Tampereen yliopisto

Heikki Kauppinen

# COMPUTERIZED SYSTEMS' ROLE DETERMINATION IN PHARMACEUTICAL MANUFACTURING PROCESS

Master of Science Thesis
Faculty of Engineering and Natural Sciences
Mikko Salmenperä
Matti Vilkko
February 2024

# ABSTRACT

Heikki Kauppinen: Computerized systems’ role determination in pharmaceutical manufacturing process
Master of Science Thesis
Tampere University
Master’s Degree Programme in Automation Engineering
February 2024

---

Pharmaceutical industry is a highly regulated industry and adoption of new technologies can be slow. However, even in pharmaceutical industry, adoption of new technologies is happening. This thesis examines a new department and its systems of a pharmaceutical production facility located in Finland. In the center of this thesis are a SCADA system and a data historian system. The goal of this thesis is to determine the roles for the systems in new department. The systems are intended to address needs and use cases that come from the target department. The needs are established by surveying employees working in different roles on the target department. The roles and the feasibility of the systems are determined by examining the user requirements of each system.

In this thesis, the special requirements and characteristics of the pharmaceutical industry are introduced. This thesis examines the SCADA and the data historian in general, before examining them from the target department’s perspective. The processes and operations of the target department are introduced, before moving into the surveying of the employees working in the target department. The answers of the survey are analyzed. Based on the answers, needs of the target department are established. In order to address these needs, the roles for each system are tried to find. This is done by analyzing the user requirements of each system.

The user requirements specifications of both systems are scrutinized, and critical requirements are summarized into six categories. The requirements in each category are analyzed in order to determine the roles for the systems. The analysis of the user requirements proved to be difficult and did not produce satisfactory results for determining the best use cases and roles for the systems. However, by examining the capabilities of each system, it was possible to partially determine the roles. Both systems can address different needs, and the parallel use of the systems could be found to be beneficial for improving the overall performance of the target department.

Keywords: pharmaceutical industry, user requirements, SCADA system, data historian

# TIIVISTELMÄ

Heikki Kauppinen: Tietojärjestelmien roolien määrittely lääketeollisessa valmistusprosessissa
Diplomityö
Tampereen yliopisto
Automaatiotekniikan DI-ohjelma
Helmikuu 2024

---

Lääketeollisuus on voimakkaasti säädelty teollisuuden ala, jossa uusien teknologioiden käyttöönotto voi olla hidasta. Lääketeollisuudessakin uusien teknologioiden käyttöönotto on kuitenkin jatkuvaa. Tämä työ tarkastelee erään Suomessa sijaitsevan lääketeollisuuden tuotantolaitoksen uutta osastoa ja sen järjestelmiä. Keskiössä ovat SCADA järjestelmä sekä data historian järjestelmä. Tavoitteena tässä työssä on selvittää järjestelmien rooleja uudella osastolla. Järjestelmät pyrkivät vastaamaan tarpeisiin ja käyttötapauksiin, jotka tulevat tarkasteltavalta osastolta. Tarpeita kartoitetaan tässä työssä haastattelemalla kohdeosastolla eri rooleissa työskenteleviä henkilöitä. Järjestelmien roolit ja soveltuvuus tarpeisiin pyritään selvittämään tarkastelemalla kunkin järjestelmän käyttäjävaatimuksia.

Tässä työssä tutustutaan lääketeollisuuden vaatimuksiin sekä erityispiirteisiin. Työssä käydään läpi SCADA järjestelmää sekä data historian järjestelmää yleisesti, ennen kuin niitä tarkastellaan kohdeosaston näkökulmasta. Kohdeosaston prosesseja ja toimintaa esitellään, kunnes siirrytään kohdeosaston henkilöstön haastatteluihin. Haastattelujen tulokset analysoidaan, joniden pohjalta kohdeosaston tarpeet muodostetaan. Viisi keskeistä tarvetta pystytään muodostamaan haastattelujen vastauksien pohjalta. Näitä tarpeita vastaamaan työssä pyritään etsimään roolit kullekin järjestelmälle. Rooleja pyritään selvittämään tarkastelemalla järjestelmien käyttäjävaatimuksia.

Järjestelmien käyttäjävaatimukset käydään läpi tarkasti, ja niistä muodostetaan kootut keskeisimmät vaatimukset kuudessa eri kategoriassa. Näiden kategoiroiden vaatimuksia analysoidaan roolien löytämiseksi. Käyttäjävaatimusten analysointi ei tuottanut haluttuja tuloksia järjestelmien roolien määrittämiseksi. Käyttäjävaatimukset osoittautuivat liian teknisiksi ja samankaltaisiksi järjestelmien välillä. Järjestelmien kyvykkyyksiä ja vaatimuksia tarkastelemalla rooleja pystyttiin kuitenkin osittain määrittämään. Molemmat järjestelmät pystyvät vastaamaan eri tarpeisiin, ja järjestelmien yhteinen käyttö voidaan todeta olevan kannattavaa kohdeosaston suorituskyvyn parantamiseksi.

Avainsanat: lääketeollisuus, käyttäjävaatimukset, SCADA järjestelmä, data historian järjestelmä

# PREFACE

This thesis was commissioned by a pharmaceutical company that has a production facility in Finland. I am grateful for the opportunity to make this thesis for the company, and I hope this thesis brings value to the company – one way or another. However, writing a master's thesis while working full time turned out to be harder than I could ever imagine. I would not have been able to overcome this challenge if it wasn't for a few people.

First, I want to thank my instructors Mikko Salmenperä and Matti Vilkko for not giving up hope on me despite occasional desperate periods. A big thanks also goes to my manager. The advice and support I received from her were significant during this process. Then, I want to thank all my friends and family and my girlfriend for their emotional support and encouragement. Finally, the biggest thanks belong to my grandmother. Her belief and support were unwavering throughout the whole process. Even at times when I did not believe in myself, my grandmother never lost hope.

