformed into the right data format and cleanness, and loaded into a data lake connected to a BI-tool. This ETL-process should ensure the value of the data before the creation of dashboards and reports. The created reports and dashboards should be validated and added to the reporting portfolio of the company. The research provided examples of reports and a dashboard created in the BI-tool Tableau. DYKA is going to utilize the BI-tool (confidential) because this tool will be the standard BI-tool of the parent organization Tessenderlo group. Data understandability and accessibility will enlarge the amount of available information, which can support discussions, co-operations, improvements, and innovations.

The fourth point is the foundation for future technologies. Modern technologies create and rely on data, so gathering and utilizing data is the key to make new industry 4.0 technologies accessible. The manufacturing environment should be transformed into an industrial internet of things (IIoT) to create smart manufacturing. Information-supported decision-making will be created by the development of an interconnected manufacturing facility. But, the first step of creating this environment is vertical integration. This connectivity of IT and technology systems is the gateway toward other promising technologies.

In conclusion, DYKA should focus on two projects that form the digital transition strategy. One project is the development of the manufacturing execution system, where the company has to find a software vendor that can deliver a system that is in line with the requirements of the company. This MES-system has to be developed into a data gathering and visualization tool that supervises operations, intending to create a manufacturing intelligence environment. The second project is developing a strategy towards data utilization in a BI-tool. This project consists of designing a proper ETL-process and researching the potential usage of the data by the creation of dashboards and reports. Both separate projects will improve the current state of utilizing available data, which will contribute to incremental steps towards a higher digital maturity.

5.2 Discussion

The discussion consists of three parts the reflection, limitations, and recommendation. This subchapter describes the significance of the findings and forms inspiration for future work.

Reflection

The goal of the research was to provide information about the development of the information system architecture and convincing the management of DYKA to focus on this digital transition to monitor, control, and improve operations. This goal is achieved because the development of MES and BI is part of the companies development projects in 2021. The research provided knowledge from the literature and insights from the data analysis. These two building blocks of the research have provided sufficient information about the required future state of IT-systems at DYKA. Reflecting on the academic goals of the research, the aim was to bridge the gap between the current academic research and the current state of IT-systems within production companies. This knowledge is created in the research and the knowledge is also generalizable for other companies. But the academic goal is achieved partly. The body of knowledge about bridging the current state of companies and the current academic focus is provided but is not specific enough to guide companies through the implementation process. The knowledge provides insights about challenges, but less on experiences during an implementation project. This information should be added to complete the case study.

Limitations

This lack of implementation experiences in the case study is the limitation of the project. The literature review provides information about the development of information system architectures, but the case study does not test these findings. The case study should be extended with the implementation project findings to provide a complete overview of challenges during such a project. The findings of the research are based on literature and the vision of the company. Another limitation of the research was that the current MES-system was not updated with the currently available features of the software company. A pilot of the future system with the current MES software wasn't possible because of this. A limitation within the research was that the data was not reliable to form an accurate overview of the current performance of the company. Therefore the research focused on examples and not on the real impact of an IT development project. Although the research could have been extended with the calculation of the impact of this development by taking the data quality for granted. This data analysis could have provided more information about the reduced downtime, increased quality, and possibly attained production performance. This is not researched because of the time limits of the project. On the other hand, this impact study was not required because the management of DYKA is already convinced that this development will deliver an increase in information and the performance of the company subsequently.

Recommendations

The knowledge base for bridging the gap between the current state of companies and the current academic research field will be enlarged if the evaluation of the case study is added to the research. Future research about attaining the vertical integrator maturity level should evaluate these implementation projects. The combination of theory, vision, and evaluation will provide a complete overview of the important decisions and challenges. Another research direction for future research is the broadening of the theoretical construct. This research focused more on providing basic knowledge, but there is a lot of research focusing on IT development within production companies. For instance elaborating research on the industry 4.0 integration standard RAMI4.0, this standard extent the ISA-95 method focusing on interconnecting people with IT and technology. Or for instance, focusing on The Open Group Architecture Framework (TOGAF) for the development of the specific business, IT, and technology architecture. The data model can also be extended with more information to create an interconnected company. This will be researched within the company in the next few years. For instance, adding data from the quality department or labor-management within the data model. This research forms the basis of knowledge and future insights, which can be extended into the theoretical direction and the practical direction in the future.

In summary, forming the data foundation is the first step toward manufacturing intelligence for DYKA. This foundation will support the company in monitoring, controlling, and improving the operational performance and forms the base for the adoption of technology in the future.

References

Abdel-Fattah, M., Helmy, Y., & Hassan, T. (2019). A Strategic Roadmap for Maximizing Big Data Return. *International Journal of Integrated Engineering*, 47-60.

Akerman, M. (2018). *Implementing shop floor IT for industry 4.0.* Gothenburg: Chalmers Reproservice.

Alukal, G. (2007). Lean Kaizen in the 21th Century. *Quality Progress*, 69-70.

APICS. (2017). Performance Management. In APICS, *Supply Chain Operations Reference Model* (pp. 374-380). Chicago IL USA: APICS.

Beerda, H., Bering, R., Schulte, H., ten Veen, S., Quist, D., van der Wal, M., et al. (2020). *Reader PVC-extrusie 2020, Maatwerk DYKA Steenwijk.* Steenwijk: vapro.

Beric, D., Stefanovic, D., Lalic, B., & Cosic, I. (2018). The implementation of ERP and MES Systems as a Support to Industrial Management Systems. *International Journal of Industrial Engineering and Management*, 77-86.

Borisovich, F., & Htun, Z. (2019). Integrated Production System using ERP and MES. *International Conference of Intelligent Robotic and Control Engineering*, 32-36.

Chen, M., Mao, S., & Yunhao, L. (2014). Big Data: A Survey. *Mobile Network Applications*, 171-209.

Clausen, P., Mathiasen, J., & Nielsen, J. (2020). Smart Manufacturing Through Digital Shop Floor Management Boards. *Springer Science+Business Media, LLC*, 1-14.

Corrales, L., Lambán, M., Korner, M., & Royo, J. (2020). Overall Equipment Effectiveness: Systematic Literature Review and Overview of Different Approaches. *Applied Sciences*, 1-20.

D'Antonio, G., Macheda, L., Bedolla, J., & Chiaber, P. (2017). PLM-MES Integration to Support Industry 4.0. *Product Lifecycle Management and the Industry of the Future* (pp. 129-137). Seville, Spain: Springer, 14th IFIP WG 5.1 International Conference.

de Carolis, A., Macchi, M., Negri, E., & Terzi, A. (2017). A Maturity Model for Assessing the Digital Readiness of Manufacturing Companies. *IFIP International Federation for Information Processing*, 13-20.

Erboz, G. (2017). How to define industry 4.0: the main pillars of industry 4.0. *Conference: Managerial trends in the development of enterprises in the globalization era*, 761-767.

Erlicher, L., & Massone, L. (2005). Human Factors in Manufacturing: New Patterns of Cooperation for Company Governance and the Management of Change. *Human Factors and Ergonomics in Manufacturing*, 403-419.

Farooqui, A., Bengtsson, K., Falkman, P., & Fabian, M. (2019). From factory floor to process models: A data gathering approach to generate, transform, and visualize manufacturing processes. *CIRP Journal of Manufacturing Science and Technology*, 6-16.

Foster, D., White, L., Adams, J., Erdil, D., Hyman, H., Kurkovsky, S., et al. (2018). Cloud computing: developing contemporary computer science curriculum for a cloud-first future. *Conference: Proceedings Companion of the 23rd Annual ACM*, 130-147.

Fucheng, P., Haibo, S., & Bin, D. (2015). Manufacturing execution system present situation and development trend analysis. *International conference on information and automation*, 535-540.

Gartner. (2020, 02 11). *Magic quadrant for analytics and business intelligence platforms*. Retrieved 02 09, 2021, from garntner.com: https://www.gartner.com/doc/reprints?id=1-1YAE9AY1&ct=200206&st=sb

Gifford, C., & Daff, D. (2020, 11 30). *ISA-95 evolves to support smart manufacturing and IIoT*. Retrieved 11 30, 2020, from International Society of Automation: https://www.isa.org/intech-home/2017/november-december/features/isa-95-to-support-smart-manufacturing-iiot

Hart, C. (1999). *Doing a Literature Review: Releasing the Research Imagination.* London: SAGE publications Ltd, London.

He, Q., & Wang, J. (2018). Statistical process monitoring as a big data analytics tool for smart manufacturing. *Journal of Process Control*, 35-43.

Heerkens, H., & Van Winden, A. (2012). *Geen Probleem, een aanpak voor alle bedrijfskundige vragen en mysteries.* Buren: Business School Nederland.

Henderson, J., & Venkatraman, N. (1993). Strategic alignment: leveraging information technology for transforming organisations. *IBM Systems Journal*, 472-484.

Hon, K. (2008). Performance and Evaluation of Manufacturing Systems. *Department of Engineering, University of Liverpool*, 1-16.

Howson, C., Richardson, J., Sallam, R., & Kronz, A. (2019). Magic Quadrant for Analytics and Business Intelligence Platforms. *Gartner*, 1-58.

Industry Directions. (2004). Meeting the manufacturing challenge, Performance Advantage of MES. *Strategic Direction*, 28-30.

Jensen, C., Pedersen, T., & Thomsen, C. (2010). *Multidimensional Databases and Data Warehousing.* Waterloo, Ontario, Canada: Morgan and Claypool.

Kamble, S., Gunasekaran, A., Ghadge, A., & Raut, R. (2020). A performance measurement system for industry 4.0 enabled smart manufacturing system in SMMEs- A review and empirical investigation. *International Journal of Production Economics*, 1-15.

Kannan, S., Suri, K., Cadavid, J., Barosan, I., van den Brand, M., & Alferez, M. G. (2017). Towards Industry 4.0: Gap analysis between current automotive MES and industry standards using model-based requirement engineering. *IEEE International Conference on Software Architecture Workshops*, 29-35.

Lee, H., Kim, Y., & Kim, K. (2017). Implementation of an efficient big data collection platform for smart manufacturing. *Journal of engineering and applied sciences*, 6304-6307.

Lesjak, C., Druml, N., Matischek, R., & Ruprechter, T. H. (2016). Security in industrial Iot - quo vadis? *Springer verlag Wien, Elektrotechnik & Informationstechnik*, 324-329.

Llave, M. (2018). Data lakes in business intelligence: reporting from the trenches. *International conference on enterprise information systems*, 516-524.

Meieran, E. S. (1993). Intelligent manufacturing systems. *Proceedings of 15th IEEE/CHMT International Electronic Manufacturing Technology Symposium*, 323-327.

Michniewicz, J., & Reinhart, G. (2014). Cyber-physical Robotics – Automated Analysis, Programming and Configuration of Robot Cells based on Cyber-physical-systems. *Procedia Technology*, 566-575.

Mladineo, M., Veza, I., Gjeldum, N., Crnjac, M., Aljinovic, A., & Basic, A. (2019). Integration and testing of the RFID-enabled Smart Factory concept within the learning factory. *9th conference on learning factories*, 384-389.

Modrák, V., & Mandul'ák, J. (2009). Mapping Development of MES Functionalities. *International Conference on Informatics in Control, Automation, and Robotics*, 244-247.

Moeuf, A., Pellerin, R., Lamouri, S., Tamayo-Giraldo, S., & Barbaray, R. (2017). The industrial management of SMEs in the era of Industry 4.0. *International Journal of Production Research*, 1118-1136.

Muchiri, P., & Pintelon, L. (2008). Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion. *International Journal of Production Research*, 3571-3535.

Obitko, M., & Jirkovský, V. (2015). Big Data Semantics in Industry 4.0 . *In: Mařík V., Schirrmann A., Trentesaux D., Vrba P. (eds) Industrial Applications of Holonic and Multi-Agent Systems.*, 217-229.

Oesterreich, T., & Teuteberg, F. (2016). Understanding the implications of digitalization and automation in the context of Industry 4.0: A triangulation approach and elements of research agenda for the construction industry. *Computers in Industry*, 121-139.

Peffers, K., Tuunanen, T., Rothenberger, M., & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, 45-77.

Piszczek, K., Tomaszewska, J., & Sterzynski, T. (2010). The influence of temperature of poly(vinyl chloride) melt on the equilibrium state of gelation process. *Polimery*, 106-110.

Prades, L., Romero, F., Estruch, A., García-Dominguez, A., & Serrano, J. (2013). Defining a methodology to design and implement business process models in BPNM according to the standard ANSI/ISA-95 in a manufacturing enterprise. *Proccedia Engineering*, 115-122.

PWC, Geissbauer, R., Vedso, J., & Schrauf, S. (2016). Industry 4.0: Building the digital enterprise. *Global Industry 4.0 Survey by PWC*, 1-36.

Qin, S. (2014). Process Data Analytics in the Era of Big Data. *American Institute of Chemical Engineers, Wiley Online Library*, 3092-3100.

Ray, P., Wang, W., Chang, J., & Hubona, G. (2011). Management of Change to Ensure IS Success: a Longitudinal Study. *Proceedings of the 2011 International Conference on Information Systems Conference, Shanghai, China*, 1-20.

Rollepaal. (2019). *About pipe extrusion*. Retrieved 2 2, 2021, from rollepaalinc.com: https://www.rollepaalinc.com/introduction.html

Rüßmann, M., Lorenz, M., Gerbert, P., Waldner, M., Justus, J., Engel, P., et al. (2015). Industry 4.0: the future of productivity and growth in manufacturing industries. *Publication of BCG Consulting*, 1-14.

Saabye, H., Kristensen, T., & Wæhrens, B. (2020). Real-Time Data Utilization Barriers to Improving Production Performance: An In-depth Case Study Linking Lean Management and Industry 4.0 from a Learning Organization Perspective. *Sustainability 2020*, 1-19.

Saturno, M., Pertel, V., Deschamps, F., & de Freitas Rocha Loures, E. (2017). Proposal of an automation solutions architecture for industry 4.0. *International Conference on Production Research*, 1-7.

Schumacher, A., Erol, S., & Sihn, W. (2016). A maturity model for assessing Industry 4.0 readiness and maturity of manufacturing enterprises. *Procedia CIRP*, 161-166.

Schumacher, A., Nemeth, T., & Shin, W. (2018). Roadmapping towards industrial digitalization based on an Industry 4.0 maturity model for manufacturing enterprises. *12th CIRP Conference on Intelligent Computation in Manufacturing Engineering*, 409-414.

TechTarget. (2020, 01). *MES (Manufacturing Execution System)*. Retrieved 02 2021, from LeMagIT: https://www.lemagit.fr/definition/MES-Manufacturing-Execution-System

Telukdarie, A. (2016). MES to ERP Integration: Rapid Deployment Toolset. *IEEE International conference on industrial engineering and engineering management*, 1030-1035.

Templier, M., & Paré, G. (2015). A Framework for Guiding and Evaluating Literature Reviews. *Communications of the Association for Information Systems*, 112-137.

The Open Group. (2019). *ArchiMate® 3.1 Specification, a standard of The Open Group*. Retrieved 02 22, 2021, from ArchiMate® 3.1 Specification: https://pubs.opengroup.org/architecture/archimate3-doc/toc.html

Turban, E., Sharda, R., Delen, D., & King, D. (2010). *Business Intelligence, a Managerial Approach.* Upper Saddle River, New Jersey, United States of America: Pearson Education.

Unver, H. (2013). An ISA-95-based manufacturing intelligence system in support of lean initiatives. *Advanced Manufacturing Technology*, 853-866.

von Wangenheim, C., Hauck, J., Salviano, C., & von Wangenheim, A. (2010). Systematic Literature Review of Software Process Capability/Maturity Models. *Proceedings of International Conference on Software Process. Improvement And Capability Determination (SPICE)*, 1-9.

White, S. (2004). Introduction to BPMN. *BPTrends*, 1-11.

Wu, P. (2011). A Mixed Methods Approach to Technology Acceptance Research. *Journal of the AIS*, 1-17.

Zhang, H., & Wang, L. (2020). Design of Data Acquisition Platform for Industrial Internet of Things. *2020 IEEE 3rd International Conference on Information Systems and Computer-Aided Education (ICISCAE)*, 613-618.

Appendices:

The Digital Transition of Manufacturing

The gathering and visualization of real-time data and historical data to monitor, control, and improve manufacturing performance

G.D. Schutte

24-02-2021

Master Thesis

Industrial Engineering and Management

UNIVERSITY OF TWENTE.

DYKA

Table of contents

References

APICS. (2017). Performance Management. In APICS, *Supply Chain Operations Reference Model* (pp. 374-380). Chicago IL USA: APICS.

Chen, M., Mao, S., & Yunhao, L. (2014). Big Data: A Survey. *Mobile Network Applications*, 171-209.

de Carolis, A., Macchi, M., Negri, E., & Terzi, A. (2017). A Maturity Model for Assessing the Digital Readiness of Manufacturing Companies. *IFIP International Federation for Information Processing*, 13-20.

Epicor. (2021). *What Is a Manufacturing Execution System (MES)?* Retrieved 02 23, 2021, from Epicor: https://www.epicor.com/en/resource-center/articles/what-is-a-manufacturing-execution-system/

He, Q., & Wang, J. (2018). Statistical process monitoring as a big data analytics tool for smart manufacturing. *Journal of Process Control*, 35-43.

Kakade, S. (2017, 11). *Manufacturing execution system (MES)*. Retrieved 02 23, 2021, from TechTarget: https://searcherp.techtarget.com/definition/manufacturing-execution-system-MES

Kamble, S., Gunasekaran, A., Ghadge, A., & Raut, R. (2020). A performance measurement system for industry 4.0 enabled smart manufacturing system in SMMEs- A review and empirical investigation. *International Journal of Production Economics*, 1-15.

Muchiri, P., & Pintelon, L. (2008). Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion. *International Journal of Production Research*, 3571-3535.

PWC, Geissbauer, R., Vedso, J., & Schrauf, S. (2016). Industry 4.0: Building the digital enterprise. *Global Industry 4.0 Survey by PWC*, 1-36.

Qin, S. (2014). Process Data Analytics in the Era of Big Data. *American Institute of Chemical Engineers, Wiley Online Library*, 3092-3100.

Raptis, T., Passarella, A., & Conti, M. (2019). Data Management in Industry 4.0: State of the Art and Open Challenges. *IEEE Access*, 97052-97093.

Rout, T. (2003). ISO/IEC 15504 - Evolution to an International Standard. *Software process improvement and practice*, 27-40.

von Wangenheim, C., Hauck, J., Salviano, C., & von Wangenheim, A. (2010). Systematic Literature Review of Software Process Capability/Maturity Models. *Proceedings of International Conference on Software Process. Improvement And Capability Determination (SPICE)*, 1-9.

Appendix A: Search strategy in Scopus

General explanations

- The following rules are used for inclusion and exclusion: Selection on the relevance of the title. Focus on production companies and software systems. Also searched for articles that will be relevant for the other subquestions

Sub-chapter	Keywords search terms	Articles after step 2	Articles after step 3	Articles after step 4	Explanation, decisions, and interesting points
2.2	TITLE-ABS-KEY ("data gathering" AND "manufacturing ")	194	28	6 (also used for other questions)	Opening the way for Industry 4.0 technologies was the main objective of most selected articles. The first step for being able to implement these technologies is data gathering. 2.2 is about the technologies, but the basis of the search is to find reasons and new opportunities emerging from gathering more data. Those opportunities are the horizon/level that can be reached in the future after forming the basis for those technologies.
2.2	Search in references of previous search action. Searching for the root article of fundamentals of industry 4.0				There are many interpretations of industry 4.0. Many articles refer to the explanation of the Boston Consulting Group, this source is used as the main pillars in this research
	TITLE ("data collection" AND "manufacturing ")	39	7	1	To prevent missing information, this synonym of data gathering is used. Search on the title is used to decrease the number of articles. The conclusion from this comparison is that data gathering and data collection have the same definition, the researcher of the article chooses one of these two words to describe the acquisition and saving of data.
2.3	TITLE-ABS-KEY ("industry 4.0" AND "maturity model")	124	5	3	Industry 4.0 technologies can be implemented after a company has obtained a certain level of digital maturity. Digital maturity is the link between a proper system information architecture and the industry 4.0 technologies
2.3	TITLE-ABS-KEY("isa-95" and "industry 4.0")	22	4	1	ISA-95 is the standard for software implementation. The relation between ISA-95 and industry 4.0 provides information about an implementation that will prepare a company for new

					technologies.
2.4	TITLE-ABS-KEY ("real-time data" AND monitor* AND manufacturing)	262	9	1	Vertical integration is important for preparing for new technologies. But the other objective of this integration is to obtain higher performances by using real-time data to support operations. Information about how to design and what are the challenges must be included in the literature search
2.4	TITLE-ABS-KEY("real-time data" and visuali* and manufacturing)	44	16	2	These keywords result in articles about how to design big data platforms. This includes interesting decisions to make in the design phase of such an information system
2.5	TITLE ("overall equipment effectiveness" AND performance)	22	3	1	Once a data gathering system is designed, this information is used to measure performance and to use this data in smart reporting. This search focuses on the manufacturing KPI
2.5	Search for literature review on production performance measurement				

Appendix B: Articles found in the systematic literature review

#	From search	Part of	Title	Authors	Year	Keywords	Journal	Aim
1	TITLE-ABS-KEY ("data gathering" AND "manufacturing ")	2.2.1 and 2.5.1	Real-Time Data Utilization Barriers to Improving Production Performance: An In-depth Case Study Linking Lean Management and Industry 4.0 from a Learning Organization Perspective	Saabye, H., Kristensen, T., & Wæhrens, B.	2020	Lean management, Industry 4.0, learning organization, enabling formalization, real-time data, OEE	Sustainability Conference 2020	Linking Industry 4.0 to Lean management principles
2	TITLE-ABS-KEY ("data gathering" AND "manufacturing ")	2, 2.2, 2.2.1,	From factory floor to process models: A data gathering approach to generate, transform, and visualize manufacturing processes.	Farooqui, A., Bengtsson, K., Falkman, P., & Fabian, M.	2019	Data acquisition, process mining, industrial robots, manufacturing systems, pattern recognition	CIRP Journal of Manufacturing Science and Technology	Development of an architecture that enables capturing robot operation data from existing and new manufacturing stations
3	TITLE-ABS-KEY ("data gathering" AND "manufacturing ")	2.3.2, 2.3.3	Integration and testing of the RFID-enabled Smart Factory concept within the learning factory.	Mladineo, M., Veza, I., Gjeldum, N., Crnjac, M., Aljinovic, A., & Basic, A.	2019	Learning factory, smart factory, RFID	9th conference on learning factories	Implementing a test smart factory, with RFID-enabled technology and integrated within the learning factory. A smart factory based on the integration of MES with the shop-floor level and de the ERP-system at the other side

4	TITLE-ABS-KEY ("data gathering" AND "manufacturing ")	2.3.3, 2.4	MES to ERP Integration: Rapid Deployment Toolset.	Telukdarie, A.	2016	Manufacturing Execution Systems (MES), Systems, Evaluation tools	IEEE International conference on industrial engineering and engineering management	The research delivers a method and case study on an internationally benchmarked, express evaluation toolset, with capacity to conduct a Business Unit (BU) evaluation in minimum time. Key value adds of the toolset includes system prioritization on a business benefit and cost basis.
5	TITLE-ABS-KEY ("data gathering" AND "manufacturing ")	2.2, 2.2.1	Big Data Semantics in Industry 4.0	Obitko, M., & Jirkovský, V.	2015	Data management, industrial networks, manufacturing, Industry 4.0	Industrial Applications of Holonic and Multi-Agent Systems.	The article can help the readers to deeply understand how data management is currently applied in networked industrial environments, and select interesting open research opportunities to pursue
6	TITLE-ABS-KEY ("data gathering" AND "manufacturing ")	2.2	Intelligent manufacturing systems	Meieran, E. S.	1993	Manufacturing Strategy Development	Proceedings of 15th IEEE/CHMT International Electronic Manufacturing Technology Symposium	Introducing new or strategic technology and applications into operating factories is not a trivial task. It requires a lot of considerations to improve factory performance into manufacturing systems technologies

7	TITLE ("data collection" AND "manufacturing "), also in TITLE-ABS-KEY ("data gathering" AND "manufacturing ")	2.4.2	Implementation of an efficient big data collection platform for smart manufacturing	Lee, H., Kim, Y., & Kim, K.	2017	Big data, big data collection platform, smart manufacturing, analysis, performance, effective	Journal of engineering and applied sciences	the study implements an effective big data collection platform for smart manufacturing
8	TITLE-ABS-KEY ("industry 4.0" AND "maturity model")	2.2.2, 2.3.1	Roadmapping towards industrial digitalization based on an Industry 4.0 maturity model for manufacturing enterprises	Schumacher, A., Nemeth, T., & Shin, W.	2018	Digital maturity, Industry 4.0	12th CIRP Conference on Intellig-ent Comp-utation in Manufact-uring Engineering	In the paper, a holistic procedure model that guides manufacturing companies from their first contact with industry 4.0 until the definition of concrete action field and their realization
9	TITLE-ABS-KEY ("industry 4.0" AND "maturity model")	2.3	A Maturity Model for Assessing the Digital Readiness of Manufacturing Companies.	de Carolis, A., Macchi, M., Negri, E., & Terzi, A.	2017	Digital readiness, Maturity model, Digital transformation, Industry 4.0	IFIP Internation al Federation for Informatio n Processing	This paper wants to illustrate a “tool” to answer this question by building a maturity assessment method to measure the digital readiness of manufacturing firms. Based on the inspiring principles of the CMMI (Capability Maturity Model Integration) framework