Turku 10.2.2024

Heikki Kauppinen

# TABLE OF CONTENTS

# LIST OF FIGURES

# ABBREVATIONS AND SYMBOLS

| | |
|---|---|
| DCS | Distributed Control System |
| EMEA | European Medicines Agency |
| ERP | Enterprise Resource Planning |
| EU | European Union |
| FDA | Food and Drug Administration |
| FIMEA | Finnish Medicines Agency |
| GAMP | Good Automated Manufacturing Practices |
| GEP | Good Engineering Practices |
| GMP | Good Manufacturing Practices |
| GUI | Graphical User Interface |
| GxP | Good “x” Practice |
| HMI | Human Machine Interface |
| ICS | Industrial Control System |
| IEEE | Institute of Electrical and Electronic Engineers |
| ISA | International Society of Automation |
| LSL | Lower Specification Limit |
| MES | Manufacturing Execution System |
| MTU | Master Terminal Unit |
| ODBC | Open Database Connectivity |
| OE | Operational Efficiency |
| OPC UA | OPC Unified Architecture |
| PLC | Programmable Logic Controller |
| QMS | Quality Management System |
| RTU | Remote Terminal Unit |
| SCADA | Supervisory Control and Data Acquisition |
| SQL | Structured Query Language |
| URS | User Requirement Specification |
| USL | Upper Specification Limit |

# 1. INTRODUCTION

Technological development and digitalization in pharmaceutical industry lag behind other industries partly because of the heavy regulations it must comply to [1]. However, digitalization and data-based value creation in pharmaceutical manufacturing play a key role in increasing efficiency, quality and productivity. In center of digitalization are various systems that collect more and more data from all phases of manufacturing processes. In addition to collecting data, utilizing the data is as important, and utilization of data can be crucial in decision making.

Pharmaceutical industry is a unique operational environment offering its own features and challenges. Pharmaceutical industry is heavily regulated and monitored by authorities like the European Medicines Agency (EMEA), the Finnish Medicines Agency (FIMEA) and the USA's Food and Drug Administration (FDA) [2,3,4]. All pharmaceutical manufacturing must follow *Good Manufacturing Practices* (GMP) according to EMEA [5]. EU legislation obligates pharmaceutical manufacturers to follow GMP [6]. Manufacturers must have a quality management system (QMS) in place that is properly documented, and its use is monitored accordingly. Quality management system and good manufacturing practices are linked to each other. Their common goal is to ensure that medicines sustain consistent high quality, are appropriate for their intended use and fulfill the requirements set by marketing or clinical trial authorization. [7]

Automation and data play an ever-increasing role in manufacturing industries. Rapid technological development has brought topics like digitalization, artificial intelligence and industry 4.0 on everyone's lips. More traditional industries that often lag behind the most novel innovations and technologies have to adapt to the changes in order to stay competitive. Especially the importance of data has vastly increased, partly because of its availability. Thus, data-based value creation is a key driver for success in competitive business environments. This thesis has a strong emphasis on data, and on how modern computerized systems bring more value and competitive advantage to the target company. At the center of this thesis are two different computerized systems used in a manufacturing process – a *Supervisory Control and Data Acquisition* (SCADA) system and a *data historian* system.

This thesis is commissioned by a pharmaceutical and life science company that has a production site in Finland. The target company has a new manufacturing department where a medicinal core for various products is made. The department that is being examined in this thesis is a huge leap forward in technology compared to the old department that was used to carry out the same processes. All the production machines are new and so is the software around the machines. The new machines and their software create whole new capabilities for the department. Along with the technological leap comes the increase in data. Now there is abundance of data available. It is a situation the department has never been in before. The challenge for the department is to maximize the potential of the new systems and the data that is now available. The danger is that the data is not utilized. This thesis finds the roles for the SCADA and the data historian systems and how their data and capabilities could be utilized the best.

This thesis examines the roles of the *SCADA* system and the *data historian* system in the target company's new production department by evaluating the *user requirements* of each system. In this thesis we investigate which system fulfills a certain use case or a need, but also the similarities between both systems are covered in this thesis. Both systems have similar capabilities, and they can both suit the same need. Topics discussed in this thesis are in the context of pharmaceutical industry and may not apply to other industries. SCADA and data historian are not always separated as two different systems. Data historian can be seen as a feature of a SCADA. However, the target company that is being examined in this thesis separates those two systems. In the target company the SCADA system and the data historian are two independent systems from different suppliers. They still have overlapping capabilities, but they are not dependent on each other.

The research problem in this thesis is a real problem and something that the target department themself has brought up. All the machines and software are new and the amount of data they create is unprecedented. During the commissioning of the department and its machines, the focus has not been on figuring out the roles of each system, and the usage of the data they generate. The objective of this thesis is to determine the roles for the SCADA and the data historian systems. Therefore, this thesis proposes a guideline for the department to utilize different systems for different use cases.

The goal of this thesis is to determine whether the SCADA system or the data historian system would be more feasible for certain needs set by the target department. User requirement specifications of both systems are thoroughly scrutinized, and key requirements are put together from each of the systems' user requirements. Those requirements are then evaluated against data based on a survey conducted in the target company. The research was done by surveying and discussing with different representatives from various functions. The survey is not done as research method but to gather background data instead. The answers of the survey will reveal the needs, concerning the systems, of the department. The feasibility of a system is measured by evaluating the requirements against the background data by analyzing the system user requirements. The research question this thesis answers is: what are the suitable roles for the SCADA and the data historian system in parallel use in the core manufacturing process?

This thesis is structured so that the first part of the thesis gives the reader enough theoretic background to understand the concepts that this thesis deals with. A depiction of the target company's manufacturing process and systems in scope follows the theoretical part. After the manufacturing process depiction and the background survey, this thesis delves into user requirements and their evaluation with the systems in scope before concluding this thesis.

Chapter two of this thesis introduces the pharmaceutical industry environment and its distinctive features. Good manufacturing practices and quality system are also covered in chapter two. Chapter two gives a general depiction of computerized systems and their validation. Chapter three describes the individual roles of different systems and their integration in relation to the International Society of Automation (ISA) ISA-95 automation standard [8]. A brief general introduction of a SCADA and a data historian is given. The chapter also touches the subject of industry 4.0 and what it means in the context of this thesis.

In chapter four of this thesis the target company's manufacturing process which is in the scope of this thesis is presented. A rough explanation of the manufacturing process and its different phases is given. The centralized SCADA system that is being used in the manufacturing process is described thoroughly. The data historian that is going to be used along with the SCADA is also examined in this chapter. Other systems in the target department are also briefly examined. In chapter five the survey questions and their corresponding answers are introduced. The answers are examined and analyzed.