10	TITLE-ABS-KEY ("industry 4.0" AND "maturity model")	2.2.3, 2.3.1,	Roadmapping towards industrial digitalization based on an Industry 4.0 maturity model for manufacturing enterprises	Schumacher, A., Erol, S., & Sihn, W.	2016	Industry 4.0, Maturity Model, Strategic Management, Change Management	Procedia CIRP	The main goal of the paper is to extend the dominating technology focus of recently developed models by including organizational aspects. There are defined 9 dimensions and assigned 62 items to them for assessing Industry 4.0 maturity
11	TITLE-ABS-KEY("isa-95" and "industry 4.0")	2.3.3	Towards Industry 4.0: Gap analysis between current automotive MES and industry standards using model-based requirement engineering.	Kannan, S., Suri, K., Cadavid, J., Barosan, I., van den Brand, M., & Alferez, M. G.	2017	Model-based development, requirement modeling, MES, Industry 4.0, Factories of the Future, ISA-95, RAMI 4.0	IEEE International Conference on Software Architecture Workshops	The paper presents a model-based requirements engineering approach along with a gap analysis process. The work is mainly divided into three phases, (i) automotive MES tool selection phase, (ii) requirements modeling phase, (iii) and gap analysis phase based on the modeled requirements
12	TITLE-ABS-KEY("real-time data" and monitor* and manufacturing)	2.4, 2.4.1, 2.4.2, 2.4.3, 2.6	Statistical process monitoring as a big data analytics tool for smart manufacturing	He, Q., & Wang, J.	2018	Statistical process monitoring, Big data, Smart manufacturing, Feature extraction, Internet of things	Journal of Process Control	a roadmap of statistical process monitoring

13	TITLE-ABS-KEY("real-time data" and visuali* and manufacturing)	2.4.1, 2.7.2	Design of Data Acquisition Platform for Industrial Internet of Things	Zhang, H., & Wang, L.	2020	Industrial internet of thing, data collection, interconnection	2020 IEEE 3rd International Conference on Information Systems and Computer Aided Education	This paper designs the data acquisition platform of the industrial Internet of things. The overall structure of the platform includes the infrastructure, the corresponding communication mode, and the necessary software
14	TITLE-ABS-KEY("real-time data" and visuali* and manufacturing)	2.4.1	Smart Manufacturing Through Digital Shop Floor Management Boards	Clausen, P., Mathiasen, J., & Nielsen, J.	2020	Shop floor management, Industry 4.0, smart manufacturing, digital shop floor management boards	Springer Science+Business Media, LLC	This paper reveals that digital SFM boards have not yet been adopted at the shop floor level, and currently, practitioners are stuck to the standardized procedures and manual processes. The forces against further adaptation are a managerial mindset stuck in an Industry 2.0 era and immature technologies to digitize the visualization of real-time data. The forces for are the need of enhancing data transparency within and across teams, which means the elimination of information silos and time-consuming manual updates of SFM boards.

15	TITLE ("overall equipment effectiveness" AND performan ce)	2.5.3	Performance measurement using overall equipment effectiveness (OEE): literature review and practical application discussion.	Muchiri, P., & Pintelon, L.	2008	Performance measurement, overall equipment effectiveness, manufacturing	International Journal of Production Research	In this paper, overall equipment effectiveness (OEE) is described as one such performance measurement tool that measures different types of production losses and indicates areas of process improvement

Appendix C: Maturity levels CMMI and ISO/IEC 15504 (SPICE)

Interesting citations about CMMI and SPICE:
Our results show that there exist a large variety of models with a trend to the specialization of those models for specific domains. We also identified that most of those models are concentrated around the CMM/CMMI framework and the standard ISO/IEC 15504 (SPICE) (von Wangenheim, Hauck, Salviano, & von Wangenheim, 2010)

ISO/IEC 15504 (SPICE), a standard for software process assessment, maturing in process management in systems and software engineering. Capability levels: Performed process, managed process, established process, predictable process, optimizing process. Definitions in appendix xd (Rout, 2003).

CMMI, Capability Maturity Model Integration. Assess the different dimensions of the company separately. Different dimensions are used to assess 5 areas in which manufacturing key processes can be grouped: (1) design and engineering, (2) production management, (3) quality management, (4) maintenance management, and (5) logistics management. Assess from low to high. Maturity levels: Initial, Managed, Defined, integrated and interoperable, digital-oriented (de Carolis, Macchi, Negri, & Terzi, 2017)

CMMI maturity levels (de Carolis, Macchi, Negri, & Terzi, 2017):

Table 1. Maturity levels' definition

Maturity level	Description
ML1 INITIAL	The process is poorly controlled or not controlled at all, process management is reactive and does not have the proper organizational and technological "tools" for building an infrastructure that will allow repeatability /usability /extensibility of the utilized solutions
ML2 MANAGED	The process is partially planned and implemented. Process management is weak due to lacks in the organization and/or enabling technologies. The choices are driven by specific objectives of single projects of integration and/or by the experience of the planner, which demonstrates a partial maturity in managing the infrastructure development
ML3 DEFINED	The process is defined thanks to the planning and the implementation of good practices and management procedures. The management of the process is limited by some constraints on the organizational responsibilities and /or on the enabling technologies. Therefore, the planning and the implementation of the process highlights some gaps/ lacks of integration and interoperability in the applications and in the information exchange
ML4 INTEGRATED AND INTEROPERABLE	Being the process built on the integration and on the interoperability of some applications and on the information exchange, it is fully planned and implemented. The integration and the interoperability are based on common and shared standards within the company, borrowed from intra- and/or cross-industry de facto standard, with respect to the best practices in industry in both the spheres of the organization and enabling technologies
ML5 DIGITAL-ORIENTED	The process is digital oriented and is based on a solid technology infrastructure and on a high potential growth organization, which supports – through high level of integration and interoperability – speed, robustness and security in information exchange, in collaboration among the company functions and in the decision making

SPICE Maturity levels (von Wangenheim, Hauck, Salviano, & von Wangenheim, 2010):

Table 2. Revised measurement framework

Capability level	Technical Report	International Standard
Level 1 – Performed process	PA 1.1 – Process performance attribute	PA 1.1 – Process performance attribute
Level 2 – Managed process	PA 2.1 – Performance management attribute	PA 2.1 – Performance management attribute
	PA 2.2 – Work product management attribute	PA 2.2 – Work product management attribute
Level 3 – Established process	PA 3.1 – Process definition attribute	PA 3.1 – Process definition attribute
	PA 3.2 – Process resource attribute	PA 3.2 – Process deployment attribute
Level 4 – Predictable process	PS 4.1 – Process measurement attribute	PA 4.1 – Process measurement attribute
	PA 4.2 – Process control attribute	PA 4.2 – Process control attribute
Level 5 – Optimising process	PA 5.1 – Process change attribute	PA 5.1 – Process innovation attribute
	PA 5.2 – Continuous improvement attribute	PA 5.2 – Process optimisation attribute

Appendix D: Maturity model, and industry 4.0 PWC

PwC maturity model - Industry 4.0 capabilities develop across seven dimensions and four stages

	1 Digital novice	2 Vertical integrator	3 Horizontal collaborator	4 Digital champion
Digital business models and customer access	First digital solutions and isolated applications	Digital product and service portfolio with software, network (M2M) and data as key differentiator	Integrated customer solutions across supply chain boundaries, collaboration with external partners	Development of new disruptive business models with innovative product and service portfolio, lot size 1
Digitisation of product and service offerings	Online presence is separated from offline channels, product focus instead of customer focus	Multi-channel distribution with integrated use of online and offline channels; data analytics deployed, e. g. for personalisation	Individualised customer approach and interaction together with value-chain partners. Shared, integrated interfaces.	Integrated Customer Journey Management across all digital marketing and sales channels with customer empathy and CRM
Digitisation and integration of vertical and horizontal value chains	Digitised and automated sub processes. Partial integration including production or with internal and external partners. Standard processes for collaboration partly in place	Vertical digitisation and standardised and harmonised internal processes and data flows within the company; limited integration with external partners	Horizontal integration of processes and data flows with customers and external partners, intensive data use through full integration across the network.	Fully digitised, integrated partner ecosystem with self-optimised, virtualised processes, focus on core competency; decentralised autonomy. Near real-time access to extended set of operative information
Data & Analytics as core capability	Analytical capabilities mainly based on semi-manual data extracts; Selected monitoring and data processing, no event management	Analytical capabilities supported by central business intelligence (BI) system Isolated, not standardised decision support systems	Central BI system consolidating all relevant internal and external information sources, some predictive analytics Specific decision support and event management systems	Central use of predictive analytics for real-time optimisation and automated event handling with intelligent database and self-learning algorithm enabling impact analysis and decision support
Agile IT architecture	Fragmented IT architecture in-house.	Homogeneous IT architecture in-house. Connection between different data cubes developing.	Common IT architectures in partner network. Interconnected single data lake with high-performance architecture	Single data lake with external data integration functionalities and flexible organisation. Partner service bus, secure data exchange
Compliance, security, legal & tax	Traditional structures, digitisation not in focus	Digital challenges recognised but not comprehensively addressed	Legal risk consistently addressed with collaboration partners,	Optimising the value-chain network for compliance, security, legal and tax
Organisation, employees and digital culture	Functional focus in "silos"	Cross-functional collaboration but not structured and consistently performed	Collaboration across company boundaries, culture and encouragement of sharing	Collaboration as a key value driver

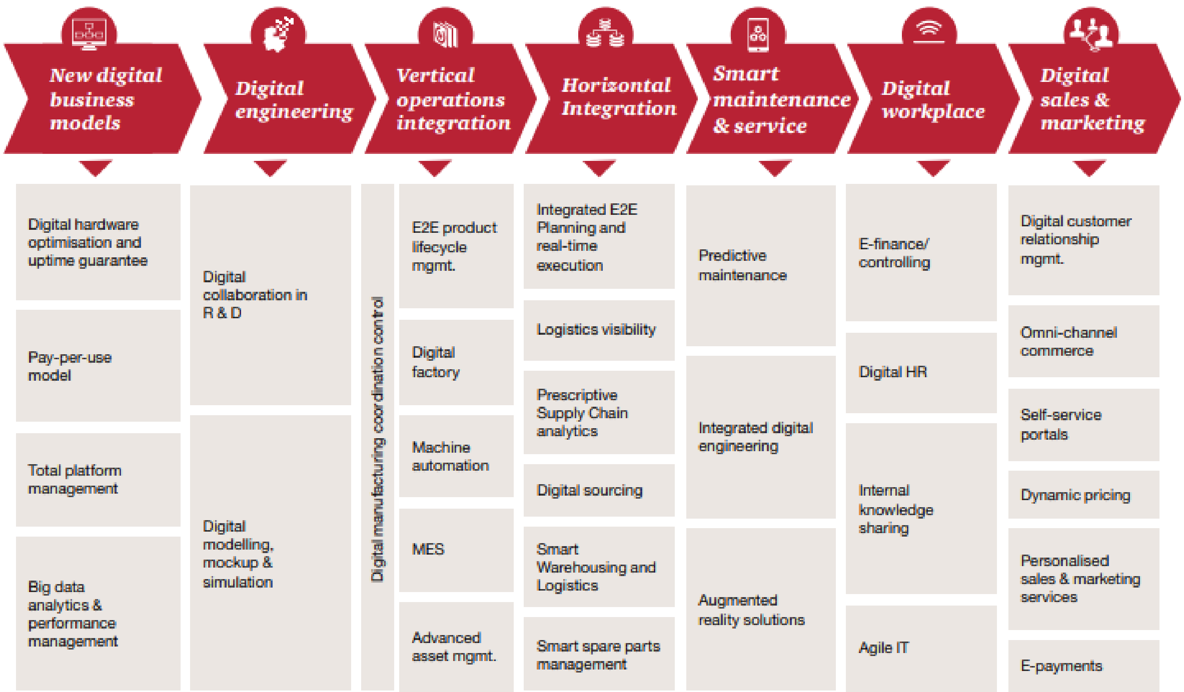
(PWC, Geissbauer, Vedso, & Schrauf, 2016)

How Industry 4.0 is delivering revenue, cost and efficiency gains

Additional revenue from:	Lower cost and greater efficiency from:
Digitising products and services within the existing portfolio	Real-time inline quality control based on Big Data Analytics
New digital products, services and solutions	Modular, flexible and customer-tailored production concepts
Offering big data and analytics as a service.	Real-time visibility into process and product variance, augmented reality and optimisation by data analytics
Personalised products and mass customisation.	Predictive maintenance on key assets using predictive algorithms to optimise repair and maintenance schedules and improve asset uptime
Capturing high-margin business through improved customer insight from data analytics	Vertical integration from sensors through MES to real-time production planning for better machine utilisation and faster throughput times
Increasing market share of core products	Horizontal integration, as well as track-and-trace of products for better inventory performance and reduced logistics
	Digitisation and automation of processes for a smarter use of human resources and higher operations speed
	System based, real-time end-to-end planning and horizontal collaboration using cloud based planning platforms for execution optimisation
	Increased scale from increased market share of core products

(PWC, Geissbauer, Vedso, & Schrauf, 2016)

Industry 4.0 pilot opportunities exist along the full vertical and horizontal operational value chains



(PWC, Geissbauer, Vedso, & Schrauf, 2016)

Appendix E: MESA-11 model description

Figure E1, MESA-11 (Kakade, 2017)

Explanation of functions of MES and the application to the DYKA case: (Epicor, 2021):

Operational sequencing: Offer your employees a global view of planned production orders and their production routing. This ensures your entire staff is on the same page and reduces errors due to miscommunication.
DYKA: Operational planning module

Dispatching production units: Manage the bidirectional flow of production data in real-time between the ERP and the workshop. This ensures production data is always accurate, consistent, and up to date.
DYKA: flow (routing), of materials to resources, material inventory management

Product tracking and genealogy: Group final parts or batches with all their corresponding manufacturing data—from the raw material to the component assembly. This data is especially useful for manufacturers that must comply with government or industry regulations.
DYKA: Information per pipe, information assigned to one specific pipe

Labor management: Easily manage your people, products, and/or operations and track any skills or authorizations they require. This ensures that you always have the right people in place at each step of the production process.
DYKA: Assigning people to specific tasks

Quality management: Manage the quality of your manufacturing process and units—including quality deviations and exceptions. This function can be integrated directly into the MES software or can use external software.