Chapter six is the research part of the thesis where the user requirements of the two systems are analyzed in order to define roles for each system. First the structure and overview of each systems' user requirement specification is given. The overview is followed by analyzing the specifications to find the critical requirements. Critical requirements are summarized into key findings from the user requirement specifications. The target of the evaluation is to find the system that best suits the specific needs. The roles for each system are tried to determine as well. The chosen approach is discussed as well. In chapter seven the results of the evaluation are discussed, and the thesis is summarized. The challenges are also addressed in this chapter. Further development for this topic is considered in chapter seven as well.

# 2. PHARMACEUTICAL INDUSTRY

Pharmaceutical industry is heavily regulated by many different authorities. Pharmaceutical regulations can be defined as a combination legislation and technical and administrative measures taken by governments [9]. Legislation means laws that are binding whereas regulations are rules and guidelines based on those laws. Patient safety and the efficacy of a medicine are the most important factors in pharmaceuticals. Bad quality or misuse of medicine can have catastrophic consequences and even lead to loss of life. Therefore, pharmaceutical regulations aim primarily to ensure that medicine are safe, of good quality, effective and have accurate product information [10].

This chapter of the thesis covers the definition of the pharmaceutical industry and explains the regulations pharmaceutical industry has to follow. One of the most important concepts of pharmaceutical manufacturing, *Good Manufacturing Practices* (GMP), is inspected in this chapter as well. This chapter also introduces the most important organizations and bodies that conducts the regulation. *Computerized systems* and their validation are also included in this chapter. In the context of this thesis pharmaceutical industry consists of prescription drugs, self-medication drugs and medical devices.

System user requirements are a key to this thesis. The research part of this thesis is to compare user requirements of two different computerized systems and analyze their similarities and differences. Therefore, this chapter also examines user requirements from the pharmaceutical industry's perspective. How user requirements affect a project in its different phases is covered in this chapter as well.

## 2.1 Quality system and good manufacturing practices

According to the European Union rules governing medicinal product [7] every pharmaceutical manufacturer must have a *Quality System* in place incorporating *Good Manufacturing Practices* and Quality Risk Management. Quality System can also be referred as Quality Management System (QMS). The goal of quality management is to ensure that the quality of a product matches their requirements and suits for its intended use throughout its life cycle. Along with GMP a GxP abbreviation is often used in pharmaceutical documentation. The ”x” is a variable that can be replaced with multiple different letters altering the meaning. For example, GEP stands for Good

Engineering Practices. Therefore, GxP encompasses all the required practices in pharmaceutical manufacturing.

Quality System is a wide collection of many requirements that guides a pharmaceutical manufacturer. Everything that can affect the quality of a product in any way is covered in the Quality System concept. GMP must be applied throughout the whole life cycle of a medicinal product. Ultimately, senior management carries the responsibility of the Quality System and should ensure that its being periodically reviewed. Quality System defining documentation including responsibilities should also be in place. Quality System also strives for continuous improvement in developing the system and the products. [7]

GMP is part of Quality Management System that sets the minimum requirements for pharmaceutical manufacturers in their production processes. Consistent high quality, fitness for intended use and compliance to requirements set by authorities are the cornerstones of GMP. GMP also requires that the pharmaceutical manufacturer has all the necessary facilities for production. These facilities include trained and qualified personnel, appropriate premises, suitable equipment, services and materials, suitable storages and transportation and procedures and instructions in accordance with the Quality System of the manufacturer [7]. Different authorities like EMEA, FIMEA and the FDA carry out periodic audits to make sure that GMP requirements are followed [2,3,4]. Companies that fail to follow GMP requirements can have serious consequences for that. Batches can be recalled, the operations may be seized and those who are accountable can be fined and even put to jail [11].

## 2.2 Computerized systems

Computerized system is a combination of software and hardware. In addition to software and hardware computerized system takes standard operating procedures, people and equipment also into account. [12] As a term, computerized system is widely used especially in pharmaceutical industry. If the system is used in GxP-processes, it must comply to according GxP-regulations and the system must be validated. Computerized systems can be anything from automated manufacturing equipment to document management [12].

To make sure that a computerized system complies with regulatory requirements and fits for its intended use, a life cycle approach is recommended to be adopted. Like a project, a computerized system can be viewed in different phases during its life cycle. The life cycle can be roughly divided into four major phases: concept, project, operation and retirement. In the concept phase requirements are set and the possible solutions are thought out. In the project phase various specifications are made and then verified and accepted to be released to operation. Identified risks are managed and removed if possible. Operation phase is the longest phase and requires managed operational processes, trained personnel and capabilities to carry out through changes. In the operation fitness for intended use and compliance must be ensured. Retirement is the final phase in the life cycle of a computerized system. Retirement phase includes questions about data retention, possible migration and destruction of the system. These processes must be carefully planned out and managed. [12]

Within GxP-regulated environments, certain roles are determined to differentiate responsibilities that e.g., a computerized system has. Every business process has a process owner whose responsibility it is to ensure that the computerized system fits for its intended use. Process owners are often the head of the function that uses the computerized system. Process owners own the data on the system and are responsible for the integrity of the data. Along with process owner there is also a system owner linked to the system. System owner supports the system and ensures its availability. System owners take care of the security of the systems and its data. System owner also takes responsibility for the maintenance activities of the system. [12]

Computerized systems in GxP-regulated environments are only one example of systems and their engineering. The same life cycle approach can be applied to all systems. There are many models for the life cycle of a system, but the main points are the same. These models include a traditional top-down model, a waterfall model, spiral model and object-oriented design for instance. In the life cycle of any system, everything starts from identifying a need. Then the system is designed based on those needs. The system is in operation until its retirement. The attention put on a system during its life cycle can be called engineering of a system. [13]

## 2.3 System validation

In GxP-regulated processes a computerized systems must be validated to ensure their fitness for their intended use. However, it is essential to understand the difference between qualification and validation. The goal in both methods is to ensure GxP-compliance and fitness for intended use, but they differ in scope. Qualification is often done for automated manufacturing equipment whereas validation is done for computerized systems [12]. If a computerized system is part of an automated manufacturing equipment, a production machine for example, the computerized system does not need to be necessarily validated separately. In that case the production machine is qualified, and the computerized system is part of the qualified machine.

Computerized systems that are GMP related should always be validated, and the validation thoroughly documented [7]. The two key things in system validation are GxP regulation compliance and fitness for intended use. In the life cycle of a computerized system the validation is part of the project phase. This phase includes planning and setting the requirements for the system (see chapter 2.4 of this thesis). Additional specifications and configuration of the system are made. The system is then tested by verifying it against the set specifications. When the specifications are tested and verified the system can be released to operation. [12] This process is depicted in the GAMP5 V-model as shown in Figure 1. GAMP stands for Good Automated Manufacturing Practices. The validation process that follows the V-model is similar to the traditional top-down model used in systems engineering generally [13]. The same principles for validating a computerized system in GxP-environment and taking any other system into use are the same.

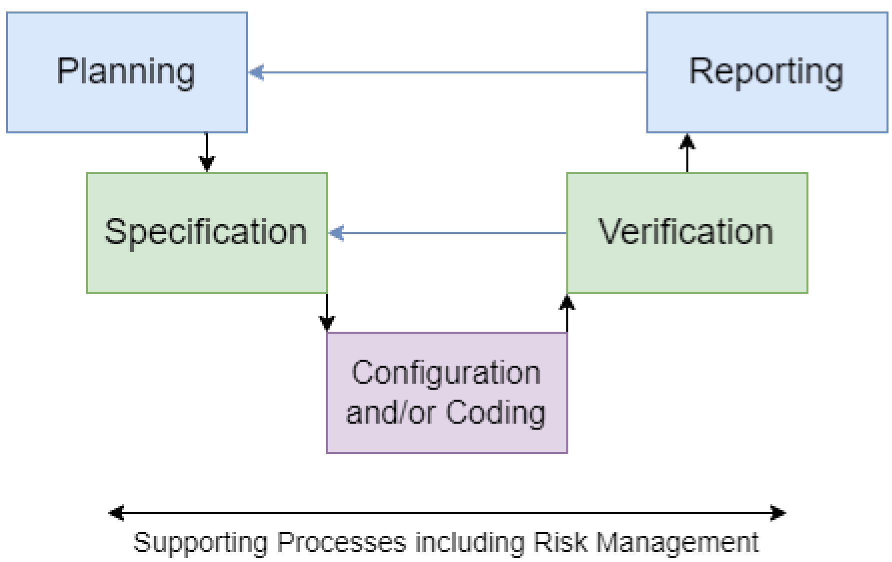


***Figure 1.*** *The GAMP5 linear V-model for achieving compliance and fitness for intended use, adapted from [12].*

Figure 1 depicts roughly the validation process. The model is V-shaped and is read from left to right and top to bottom. Everything starts with planning. Based on plans from the previous phase, specifications are made. According to the specifications, the necessary configuration and coding is done to the system. Then the system is put to test and verified against the specifications. If the system meets the requirements and specifications, it can be put into operation. During operation the system is monitored, and reporting is carried out. When a computerized system is more custom and complex, more specification is needed. Specifications can be divided into requirements specification, functional specification and design specification. Occasionally, additional specifications are needed, for example hardware design specification and software design specification. [12]

## 2.4 System user requirements

User requirements are what an end user wants from a system to meet the intended use. User requirements are defined in the early stages when projects are planned. A document called *user requirement specification* (URS) is created based on the carefully drafted requirements. Later in the project the requirements are verified against the specification to ensure that the system meets the intended specifications. User requirements should always focus on the business process – what the business would want from the system. Requirements need to be compliant to GxP regulations, and thus must ensure that patient safety, product quality and data integrity are addressed. [12]

Requirements for computerized systems should reflect the business need, processes and data workflows. Good characteristics for user requirements follow a mnemonic called SMART. The letters stand for specific, measurable, achievable, realistic and testable. Along with SMART rules, the more detailed the requirements are the better. The types and number of requirements can vary. Different types of requirements can be for example operational requirements, functional requirements, technical requirements and data requirements. However, typically there are a lot more requirements listed in the scope of a computerized system. [12]

Requirements in GxP regulated environments are defined the same way they are defined in systems engineering generally. In a system's life cycle, everything starts with addressing the needs and objectives for the system. These objectives and needs will determine the requirements. Furthermore, the requirements grant the means for a validation of a system's design in the qualification process. In systems engineering the requirements can be seen as indicators that measure the success of the system design. [13]

The requirements can be divided hierarchically. The top-level requirements are called mission requirements. The mission requirements reflect the needs and objectives of the stakeholder in a bigger picture, and not just the system that is being designed. The stakeholder in this context means the user of the system. Stakeholders' requirements are defined next in the context of the mission requirements. Stakeholders' requirements set the constraints and performance parameters for the system to be designed. From the stakeholders' requirements system requirements are derived. System requirements are more detailed and technical. Finally, the system requirements are divided into components and configuration items for the system. The requirements are then refined into a system requirements document. [13]

In this thesis user requirement specifications of two different systems made by the target company are analyzed and compared. The whole contents of those user requirement specifications are not covered in this thesis but there are certain segments every user requirement specification should have. Introduction should state who is the author of the document and for what purpose is the document made for. An overview should provide information about the overall goal of the system, the scope of the system and how it impacts patient safety, product quality and data integrity. Operational requirements are the part that is later verified to make sure a system meets its requirements. Operational requirements should have the functional, technical and data requirements depicted. Interfaces, nonfunctional attributes and environment requirements can also be part of the operational requirements. Requirements that are GxP related must be applicable to regulations and be considered carefully. [12]

# 3. ENTERPRISE AND CONTROL SYSTEM INTEGRATION

*Industrial control systems* (ICS) can be found in every industry and pharmaceutical industry is not an exception. Industrial control system is an umbrella term for various control system types. *Supervisory control and data acquisition* (SCADA), distributed control systems (DCS) and programmable logic controllers (PLC) are all part of an ICS. An ICS consists of various electrical, mechanical, hydraulic or pneumatic components that control a desired process. The control can be either automated or manual and can work in open or closed control loops, and a single control system can have countless control loops. [14] This thesis focuses particularly on SCADA in the context of industrial control systems. However, the difference between SCADA and ICS in general helps understand the bigger picture in control system integration.

This chapter of the thesis gives a brief introduction to industrial control systems and their general role in relation to the ISA95 automation standard. We take a look at the characteristics of data at different levels in the automation hierarchy. In this chapter SCADA and data historian are covered at a general level and are given a definition. Further in this thesis both systems are discussed in the context of the target company's manufacturing process. This chapter also discusses the concept of industry 4.0 and how it changes the way we perceive manufacturing.