DYKA: Controlling automatic and manual gathered data

Maintenance management: More easily and accurately plan preventative machine maintenance to reduce downtime and production interruptions.
DYKA: Interfacing with production and maintenance department

Data collection and acquisition: Track and gather essential data and easily recover that data when you need it.
DYKA: Shopfloor data gathering, process data, and quality data

Process management: Provide process routing and operational sequencing—including full production traceability.
DYKA: Monitoring function of MES. Which checks the operational performance on process and quality parameters and provides alarms

Performance analysis: Consolidate data to calculate key performance indicators (KPIs) like rework, scrap, process capability, OEE, and more. This lets you know how your production process is working and how it could be improved.
DYKA: connection with the BI-tool, which is another IT project of DYKA

Document control: Provide a simple way for your operators to access important documents—including instructions, drawings, notes, and more—when they need them. This saves you and your employees time by not having to search through file cabinets for the information you need.
DYKA: Administration and updating of process lists

Resource allocation and status: Define and track the status of your resources and how they are used in the production process.
DYKA: Tooling management of DYKA

Appendix F: Interesting analytics after the creation of a big data platform

This research question was part of the project plan but is out of scope because of the focus on the creation of connected data (the big data platform).

What are potential future projects to improve the overall manufacturing process efficiency focusing on process mining and analytics?

General remarks/citations about analytics:

- Big data analytics offer many opportunities to evaluate data in all layers of the industrial installations, for example, to identify preferences from end-users, to better understand technological enablers' behaviors, or to relate issues derived from combined and statistical processing of data (Raptis, Passarella, & Conti, 2019)

- Data analytics can be viewed as the science and engineering of examining data to uncover hidden patterns, unknown correlations, and other useful information that can be used to make better decisions or to develop effective solutions. (He & Wang, 2018)
- Emerging opportunities: Prognostics, anomalies detection, fault diagnosis, multi-agent systems, decision making, job scheduling, energy management, machine learning, big data analytics, data semantics (relations), human-in-the-loop services, camera/vision, augmented reality, virtual reality security, energy management, the cloud. (Raptis, Passarella, & Conti, 2019)

The three generations of Statistical process control according to (He & Wang, 2018)**:**

1. Statistical process control (SPC) using mean and variance (quality control based on product quality distribution)
2. Multivariate statistical process monitoring (MSPM) extending variance with the covariance of both product quality variables and process variables, this added information allows the detection of those faults that cannot be detected by SPC methods, therefore enables significantly improved monitoring performance (exploring correlations among process and quality variables)
3. Directly address special characteristics (e.g., dynamics, nonlinearity); self-adaptive; etc.

Data mining techniques

- Cluster Analysis: is a statistical method for grouping objects, and specifically, classifying objects according to some features (Chen, Mao, & Yunhao, 2014)
- Correlation Analysis: is an analytical method for determining the law of relations, such as correlation, correlative dependence, and mutual restriction, among observed phenomena and accordingly conducting forecast and control (Chen, Mao, & Yunhao, 2014)
- Regression Analysis: is a mathematical tool for revealing correlations between one variable and several other variables (Chen, Mao, & Yunhao, 2014)
- A/B Testing: also called bucket testing. It is a technology for determining how to improve target variables by comparing the tested group (Chen, Mao, & Yunhao, 2014)
- Statistical Analysis: Statistical analysis is based on the statistical theory, a branch of applied mathematics (Chen, Mao, & Yunhao, 2014)
- Data Mining Algorithms: Data mining is a process for extracting hidden, unknown, but potentially useful information and knowledge from massive, incomplete, noisy, fuzzy, and random data (Chen, Mao, & Yunhao, 2014)
- Indexing (query by content): Given a query time series and some similarity/dissimilarity measurement, it finds the nearest matching time series in the database. (Chen, Mao, & Yunhao, 2014)
- Classification: Given an unlabeled time series, it assigned it to one of two or more predefined classes. (Chen, Mao, & Yunhao, 2014)
- Segmentation: Given a time series containing plenty of data points, it constructs a model with smaller piecewise segments such that the latter closely approximates the former. This representation makes the storage, transmission, and computation of the data more efficient. Specifically, in the context of data mining, piecewise linear representation is used.

- Dynamic time warping: This measures the distance between two-time series after first aligning them in the time axis (Qin, 2014).
- Metamodeling: The use of data-analytics techniques to generate surrogate or reduced-dimension process models that are accurate and computationally efficient (Qin, 2014).
- Derivative-free optimization: Optimization based on techniques that make use of data or their effective representation (Qin, 2014).
- Optimization under uncertainty: The characterization of uncertainty with data-driven machine-learning tools to enhance optimization techniques to work with uncertainty (Qin, 2014).

Appendix G: Explanation and information of the 4V's of big data

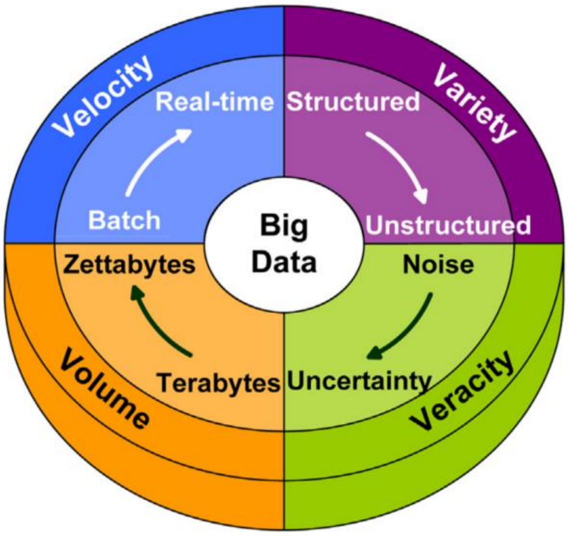


Figure x, decision range of the 4V's (He & Wang, 2018)

More information about the 4V's,
Interesting citations to consider during an implementation project
Volume:
- A few key features of the process can be identified to capture the important process characteristics, then monitoring these key features can not only effectively address the large volume of the big data, but also provide enhanced monitoring performance by capturing the important process characteristics (e.g., nonlinear behavior or special dynamics) that are embedded among different process variables. (He & Wang, 2018)
- The size of the data to be circulated in a network environment is of crucial importance to the network design and the technological enablers used in the deployment. In an industrial networked environment, there can be a diversity of data volumes, depending on the scope of each use case (Raptis, Passarella, & Conti, 2019)

Variety:
- Data variety is a measure of the richness of the data representation – text, images video, audio, etc. (Abdel-Fattah, Helmy, & Hassan, 2019)

- Manufacturing operations generate a different form of data, such as process or machine data and product quality or metrology data, each could take different forms, monitor different parts of the system, measure different phases of the process. (He & Wang, 2018)

Velocity:

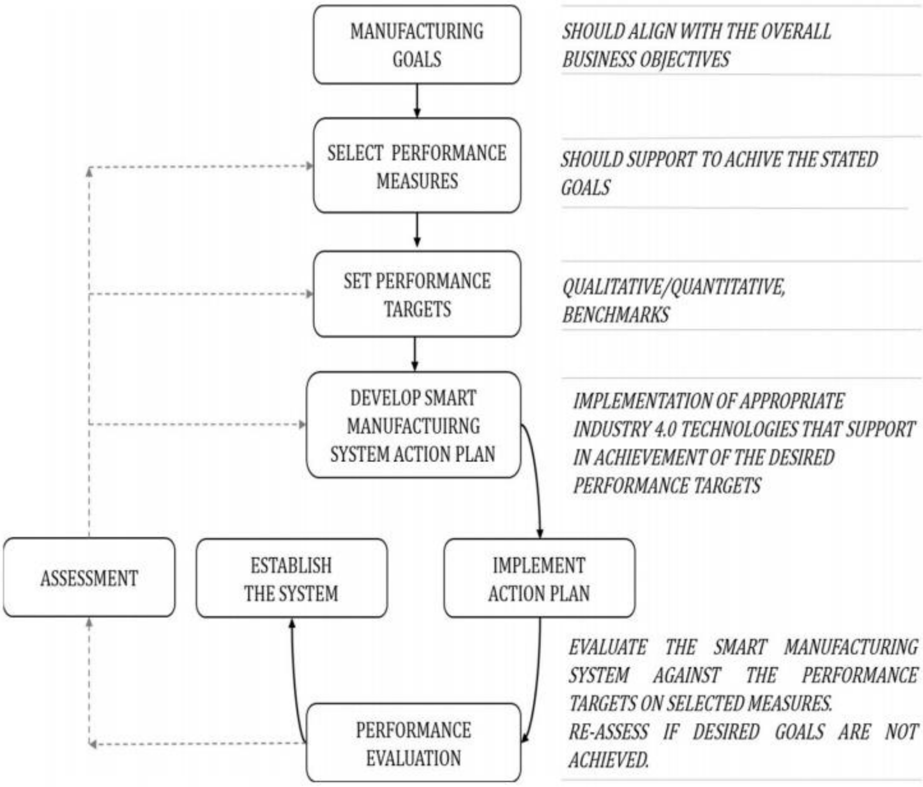
- In the era of big data, there will be at least three modes of data analytics: streaming mode, batch mode, and mixed-mode. It is expected that different modes of data analytics will be used for different purposes, through developing different types of models. For example, batch models could be used for optimizing product/process design; streaming or online models for process monitoring (e.g., diagnostics and prognostics) and control; and mixed-mode models for predictive maintenance. (He & Wang, 2018)
- velocity is the speed of data generation (real-time data) (Akerman, 2018)
- The streaming model considers the data flowing in as a steady stream. When new data arrive, they are processed immediately. The Batch-mode is originally designed to accomplish parallel processing of a large amount of data through largescale, low-cost server clusters. The third mode is mixed, improve the processing ability for continuous data. (Qin, 2014)
- To address the large volume of streaming data for real-time statistical analysis and online monitoring, effective incremental or iterative modeling approaches will be needed (He & Wang, 2018)

Veracity :

- In the process industry, veracity means data quality or cleanness issues such as missing data, outliers, noises, delays, and data asynchronism (He & Wang, 2018)
- While the traditional MSPM emphasize the cleanness of the data to prevent potentially misleading conclusions, it has been suggested that the next-generation data analytics tools should consider data errors or messiness as unavoidable, and use massive data to develop solutions that are robust to the imperfections in the data (He & Wang, 2018)
- This inherent importance separates the data into two major categories, critical and non-critical data. We label the first category as da-ta of high Criticality and the second category as data of low criticality (Raptis, Passarella, & Conti, 2019).

Appendix H: Designing a performance measurement portfolio

Designing a manufacturing performance portfolio



(Kamble, Gunasekaran, Ghadge, & Raut, 2020)

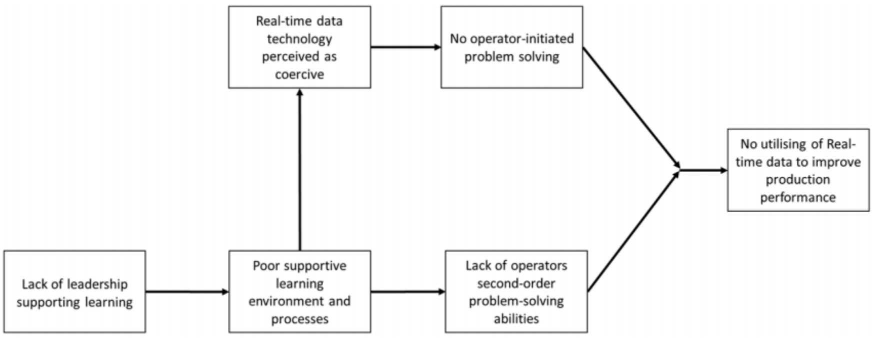


Figure 3. Barriers to utilizing real-time operational data in a manufacturing context, from a LO perspective.