## 3.1 Automation and data

Good approach to examine different levels of automation is the ISA95 automation pyramid. The pyramid introduces different levels of automation and operation as depicted in Figure 2. The pyramid has often five levels and the levels are often presented from zero to four. Data is gathered on all levels of the pyramid. The data is sent to the upper level and decisions are made based on that information. The ISA95 pyramid model is a good way to define the integration of control system functions and enterprise functions. However, with the coming of Industry 4.0, which is said to be the fourth industrial revolution, the relevance of the pyramid depiction can be questioned. The new approach could be a network-structured architecture where the hierarchy of levels is not present. The new approach would still be in compliance with the ISA95 standard [15].

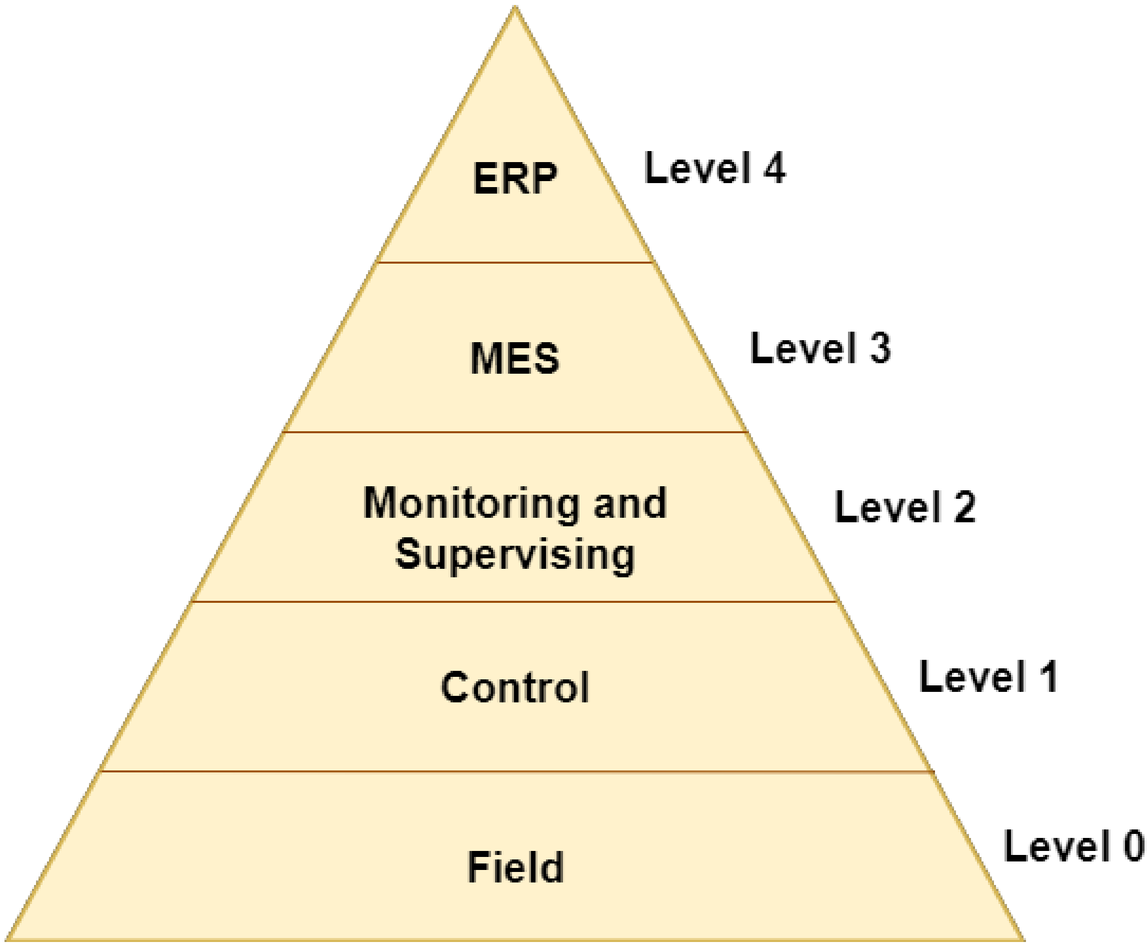


***Figure 2.*** *ISA-95 Automation pyramid, adapted from [15].*

In Figure 2 the levels of automation are presented in a form of a pyramid. On the bottom level of the pyramid is field level which is closest to the physical processes done by production machines for instance. The field level includes sensors and other actuators that directly affect the production process. The second level from the bottom is control. Control is the level where the PLCs are that control the process. The PLCs get information from the sensors and makes decisions based on that. On top of control level is monitoring and supervising. Monitoring and supervising level includes SCADA systems. SCADA systems are used to supervise, control and monitor the production processes. SCADA systems also gather data from the production processes and that data can be sent to other systems. On top of monitoring and supervising level is often another system called MES. MES stands for *manufacturing execution system*. MES often collects and maintains records of the production process. MES can also be used to optimize the production processes. On the highest level is ERP which stands for *enterprise resource planning*. ERP often covers a whole facility and other systems are integrated to the company's ERP system. ERP collects data from other systems, and based on that data, decisions about sales, shipping and storage are made. [15]

One way to separate the different levels of automation is to view the time it takes for functions in every level of the automation pyramid. When going up the pyramid from the bottom the time for functions inside each level increases. The lower on the pyramid the more real-time the operations are. On the lowest level where the physical processes happen, functions can only take microseconds. On the second and third level time frames for functions can be seconds, minutes, hours or whole shifts and batches. For MES systems functions often cover the time to produce one batch for example. Thus, the time frame is days or weeks. On the ERP level decisions are made with a long scope in mind and the decisions can have impact many months. [15]

## 3.2 Supervisory control and data acquisition

SCADA stands for *supervisory control and data acquisition*. SCADA is a vast term, and it does not have a single factual definition. Industrial control systems (ICS), Distributed control systems (DCS) and SCADA are all related to each other and the lines between them can get blurred but there are differences. Industrial control systems (ICS) encompass the other control system types and SCADA is one type of a control system [14]. However, SCADA can be defined as a system that is a combination of software and hardware [16]. SCADA systems are used to control assets and to gather data. Both roles are as important for a SCADA system. [14]

SCADA system comprises of software and hardware components. The scope and sophistication may vary but generally there are a few main components in a general SCADA system. A SCADA system has a control server, or a master terminal unit (MTU), which can be located in a control center. In the control center connected to the control server are *human machine interfaces* (HMI) where users can monitor the activities. Some engineering workstations and different kind of communication equipment are also part of the control center. On the shop floor near the physical processes are remote terminal units (RTU) and PLCs which are used to control and monitor the sensors and actuators. [14]

A depiction of a general SCADA system is presented in Figure 3. Control system including HMI, engineering workstation, control and data server can be seen in the control center. Control center is connected to a field site with network communication which allows information to be exchanged between the two. A PLC can be found on the field site, and it can be operated from the RTU.

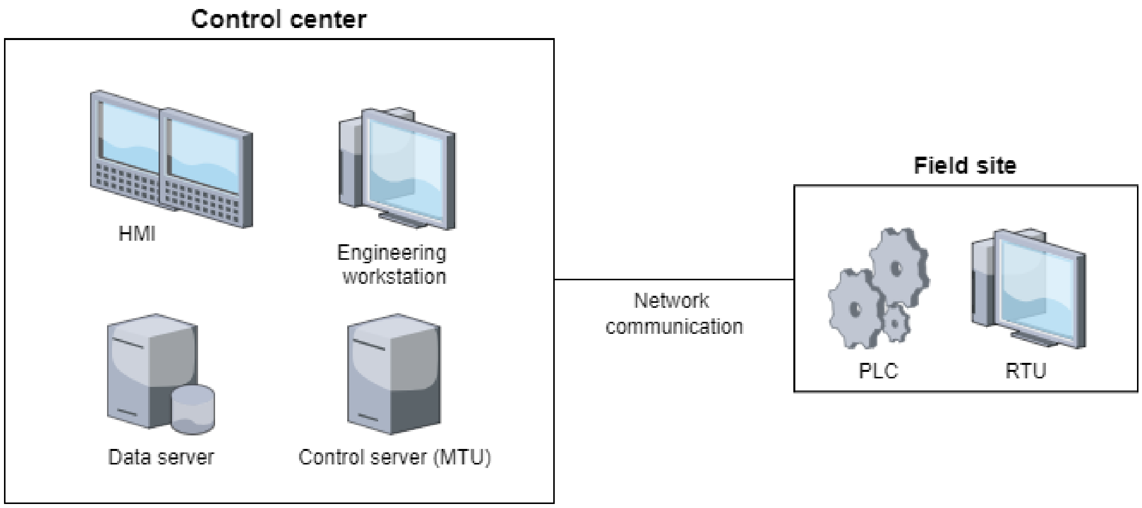


***Figure 3.*** *General SCADA system architecture, adapted from [14].*

One of the SCADA system's tasks is to collect data from the field and transfer it to the control center. In the control center the data can be processed and then graphically presented to an operator. Based on that information the operator can then monitor and control the processes in almost real time. Control center logs all information collected from the field and displays it on the HMIs. This information can be used to make actions based on events noticed in the log. The control center also collects information about alarms and trends. For example, if a certain parameter is out of its range, an alarm is shown on the HMI in the control center and corrective decisions can be made upon that. [14]

SCADA systems are used for various purposes. A SCADA system can be used to control and monitor a single production machine and its production process. On the other hand, a SCADA system can be hugely complex and have countless of components and it can be used to control and monitor an entire smart electric grid. Not all SCADA systems follow the same architecture and have all the same components. On some occasions a physical control center may not be necessary. All the SCADA applications can run on a server and are accessible remotely. The HMIs and engineering stations can be situated on the shop floor, and they can work as clients to the applications that run on the servers. All the necessary information and functions can be displayed on the HMIs on the shop floor without a need for a control center.

## 3.3 Data historian

Data historians can be recognized as a part of a SCADA system [14,16]. At its simplest, a data historian can be the database of the SCADA system that just stores the data collected by the

SCADA system [16]. A slightly more sophisticated description of a data historian could be a database that has tools for data analysis and statistical process control [14]. The latter is closer to the one that is being used in the target company. When the data historian is more than a database and has features on its own and is not dependent on the SCADA system, it can be viewed as its own system. In the scope of this thesis data historians are viewed as such. Data historian in the target company is further examined in chapter 4.3.

Unlike SCADA systems, data historians do not have a large variety of hardware components. Data historians can be a set of software modules that collect and store information and then process the information for it to be further used [17]. There is no single way to depict the hardware components of a data historian since most of it is just software. What separates data historians from others is what data they collect, where they collect it, how they collect it and what they do with the collected data. A simple way to depict a data historian is to divide it to three main components: data collectors, historian server and clients [18]. Data collectors collect and analyze the data, historian server is where the data is stored, and clients retrieve the data from the server to be further analyzed.

Along with the hardware components, there are a variety of services in a data historian configuration. In a data historian architecture, there is an interface which is used to collect data points. From the interface the data is passed to a data archive server. In the architecture is also an asset framework server with its own relational database. The asset framework server is used to structure the data from different sources. Along with the different databases, there can be other services in the configuration such as analysis service. The analysis service is used to run and configure analyses from the data. Other services include notification services and management tools. Altogether they make a data historian. [19] A similar configuration is used in the target company. A closer look of the system is taken in chapter 4.3.

Data stored in the data historian is valuable and it has a lot of beneficial use cases. Many different functions can benefit from the data and analyzes based on the data. Data historians can be used as a powerful tool for decision making for the management. However, not only management can utilize the data. The data can be processed and visualized in forms that can help the operators and engineers in their daily operations. Analyzing trends and unexpected downtimes for example, can help increase the performance of a process. Management, on the other hand, can use the data to calculate key performance indicators (KPI). [17]

## 3.4 Industry 4.0

In the last decades technology has rapidly developed. The traditional ISA95 pyramid model may not be applicable in modern industrial manufacturing. New models are made to replace the old hierarchical pyramid model. Discussion about digitalization and artificial intelligence can be heard anywhere. Manufacturing industries have not avoided those topics either, although technological trends tend to affect industries with a delay. In the context of industrial manufacturing the recent technological development and its consequences can be called industry 4.0. It is said to be the fourth industrial revolution, hence the name industry 4.0. It is a change that revolutionizes manufacturing as we know it [20].