Appendix I: Overall equipment effectiveness information

More information about each specific loss of the OEE (Muchiri & Pintelon, 2008):

Downtime losses:

(1) Breakdown losses are categorized as time losses and quantity losses caused by equipment failure or breakdown. For example, a breakdown of palletizing plant motor in a brewery leads to downtime and thus production loss.

(2) Set-up and adjustment losses occur when production is changing over from the requirement of one item to another. In a brewery plant, this type of loss is encountered during set-ups between different products, testing during start-ups, and fine-tuning of machines and instruments.

Speed losses:

(3) Idling and minor stoppage losses occur when production is interrupted by a temporary malfunction or when a machine is idling. For example, dirty photocells on palletizing machines cause minor stoppages. Though they are quickly fixed, much capacity is lost due to their frequency.

(4) Reduced speed losses refer to the difference between equipment design speed and actual operating speed. In a palletizing plant, the use of unadapted pallets leads to longer processing times for the same number of bottles leading to speed losses.

Quality losses:

(5) Quality defects and rework are losses in quality caused by malfunctioning production equipment. For example, some pallet types get stuck in between depalletizer and unpacker and are damaged.

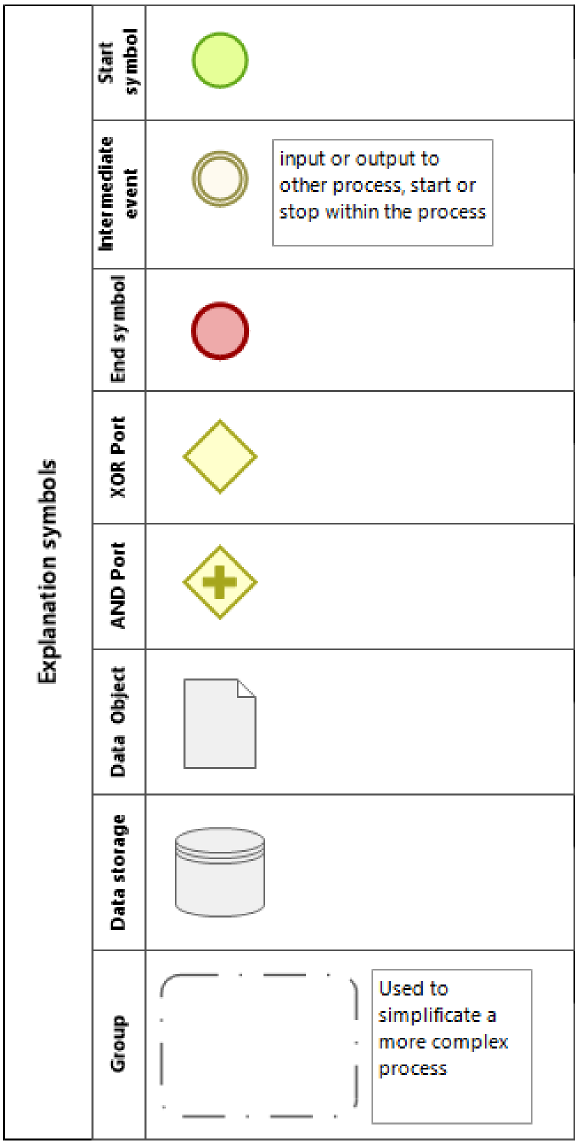
(6) Reduced yield during start-up is yield losses that occur from machine startup to stabilization. For example, in the brewery, poor preparation for morning shift by night shift leads to problems with the filling taps and thus leads to reduced yields.

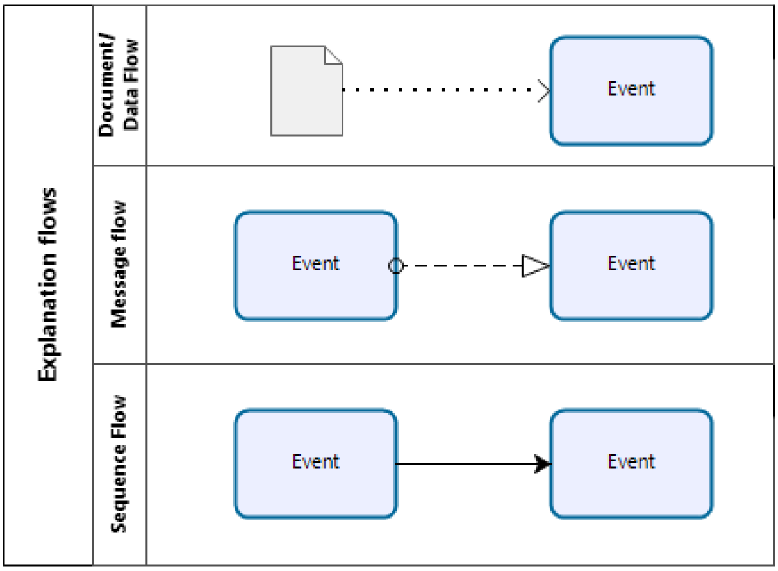
Appendix J: Reporting system updating APICS

(APICS, 2017) updating the reports in the report-portfolio should be a continuous process

Step	Activity	Result and input next step
0	Performance measurement problem	Set reporting requirements, risk monitoring
1	Initiate reporting on performance measurements	Quarterly, Daily, monthly, weekly, annually, real-time reports
2	Analyze reports, receive and plan maintenance requests.	Detailed performance gap
3	Find root causes, root cause finding methods, and techniques include: - Adding commentary to reported data - Metrics decomposition using diagnostic relationships of metrics - Time studies, sampling, audits - 5 Why's/Cause and effect analysis - Statistical Analysis Techniques: Histogram, Scatter Plots ANOVA	The root cause, data requirement, or a data analysis requirement
4	Maintenance request, IT request, or prioritizing of root causes. The process of sorting root causes by relative contribution and prioritizing root causes.	Prioritized Root Cause and Performance root causes
5	Develop corrective actions	Corrective action, Skill/Resource Gap (identify Skill/resource requirement or determine training/education), Capacity Gab (schedule asset management activities), Network reconfiguration opportunity (select scope of the project), Network gap (develop strategy and plan)
6	Approve and launch corrective action. The process of obtaining approvals, prioritizing, communicating, and launching the corrective actions	reporting requirements, restart the cycle at step 1

Appendix K: BPMN symbols and flows explanation





Appendix L: Production lines and manufacturer

Confidential

Appendix M: Organizational chart of the department

Confidential

Appendix N: Example process list

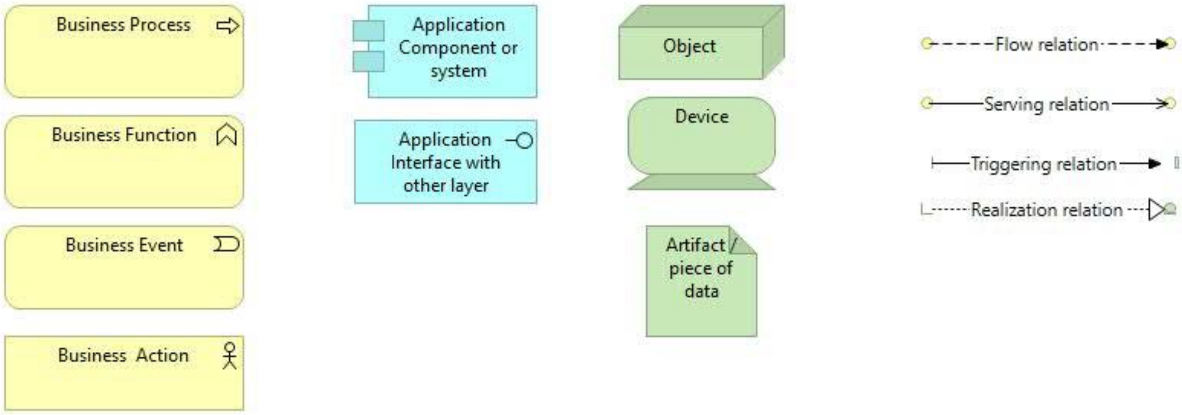
Confidential

Appendix O Downtime alarms from production line 19

Confidential

Appendix P: Archimate Model and modeling explanation

Confidential


Business Process
Business Function
Business Event
Business Action
Application Component or system
Application Interface with other layer
Object
Device
Artifact / piece of data
Flow relation
Serving relation
Triggering relation
Realization relation

Appendix Q: Figures of current MES-system applications

Confidential

Figure xf1, planboard dashboard

Confidential

Appendix R: Declaration statuses, active, downtime, quality

Confidential

Confidential

Appendix S: Current OEE calculations

Confidential

Confidential

Confidential

Confidential

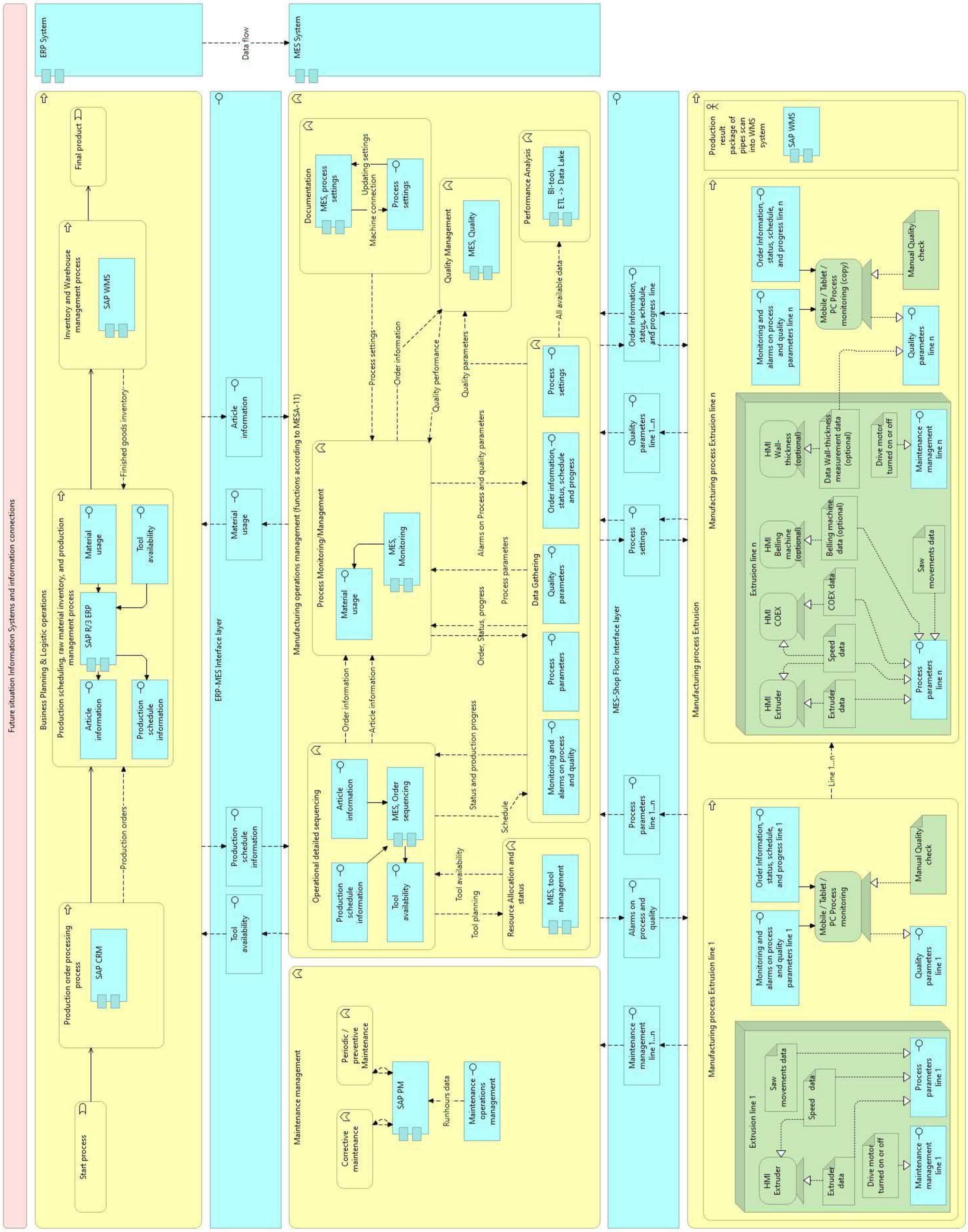
Appendix T: Software improvements on data gathering MES