The big concept behind industry 4.0 is internet of things and industry 4.0 is its application in industrial context. Industry 4.0 means digitalization in an industrial scale and the increase of information and communication technology in places where it has not been utilized before. At the center of industry 4.0 is that everything is becoming data driven. The physical and digital worlds collide into cyber-physical systems. In cyber-physical systems the production machines, actuators and sensors generate data all the time and send it to platforms that can analyze the data in real time. These systems can share data with each other and even make their own decisions. This kind of development shape factories to become so called smart factories. [20]

The key drivers in industry 4.0 are the ever cheapening and more powerful microchips and sensors, the number or devices that have access to internet and cheaper price for data storage. The combination of all these drivers have made analyzing huge amounts of data in almost real time. The data can be used to optimize automated processes to make them more efficient or it can be used to predict the need for maintenance in the production equipment for example. When an asset can be completely monitored and controlled remotely, predictive maintenance is possible. With predictive maintenance it is possible to increase the life span of a production machine and decrease the upkeep costs. [20] Part of a development like this are the systems like a centralized SCADA and a data historian. These systems collect a lot of data and makes it possible to control assets remotely.

The benefits of industry 4.0 are obvious. It makes manufacturing more cost-efficient, safer and by increasing efficiency also increases revenue. There are three ways how industry 4.0 can help companies financially. Firstly, it increases sales of the existing products, and industry 4.0 generates totally new products, services and business models. Secondly it decreases the operating expenses by making production more efficient and by decreasing service costs for example. Thirdly it decreases capital expenditures by optimizing the fixed assets and reducing storages.

Apart from the financial benefits, the qualitative benefits and the otherwise operational environment revolutionizing benefits cannot be undermined. [20] Industry 4.0 seems inevitable for every are of industry sooner or later. The sooner an industry is ready to adopt the change the better are the competitive opportunities.

# 4. CORE MANUFACTURING PROCESS

In the context of this thesis, it is important to understand the environment that we are investigating and especially the challenges it brings. Pharmaceutical manufacturing has a lot of specific requirements like as presented in the earlier chapters. The requirements go all the way to the rooms where the production is done. This sets special requirements for all the production equipment and the personnel alike. This chapter roughly describes the manufacturing process of the core of the product that is being made in the target company. The target company has a specific department for this process and that department, and its processes are in the scope of this thesis. This chapter then examines the SCADA system and the data historian from the perspective of the target company.

## 4.1 Introduction to the core process

The target company is a big pharmaceutical and life science company that has manufacturing sites all around the world. One of the sites is located in Finland and it manufactures certain pharmaceutical products. This thesis is made on behalf of that company to help the target site become more data driven and to help increase digital maturity. In the center of digital maturity and data leadership are the new computerized systems. This thesis focuses on one of the production departments in the target company's manufacturing site. The department is used for manufacturing core, made of a medicinal substance, for multiple pharmaceutical products manufactured on the same target site.

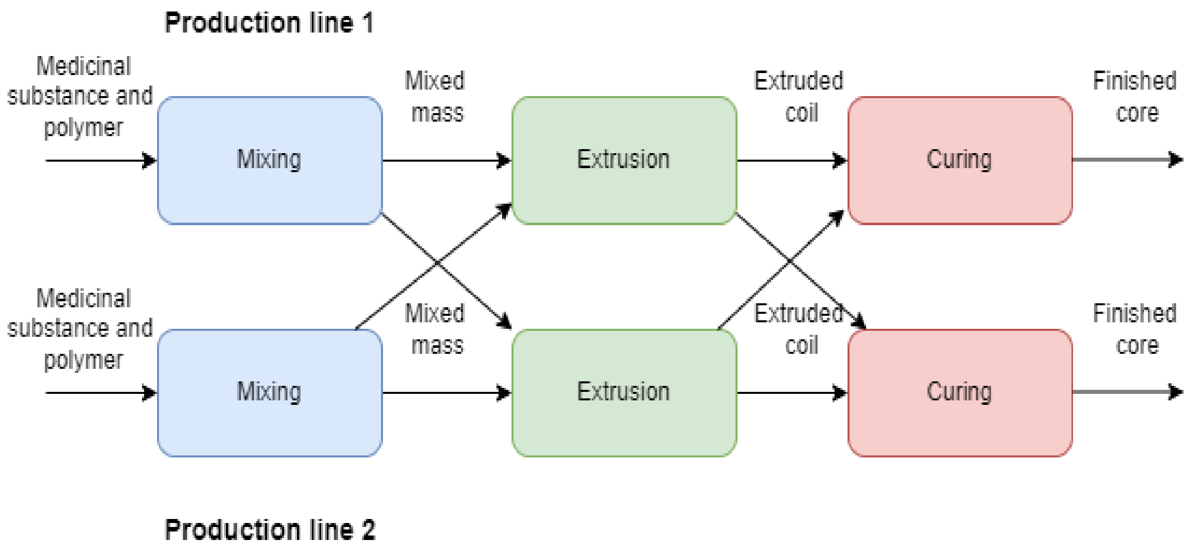


***Figure 4.*** *The three phases of the core manufacturing process on two production lines.*

The core manufacturing process can be divided into three main phases: mixing, extrusion and curing. Each phase has its own production machine, but they all share the same centralized SCADA system. The three phases are depicted in Figure 4. The department has two identical production lines that can be cross driven if needed. The first phase in the process is mixing where two main substances are mixed together into one mass in big, isolated mixers. In the second phase an extruder is used for the mass to be extruded through different kind of nozzles depending on the product. In the extruding process the core mass is also shock cured with ovens to modify the material characteristics. The output of the extrusion phase is a long and round shaped coil. In the last and third phase the coils are cured in an oven for a long time in a certain temperature in order to modify the material characteristics of the core mass in the coils. When the product has achieved desired characteristics, it is ready to be sent to assembly departments where the core and other components are assembled to a finished product.

The process is very complex and is done in challenging environments. One of the challenges is the need for cleanrooms. Production of sterile products such as pharmaceutics require very clean conditions. The idea of cleanrooms is to keep the number of particles in the air as low as possible. Therefore, cleanrooms require airlocks for material and personnel. Personnel are also required to dress appropriately in the cleanrooms to minimize all particles.

GMP legislation has requirements for cleanrooms, and a grading system can be used to determine the requirements of a specific cleanroom. The cleanrooms can be typically divided into four different grades where A is the “cleanest” and D is the “least clean”. The difference between the grades can be determined by the number of particles in air while in production and on the other hand while at rest. Manufacturing activities can happen in a C graded cleanroom for example. Such room requires that there are less than 352 000 particles that are larger than 0.5 $\mu m^3$ but smaller than 5 $\mu m^3$ and less than 2930 particles larger than 5 $\mu m^3$ while there is no operation in the room. [21]

Cleanrooms also set requirements for the equipment. From small sensors to big production machines, everything must comply with the cleanroom requirements. For example, all peripheral devices such as monitors, keyboards and mice must be qualified to use in cleanroom. Such devices are often many times more expensive compared to an equivalent regular device. The manufacturing processes for the product made in the target company is also unique and thus requires very specific and innovative production machines. The machines must be very robust and reliable. Modern production machines also generate a lot of data which is increasingly important in order to improve efficiency. Data is also a central part of this thesis and is examined closer in following chapters.