Confidential

Confidential

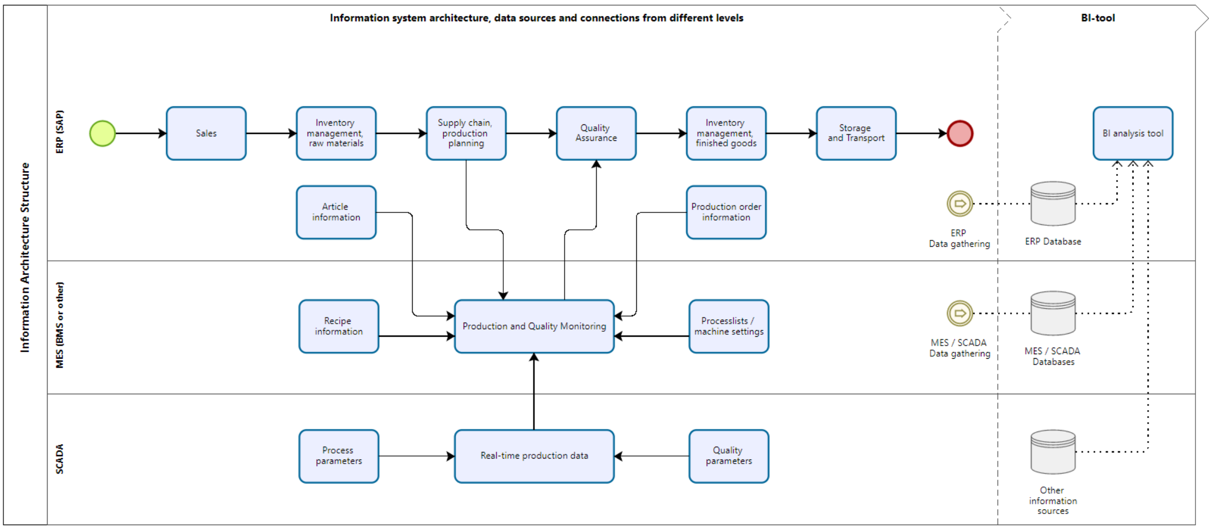
Appendix U: Future Information systems architecture

Information System architecture, with processes in SAP, functions in MES, and data gathering on the shopfloor

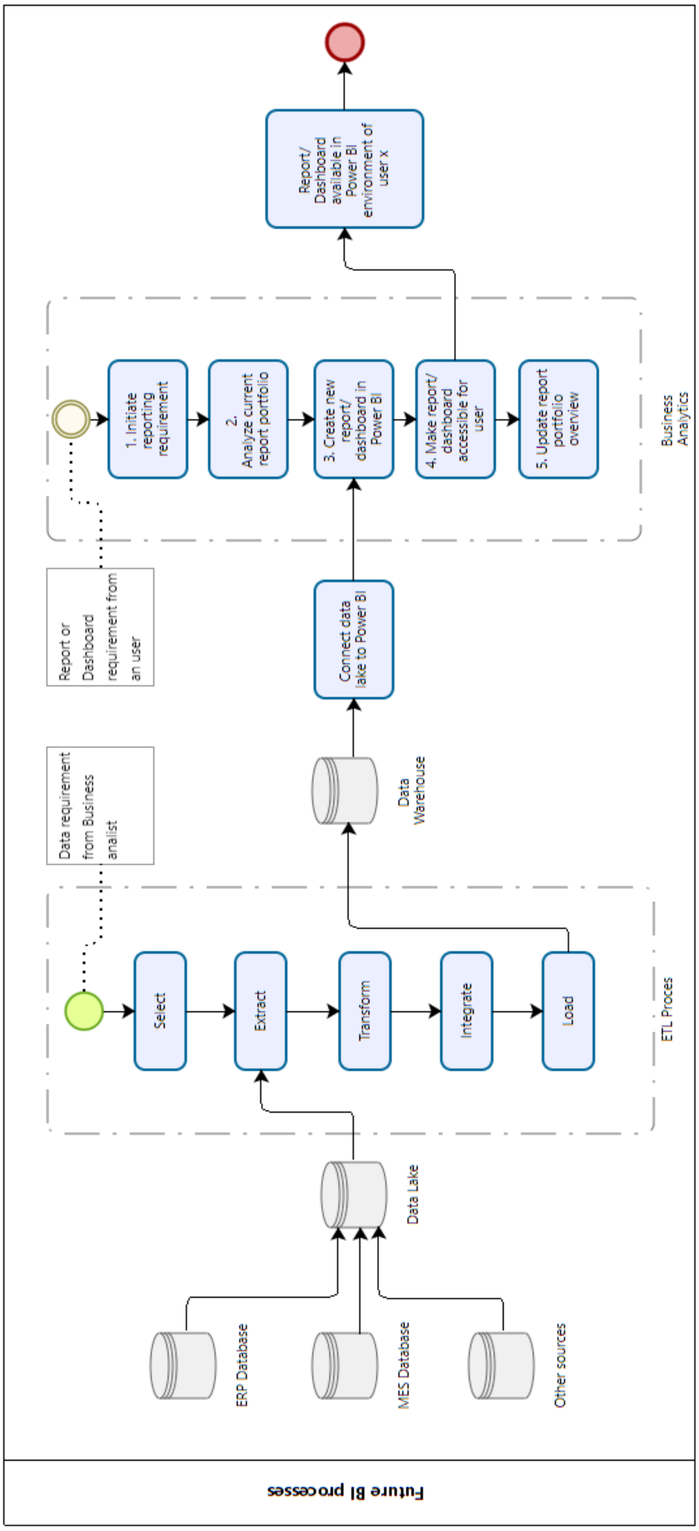


Simplified figure with information flows through the systems:

The left pool visualizes the information system architecture and the information sharing between different modules in a system. The right pool presents the architecture for the BI-tool. This BI-tool is the front dashboard that presents data from multiple sources, it selects the required data from the 'data lake'.



Future data process of information creation in the BI-tool Power BI:



Appendix V: Process parameters and Quality measurements

Process parameters currently gathered in extrusion line 1, extracted with USB from the machine:

Confidential

Confidential

Confidential

Confidential

Appendix W: Magic quadrant of Gartner

Magic Quadrant

Figure 1. Magic Quadrant for Analytics and Business Intelligence Platforms

Source: Gartner (February 2020)

Appendix X: Constructing data visualizations in BI-tool

Confidential

Confidential

Confidential

Confidential

Confidential

Confidential

Confidential

Confidential

Confidential

Confidential

Confidential

Confidential

Confidential

Appendix AA: Data processing method

Data processing process (Chen, Mao, & Yunhao, 2014)

1. Data Gathering
2. Data transportation, Upon the completion of raw data collection, data will be transferred to a data storage infrastructure for processing and analysis
3. Data pre-processing
    a. Integration: Data integration is the cornerstone of modern commercial informatics, which involves the combination of data from different sources and provides users with a uniform view of data. Historically, two methods have been widely recognized: data warehouse and data federation. Data warehousing includes a process named ETL (Extract, Transform and Load). Extraction involves connecting source systems, selecting, collecting, analyzing, and processing necessary data. Transformation is the execution of a series of rules to transform the extracted data into standard formats
    b. Cleaning: data cleaning is a process to identify inaccurate, incomplete, or unreasonable data, and then modify or delete such data to improve data quality. Generally, data cleaning includes five complementary procedures [68]:
        i. defining and determining error types,
        ii. searching and identifying errors,
        iii. correcting errors,
        iv. documenting error examples and error types
        v. modifying data entry procedures to reduce future errors.
    c. Redundancy elimination: data redundancy refers to data repetitions or surplus, which usually occurs in many datasets. Data redundancy can increase the unnecessary data transmission expense and cause defects on storage systems, e.g., waste of storage space, leading to data inconsistency, reduction of data reliability, and data damage. Therefore, various redundancy reduction methods have been proposed, such as redundancy detection, data filtering, and data compression. redundancy reduction may also bring about certain negative effects. For example, data compression and decompression cause additional computational burdens.
4. Data storage, To use a distributed system to store massive data, the following factors should be taken into consideration:
    a. Consistency: a distributed storage system requires multiple servers to cooperatively store data
    b. Availability: a distributed storage system operates in multiple sets of servers
    c. Partition Tolerance: multiple servers in a distributed storage system are connected by a network.
5. Data analytics
    a. Traditional (also used in the literature review)
        i. Cluster Analysis: is a statistical method for grouping objects, and specifically, classifying objects according to some features
        ii. Correlation Analysis: is an analytical method for determining the law of relations, such as correlation, correlative dependence, and mutual

restriction, among observed phenomena and accordingly conducting forecast and control

iii. Regression Analysis: is a mathematical tool for revealing correlations between one variable and several other variables.
iv. A/B Testing: also called bucket testing. It is a technology for determining how to improve target variables by comparing the tested group
v. Statistical Analysis: Statistical analysis is based on statistical theory, a branch of applied mathematics.

Data Mining Algorithms: Data mining is a process for extracting hidden, unknown, but potentially useful information and knowledge from massive, incomplete, noisy, fuzzy, and random data. (Chen, Mao, & Yunhao, 2014)

Appendix AB: Data transformation code Rstudio

```
#### Creation of SAP.EX
  ### By: Guus Schutte
  ### Date: 02-12-2020 adjusted at 23-12-2020

  ## Add packages
  library(DBI)
  library(RPostgreSQL)
  library(readr)
  library(dplyr)
  library(lubridate)
  library(readxl)
  library(tibble)
  library(readxl)
  library(rJava)
  library(writexl)
  library(xlsx)

  ## Insert Data SAP-EX-GEN
    # Insert Data
    #(check file name and location is stated properly)
    SAP.EX.GEN <- read_excel("C:\\Users\\guuss\\OneDrive\\Bureaublad\\DataDyka\\Data\\
                             SAP-EX-GEN.xlsx", sheet = 1,col_names = TRUE, col_types = NULL,
                             na = "", skip = 0)

  ##DATA cleaning and formatting
    #delete undesired columns
    SAP.EX.GEN$Timestamp = NULL

    #change column order
    SAP.EX.GEN <- SAP.EX.GEN[,c(1,3,13,12,4,5,6,7,8,9,10,11,14,2)]

    #Rename 'Dag' column to 'Datum' and `Order nummer` to 'order'
    SAP.EX.GEN <-SAP.EX.GEN %>% rename(Datum = Dag)
    SAP.EX.GEN <-SAP.EX.GEN %>% rename(Order = `Order nummer`)

    #change time format
    SAP.EX.GEN$Tijd <- format(as.POSIXct(SAP.EX.GEN$Tijd, format='%m/%d/%Y %H:%M:%S'),
                              format='%H:%M:%S')
    SAP.EX.GEN$Datum <- format(as.POSIXct(SAP.EX.GEN$Datum, format='%Y/%m/%d %H:%M:%S'),
                               format='%Y-%m-%d')

    #arrange data by time
    SAP.EX.GEN <- SAP.EX.GEN %>% arrange(Tijd)
    SAP.EX.GEN <- SAP.EX.GEN %>% arrange(Datum)

    #change data format machine
    SAP.EX.GEN$Afd <- paste('EX',substr(SAP.EX.GEN$Afd,2,3),sep='')

  ## Insert Data SAP-EX-HU
    #(check file name and location is stated properly)
    SAP.EX.HU <- read_excel("C:\\Users\\guuss\\OneDrive\\Bureaublad\\DataDyka\\Data\\
                            SAP-EX-HU.xlsx", sheet = 1,col_names = TRUE, col_types = NULL,
                            na = "", skip = 0)

  ##DATA cleaning and formatting
    #change column order
    SAP.EX.HU <- SAP.EX.HU[,c(1,2,3,9,7,8,4,6,5,10)]

    #Rename 'Ingevoerd op' column to 'Datum' and 'Externe key' to 'HU-nummer'
    SAP.EX.HU <-SAP.EX.HU %>% rename('Datum' ='Ingevoerd op')
    SAP.EX.HU <-SAP.EX.HU %>% rename('HU-nummer'='Externe key')
    SAP.EX.HU <-SAP.EX.HU %>% rename('Hoev' = 'Teruggem.goede hoevh')

    #change time format
    SAP.EX.HU$Tijd <- format(as.POSIXct(SAP.EX.HU$Tijd, format='%m/%d/%Y %H:%M:%S'),
                             format='%H:%M:%S')
    SAP.EX.HU$Datum <- format(as.POSIXct(SAP.EX.HU$Datum, format='%Y/%m/%d %H:%M:%S'),
                              format='%Y-%m-%d')
```

```
    #arrange data by time
    SAP.EX.HU <- SAP.EX.HU %>% arrange(Tijd)
    SAP.EX.HU <- SAP.EX.HU %>% arrange(Datum)

    #Adjust HU-number to
    SAP.EX.HU$'HU-nummer' = substr(SAP.EX.HU$'HU-nummer',1,18)

    #change data format
    SAP.EX.HU$'Storno terugmelding' <- as.numeric(as.character(SAP.EX.HU$'Storno terugmelding'))

    #Multiply by -1 if package is storno
    SAP.EX.HU$test <- NA
    SAP.EX.HU$test[SAP.EX.HU$'Storno terugmelding' != 0] <- (-1)
    SAP.EX.HU$test[SAP.EX.HU$'Storno terugmelding' == 0] <- (1)
    SAP.EX.HU$Hoev <- SAP.EX.HU$Hoev * SAP.EX.HU$test
    SAP.EX.HU$test <- NULL

    #Combine both tables to one
    mergeCols <- c("Datum","Tijd", "Order","Hoev" )
    SAP.EX <- merge(SAP.EX.GEN, SAP.EX.HU, by = mergeCols, all = TRUE)
    rm(mergeCols)

    #Combine date and time to fit the format for the validation of orderstart
    SAP.EX$Date <- NA
    SAP.EX$Date <- as.POSIXct(paste(SAP.EX.HU$Datum, SAP.EX.HU$Tijd), format='%Y-%m-%d %H:%M:%S')

    # rename Order to job
    SAP.EX <- SAP.EX  %>% rename('Job' = Order)

  ## delete GEN and HU
      rm(SAP.EX.GEN)
      rm(SAP.EX.HU)

 ##endtable


#### Creation of Production.Eventlog PROD-EX-Eventlog
  ### By: Guus Schutte
  ### Date: 08-12-2020

  ## Add packacges
  library(DBI)
  library(RPostgreSQL)
  library(readr)
  library(dplyr)
  library(lubridate)
  library(readxl)
  library(rJava)
  library(tibble)
  library(readxl)
  library(writexl)
  library(xlsx)
  library(tidyr)
  library(data.table)

    ## Insert Data Production-EX01-Eventlog
      #(check file name and location is stated properly)
      PROD.EX.Eventlog <- read_excel("C:\\Users\\guuss\\OneDrive\\Bureaublad\\DataDyka\\Data\\
                                 PROD-EX-Eventlog.xlsx", sheet = 1,col_names = TRUE,
                                 col_types = NULL, na = "", skip = 0)

      #Resize Machine number format
      PROD.EX.Eventlog$Machine <- paste('EX',substr(PROD.EX.Eventlog$Machine,2,3),sep='')

```

```
    # rename table headers
    PROD.EX.Eventlog <-PROD.EX.Eventlog %>% rename('Job Start Datum' = Starttijd)
    PROD.EX.Eventlog <-PROD.EX.Eventlog %>% rename('Job Sluit Datum' = Eindtijd)

  ## Insert Data Production-EX01-ProdCount
    #(check file name and location is stated properly)
    PROD.EX.ProdCount <- read_excel("C:\\Users\\guuss\\OneDrive\\Bureaublad\\DataDyka\\Data\\
                                    PROD-EX-ProdCount.xlsx", sheet = 1,col_names = TRUE,
                                    col_types = NULL, na = "", skip = 0)

    # rename table headers
    PROD.EX.ProdCount <-PROD.EX.ProdCount %>% rename('Job Sluit Datum' = 'Job End Date')
    PROD.EX.ProdCount <-PROD.EX.ProdCount %>% rename('Machine' = 'Machine nummer')

    #Resize Machine number format
    PROD.EX.ProdCount$length <- NA
    PROD.EX.ProdCount$length = nchar(as.character(PROD.EX.ProdCount$`Machine`))
    PROD.EX.ProdCount$`Machine` <-  if_else(PROD.EX.ProdCount$length > 1, paste('EX',
                                    PROD.EX.ProdCount$`Machine`,sep=''), paste('EX0',
                                    PROD.EX.ProdCount$`Machine`,sep=''))
    PROD.EX.ProdCount <- PROD.EX.ProdCount[-c(20)]

  ## Insert jobnumber in eventlog
    #add table that updates the values, orders the lines by date to fill order in eventlog
    mergeCols <- c("Job Start Datum","Job Sluit Datum", "Machine")
    PROD.EX.Eventlog2 <- merge(PROD.EX.Eventlog, PROD.EX.ProdCount, by = mergeCols, all = TRUE)
    rm(mergeCols)

    PROD.EX.Eventlog2<-  PROD.EX.Eventlog2 %>% arrange("Job Start Datum")
    PROD.EX.Eventlog2<-  PROD.EX.Eventlog2 %>% arrange(Machine)

    #copy ordernumber into eventlog
    PROD.EX.Eventlog2 <-   PROD.EX.Eventlog2 %>% fill(Job)
    # delete prodcount values
    PROD.EX.Eventlog2 <-  PROD.EX.Eventlog2[is.na(PROD.EX.Eventlog2$`Ploeg startdatum`),]
    # delete redundant columns and overwrite eventlog
    PROD.EX.Eventlog2 <- PROD.EX.Eventlog2[-c(7,8,9,10,12,13,14,15,16,17,18,19,20,21,22,23,
                                              24,25,26,27)]
    PROD.EX.Eventlog <- PROD.EX.Eventlog2
    rm(PROD.EX.Eventlog2)

  ## data preparation for merge with production data
    # add column that fills na with running status
    PROD.EX.Eventlog$Status <- PROD.EX.Eventlog$Omschrijving
    PROD.EX.Eventlog$Status <- PROD.EX.Eventlog$Status %>% replace_na("Running")
    PROD.EX.Eventlog <- PROD.EX.Eventlog[ -c(6) ]


    # Add eventlog ID for equencing the event that occur in the same minute
    PROD.EX.Eventlog <- PROD.EX.Eventlog %>% mutate(Eventid = row_number())
    PROD.EX.Eventlog <- PROD.EX.Eventlog[,c(8,1,2,3,4,5,6,7)]

    # change date format from date-seconds to date-minute
    PROD.EX.Eventlog$'Job Start Datum'  <- format(as.POSIXct(PROD.EX.Eventlog$'Job Start Datum',
                                           format='%d.%m.%Y %H:%M:%S'),format='%Y-%m-%d %H:%M')
    PROD.EX.Eventlog$'Job Sluit Datum'  <- format(as.POSIXct(PROD.EX.Eventlog$'Job Sluit Datum',
                                           format='%d.%m.%Y %H:%M:%S'),format='%Y-%m-%d %H:%M')

    # Rename column for merge with production data
    names(PROD.EX.Eventlog)[names(PROD.EX.Eventlog) == "Job Start Datum"] <- "Date"

    # resize Ordernumber
    PROD.EX.Eventlog$Job <- substr(PROD.EX.Eventlog$Job, 1,7)

    # Delete redunant columns of running time and stop time
    PROD.EX.Eventlog <- PROD.EX.Eventlog[ -c(3,5,6) ]
```

```
    # Group delete duplicates
    PROD.EX.Eventlog<- PROD.EX.Eventlog %>% arrange(Eventid)
    PROD.EX.Eventlog <- PROD.EX.Eventlog[!duplicated(data.table::rleidv(PROD.EX.Eventlog,c("Job",
                                                                                 "Status"))), ]

    ##archive PRODCOUNT
    Z.OLD.PROD.EX.ProdCount <- PROD.EX.ProdCount
    rm(PROD.EX.ProdCount)
    rm(Z.OLD.PROD.EX.ProdCount)

    # resize eventlog ID for equencing the event that occur in the same minute
    PROD.EX.Eventlog<- PROD.EX.Eventlog %>% arrange(Eventid)
    PROD.EX.Eventlog <- PROD.EX.Eventlog %>% mutate(Eventid = row_number())


    #Eventlog with job number

#### Validation of Production.Eventlog with sapdata SAP-EX01-VAL
  ### By: Guus Schutte
  ### Date: 15-12-2020
  ### Valiation iss used to provide information, not to clean the data

    ## Add packacges
    library(DBI)
    library(RPostgreSQL)
    library(readr)
    library(dplyr)
    library(lubridate)
    library(readxl)
    library(rJava)
    library(tibble)
    library(readxl)
    library(writexl)
    library(xlsx)
    library(tidyr)
    library(data.table)

    ##run SAP.EX.R and PROD.EX01-Eventlog

    ##Table SAP.EX Validates the start date of a job

    ## Create validation table of SAP.EX01 -> SAP.EX.VAL
    SAP.EX.VAL <- SAP.EX

    ## merge date and time
    SAP.EX.VAL$Date <- as.POSIXct(paste(SAP.EX.VAL$Datum, SAP.EX.VAL$Tijd), format=
                                                             "%Y-%m-%d %H:%M:%S")

    ## sort on time then date then on order
    SAP.EX.VAL<- SAP.EX.VAL %>% arrange(Tijd)
    SAP.EX.VAL<- SAP.EX.VAL %>% arrange(Date)
    SAP.EX.VAL<- SAP.EX.VAL %>% arrange(Job)

    ## remove dublicates to create earliest start table (first package of an order in the system)
    SAP.EX.VAL <- SAP.EX.VAL[!duplicated(data.table::rleidv(SAP.EX.VAL,c("Job"))), ]

    ## sort on time then date
    SAP.EX.VAL<- SAP.EX.VAL %>% arrange(Tijd)
    SAP.EX.VAL<- SAP.EX.VAL %>% arrange(Date)

    ###insert article data
    ART.EX <- read_excel("C:\\Users\\guuss\\OneDrive\\Bureaublad\\DataDyka\\Data\\ART-EX.xlsx",
                         sheet = 1,col_names = TRUE, col_types = NULL, na = "", skip = 0)

    ## Change format machine
    ART.EX <- ART.EX %>% rename('Afd' = Resource)
```

```
        ART.EX$`Afd` <- paste('EX',substr(ART.EX$`Afd`, 2, 3),sep='')

        ##add time to make a pipe
        mergeCols <- c("Artikel","Afd")
        SAP.EX.VAL <- merge(x = SAP.EX.VAL, y = ART.EX, by = mergeCols, all.x = TRUE)
        rm(mergeCols)

        ##select important collumns
        SAP.EX.VAL  <- SAP.EX.VAL [,c(1,2,3,4,5,6,7,21,46,49)]

        ## calculate time to complete package
        SAP.EX.VAL$Pakket <-    SAP.EX.VAL$Machine * SAP.EX.VAL$Hoev

        ## calculate earlist start
        SAP.EX.VAL$maximumstartjob <-  as.POSIXct(SAP.EX.VAL$Date, format = '%Y.%m.%d %H:%M:%S') -
                                                                       (SAP.EX.VAL$Pakket *60)

        ### ADD earlist start to eventlog
        SAP.EX.VAL  <- SAP.EX.VAL [,c(1,2,5,12)]
        SAP.EX.VAL$Status <- 'Running'


        ## left join eventlog with sap.ex01.val
        mergeCols <- c("Job", "Status")
        PROD.EX.Eventlog <- merge(x = PROD.EX.Eventlog, y = SAP.EX.VAL, by = mergeCols, all.x = TRUE)
        rm(mergeCols)
        PROD.EX.Eventlog<- PROD.EX.Eventlog %>% arrange(Date)
        PROD.EX.Eventlog<- PROD.EX.Eventlog %>% arrange(Afd)

        ##delete earlystartdate if a job has already started
        PROD.EX.Eventlog<- PROD.EX.Eventlog %>% arrange(maximumstartjob)
        PROD.EX.Eventlog$maximumstartjob[duplicated(PROD.EX.Eventlog$maximumstartjob)] <- NA
        PROD.EX.Eventlog  <- PROD.EX.Eventlog [,c(3,2,1,7,6,4,8)]

        ## Arrange eventlog
        PROD.EX.Eventlog<- PROD.EX.Eventlog %>% arrange(Eventid)

        ## Fill article and Afd
        PROD.EX.Eventlog <- PROD.EX.Eventlog %>% fill(Afd, Artikel)

        ##change date format
        PROD.EX.Eventlog$maximumstartjob  <- format(as.POSIXct(PROD.EX.Eventlog$maximumstartjob,
                                              format='%d.%m.%Y %H:%M:%S'),format='%Y-%m-%d %H:%M')

        ## Delete SAP.EX.01.VAL
        rm(SAP.EX.VAL)

        ##Compare jobchange to late executed in minutes
        PROD.EX.Eventlog$JobChangeTooLate  <- difftime(PROD.EX.Eventlog$Date,
                                              PROD.EX.Eventlog$maximumstartjob, units = "mins")
        PROD.EX.Eventlog$JobChangeTooLate[PROD.EX.Eventlog$JobChangeTooLate < 0] <- "0"
        PROD.EX.Eventlog<- PROD.EX.Eventlog %>% arrange(Eventid)

        PROD.EX.Eventlog$JobChangeTooLate  <- if_else(difftime(PROD.EX.Eventlog$Date,
                                              PROD.EX.Eventlog$maximumstartjob, units = "mins") > 0,
                                              difftime(PROD.EX.Eventlog$Date,
                                               PROD.EX.Eventlog$maximumstartjob, units = "mins"),0)


        #endtable

#### Creation and cleaning  of Production data PROD-EX01-PROD
### By: Guus Schutte
### Date: 08-12-2020
```

```r

        ## Add packages
        library(DBI)
        library(RPostgreSQL)
        library(readr)
        library(dplyr)
        library(lubridate)
        library(readxl)
        library(rJava)
        library(tibble)
        library(readxl)
        library(writexl)
        library(xlsx)
        library(tidyr)

        ## Insert Data Production-EX01-Eventlog
        # Insert Data
        #(check file name and location is stated properly)
        PROD.EX01.PROD <- read_excel("C:\\Users\\guuss\\OneDrive\\Bureaublad\\DataDyka\\Data\\
            PROD-EX01-PROD.xlsx", sheet = 1,col_names = TRUE, col_types = NULL, na = "", skip = 0)

        ## Data cleaning
        # Reset time difference machine and real time
        PROD.EX01.PROD$Date <-  as.POSIXct(PROD.EX01.PROD$Date, format = '%d.%m.%Y %H:%M:%S') -
                                                  minutes(69) #69 minutes difference

        # format cells to date-minute
        PROD.EX01.PROD$Date <- format(as.POSIXct(PROD.EX01.PROD$Date, format='%d.%m.%Y %H:%M:%S'),
                                                                 format='%Y-%m-%d %H:%M')

        # sorting
        PROD.EX01.PROD <- PROD.EX01.PROD %>% arrange(Date)

        #Remove duplicates
        PROD.EX01.PROD2 <-PROD.EX01.PROD[!duplicated(PROD.EX01.PROD$Date), ]

        PROD.EX01.PROD <- PROD.EX01.PROD2
        PROD.EX01.PROD <- PROD.EX01.PROD %>% arrange(Date)
        rm(PROD.EX01.PROD2)

        #Add machine number
        PROD.EX01.PROD$Afd <- NA
        PROD.EX01.PROD$Afd <- PROD.EX01.PROD$Afd  %>% replace_na("EX01")

        # Change column order
        PROD.EX01.PROD  <- PROD.EX01.PROD [,c(1,62,2,22,23,24,25,47,48,50,57,58,59,61,26,45,46,3,4
                             ,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,27,28,29,30,31,
                             32,33,34,35,36,37,38,39,40,41,42,43,44,49,51,52,53,54,55,56,60)]

# end production table PROD-EX01
#### Creation and  Production data combination eventlog and production date
### By: Guus Schutte
### Date: 08-12-2020

        ## Add packacges
        library(DBI)
        library(RPostgreSQL)
        library(readr)
        library(dplyr)
        library(lubridate)
        library(readxl)
        library(rJava)
        library(tibble)
        library(readxl)
        library(writexl)
        library(xlsx)
```

```
      library(tidyr)

      ### first run PROD-EX01-Eventlog and PROD-EX01-PROD

      ### Combine both table
      ## Create list with minutes
      PROD.EX01 <- data.frame(Date = seq(as.POSIXct("2020-08-01"),as.POSIXct("2020-12-01"),by=(60)))

      #Add machine number (change this value if you add data from another machine)
      PROD.EX01$Afd <- NA
      PROD.EX01$Afd <- PROD.EX01$Afd  %>% replace_na("EX01")

      ## format cells to date-minute
      PROD.EX01$Date <- format(as.POSIXct(PROD.EX01$Date, format='%d.%m.%Y %H:%M:%S'),
                                                      format='%Y-%m-%d %H:%M')

      ## left outer join PROD.EX01 <- PROD.EX01.Eventlog
      mergeCols <- c("Date", "Afd")
      PROD.EX01 <- merge(x = PROD.EX01, y = PROD.EX.Eventlog, by = mergeCols, all.x = TRUE)
      rm(mergeCols)

      ## fill values for each minute in  entire table
      PROD.EX01 <-   PROD.EX01 %>% fill(Afd, Job, Status, Artikel)

      ##Delete redundant columns
      PROD.EX01  <- PROD.EX01 [,c(1,2,4,5,6)]

      ##delete row to get required time interval
      PROD.EX01 <- PROD.EX01 %>% filter(Date > "2020-09-01")

      ## left outer join PROD.EX01 <- PROD.EX01.PROD
      mergeCols <- c("Date", "Afd")
      PROD.EX01 <- merge(x = PROD.EX01, y = PROD.EX01.PROD, by = mergeCols, all.x = TRUE)
      rm(mergeCols)

      ##Remove unnecessary columns
      rm(PROD.EX01.PROD)
      #rm(ART.EX)

### end production table BMS-EX01

#### Creation data from BMS to provide insight in the declaration behavior
### By: Guus Schutte
### Date: 22-12-2020

    ## Add packacges
    library(DBI)
    library(RPostgreSQL)
    library(readr)
    library(dplyr)
    library(lubridate)
    library(readxl)
    library(rJava)
    library(tibble)
    library(readxl)
    library(writexl)
    library(xlsx)
    library(tidyr)

    ## Insert Data BMS-EX data drom GUUS PROD COUNT in BMS
      #(check file name and location is stated properly)
      BMS.EX <- read_excel("C:\\Users\\guuss\\OneDrive\\Bureaublad\\DataDyka\\Data\\
                           BMS-EX.xlsx", sheet = 1,col_names = TRUE, col_types = NULL,
                           na = "", skip = 0)
```

```
    # rename table headers
    BMS.EX <- BMS.EX %>% rename('Job Sluit Datum' = 'Job End Date')
    BMS.EX <- BMS.EX %>% rename('Machine' = 'Machine nummer')

    #Resize Machine number format
    BMS.EX$length <- NA
    BMS.EX$length = nchar(as.character(BMS.EX$`Machine`))
    BMS.EX$`Machine` <-  if_else(BMS.EX$length > 1, paste('EX',BMS.EX$`Machine`,sep=''),
                                                   paste('EX0',BMS.EX$`Machine`,sep=''))
    BMS.EX <- BMS.EX[-c(20)]
    BMS.EX <- BMS.EX %>% rename('Afd' = 'Machine')
    BMS.EX <- BMS.EX %>% rename('Artikel' = 'Product')
    BMS.EX <- BMS.EX %>% rename('Artikelbeschrijving' = 'Productbeschrijving')

    # resize Ordernumber
    BMS.EX$Job <- substr(BMS.EX$Job, 1,7)

    ##rename PROD.EX.Eventlog to BMS-EX-Eventlog
    BMS.EX.Eventlog <- PROD.EX.Eventlog
    rm(PROD.EX.Eventlog)


#### Creation and cleaning  of Production data PROD-EX23-PROD
### By: Guus Schutte
### Date: 13-01-2021

  ## Add packacges
  library(DBI)
  library(RPostgreSQL)
  library(readr)
  library(dplyr)
  library(lubridate)
  library(readxl)
  library(rJava)
  library(tibble)
  library(readxl)
  library(writexl)
  library(xlsx)
  library(tidyr)

  ## Insert Data Production-EX01-Eventlog
    # Insert Data
    #(check file name and location is stated properly)
    PROD.EX23.PROD <- read_excel("C:\\Users\\guuss\\OneDrive\\Bureaublad\\DataDyka\\Data\\
                              PROD-EX23-PROD.xlsx", sheet = 1,col_names = TRUE, col_types =
                                NULL, na = "", skip = 0)

  ## Data cleaning
    # format cells to date-minute
    PROD.EX23.PROD$Date <- format(as.POSIXct(PROD.EX23.PROD$Date, format='%d.%m.%Y %H:%M:%S'),
                               format='%Y-%m-%d %H:%M') #time difference is adjusted in excel

    # sorting
    PROD.EX23.PROD <- PROD.EX23.PROD %>% arrange(Date)

    #Remove duplicates
    PROD.EX23.PROD2 <-PROD.EX23.PROD[!duplicated(PROD.EX23.PROD$Date), ]

    PROD.EX23.PROD <- PROD.EX23.PROD2
    PROD.EX23.PROD <- PROD.EX23.PROD %>% arrange(Date)
    rm(PROD.EX23.PROD2)

    #Add machine number
```

```r
      PROD.EX23.PROD$Afd <- NA
      PROD.EX23.PROD$Afd <- PROD.EX23.PROD$Afd  %>% replace_na("EX23")

      # Change column order
      PROD.EX23.PROD  <- PROD.EX23.PROD [,c(1,108,2,55,56,58,88,90,96,103,104,54,57,59,84,85,86,
                                             87,89,94,95,105,106,107,3,4,5,6,7,8,9,10,11,12,13,14,
                                             15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,
                                             33,34,35,36,37,38,39,40,41,42,43,44,45,46,47,48,49,50,
                                             51,52,53,54,60,61,62,63,64,65,66,67,68,69,70,71,72,73,
                                             74,75,76,78,79,80,81,82,83,91,92,93,94,95,97,98,99,100,
                                             101,102)]

#### Creation and  Production data combination eventlog and production date
### By: Guus Schutte
### Date: 14-01-2021

      ## Add packacges
      library(DBI)
      library(RPostgreSQL)
      library(readr)
      library(dplyr)
      library(lubridate)
      library(readxl)
      library(rJava)
      library(tibble)
      library(readxl)
      library(writexl)
      library(xlsx)
      library(tidyr)

      ### first run PROD-EX-Eventlog and PROD-EX01-PROD

      ### Combine both tables
      ## Create list with minutes
      PROD.EX23 <- data.frame(Date = seq(as.POSIXct("2020-08-01"),as.POSIXct("2020-12-01"),by=(60)))

      #Add machine number (change this value if you add data from another machine)
      PROD.EX23$Afd <- NA
      PROD.EX23$Afd <- PROD.EX23$Afd  %>% replace_na("EX23")

      ## format cells to date-minute
      PROD.EX23$Date <- format(as.POSIXct(PROD.EX23$Date, format='%d.%m.%Y %H:%M:%S'),
                                                         format='%Y-%m-%d %H:%M')

      ## left outer join PROD.EX23 <- PROD.EX23.Eventlog
      mergeCols <- c("Date", "Afd")
      PROD.EX23 <- merge(x = PROD.EX23, y = BMS.EX.Eventlog, by = mergeCols, all.x = TRUE)
                                                    # name of PROD.EX.Eventlog Has changed
      rm(mergeCols)

      ## fill values for each minute in  entire table
      PROD.EX23 <-    PROD.EX23 %>% fill(Afd, Job, Status, Artikel)

      ##Delete redundant columns
      PROD.EX23   <- PROD.EX23 [,c(1,2,4,5,6)]

      ##delete row to get required time interval
      PROD.EX23 <- PROD.EX23 %>% filter(Date > "2020-09-01")

      ## left outer join PROD.EX23 <- PROD.EX23.PROD
      mergeCols <- c("Date", "Afd")
      PROD.EX23 <- merge(x = PROD.EX23, y = PROD.EX23.PROD, by = mergeCols, all.x = TRUE)
      rm(mergeCols)
```

```
        ##Remove unnecessary columns
        rm(PROD.EX23.PROD)
        rm(PROD.EX23.PRODbackup)

        #end table PROD.EX23

  #### Add and rewrite BMS-QC-EX data

        BMS.QC.EX <- read_excel("C:\\Users\\guuss\\OneDrive\\Bureaublad\\DataDyka\\Data\\
                                BMS-QC-EX.xlsx", sheet = 1,col_names = TRUE, col_types = NULL,
                                na = "", skip = 0)


 #### Save data
        write_xlsx(PROD.EX01,"./Data/PROD-EX01-end.xlsx")
        write_xlsx(PROD.EX23,"./Data/PROD-EX23-end.xlsx")

        write_xlsx(BMS.EX,"./Data/BMS-EX-end.xlsx")
        write_xlsx(BMS.QC.EX,"./Data/BMS-QC-EX-end.xlsx")
        write_xlsx(BMS.EX.Eventlog,"./Data/BMS-EX-Eventlog-end.xlsx")
        write_xlsx(ART.EX,"./Data/ART-EX-end.xlsx")
        write_xlsx(SAP.EX,"./Data/SAP-EX-end.xlsx")
```

Appendix AC: Evaluation of BI-model with future users

Function	Required dashboard/information from data
Quality Engineer and Manager	A dashboard on Quality data
General engineering	Data available for analysis
Quality controller extrusion	(not a future user) (maybe a dashboard for production history)
Business controller Financial	Data available for analysis and the creation of dashboards
Business controller Operations	Data available for analysis and the creation of dashboards
Technical assistant Injection Molding	Data available for analysis
Technical department	Data available for analysis
Process engineer Extrusion	Data available for analysis and dashboards for process and quality KPI's
Technical assistant extrusion	Data available for analysis
Shift leader extrusion	dashboards OEE performance and for process and quality KPI's
Manager injection molding	dashboards OEE performance and for process and quality KPI's
Manager extrusion	dashboards OEE performance and for process and quality KPI's
Director sales	Dashboard reporting costs/margin performance
Director Finance	Dashboard reporting operational performance and costs/margin performance
Site director operations	Dashboard reporting costs/margin performance