The three phases presented are a rough simplification of the process. There are many supportive processes happening around the manufacturing process itself. This thesis focuses solely on the manufacturing process itself, but nothing would be possible without the supporting functions and processes. Environmental monitoring for the conditions in the cleanrooms for example, is essential for the operations. If the condition requirements are not met, the operations will halt. Throughout the manufacturing process a lot of electronic documentation and reporting is made. Batch reports for example, are important documents that are used to track the performance of each batch. The department requires the contribution of all functions to operate the production, and new computerized systems are adopted to support everyone’s contribution.

## 4.2 SCADA system in the core manufacturing process

The core manufacturing process utilizes multiple systems, but the SCADA system plays the biggest role. The SCADA system used in the process is a commercial product that has been customized to the core process’ specific needs. The SCADA system also has multiple add-ons that improves it capabilities. The SCADA system is in place and its being used but at the time of writing this thesis, it has not been connected to any other major systems such as a data historian or a MES system. However, both systems are going to be integrated at some point. Because the data historian system has overlapping capabilities with the SCADA system, this thesis tries to discover the benefits of integrating the data historian system. The SCADA system used in the core manufacturing process is roughly depicted in Figure 5.

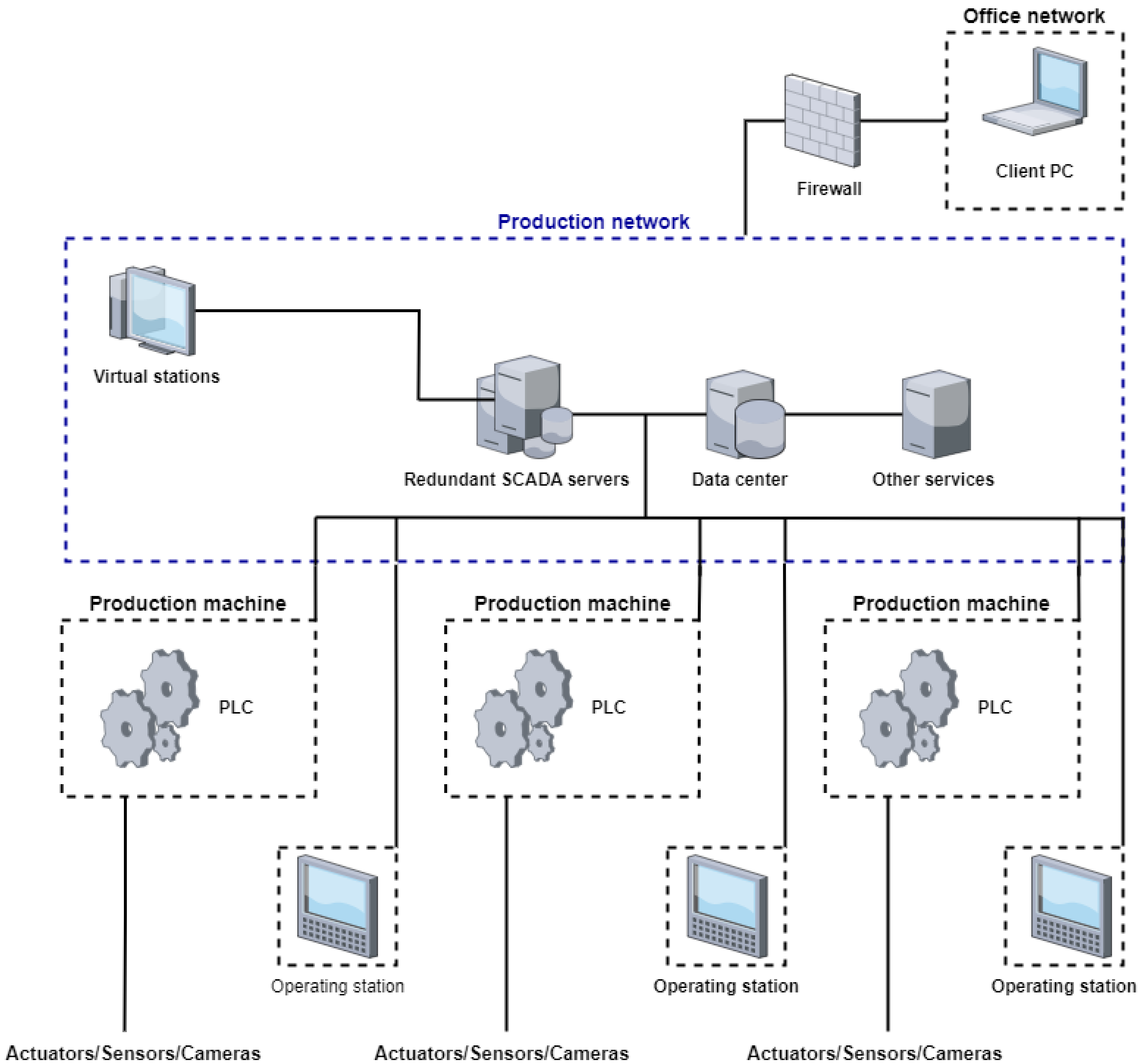


***Figure 5.*** *Centralized SCADA system architecture in the core manufacturing process.*

Figure 5 shows an example of a centralized SCADA system. The SCADA is divided into multiple components. In the SCADA example office and production networks are separated and production machines are also shown as their own entities. The idea with a solution like this is that one can have multiple production machines under one SCADA system. The one SCADA system controls the operation of all machines and collectively gathers data from all machines and their operations. This kind of architecture may classically resemble more like a DCS than a SCADA because of the complexity and scope of this architecture [14]. However, the distinction between SCADA and DCS may not always be clear, and their functions may overlap [22].

The SCADA system works on a server-client principle. The SCADA application runs on the virtualized redundant servers. Redundancy means that if one of the two servers faults the other one is used without any significant delay. In other words, the servers are identical servers that work in parallel. On the shop floor the operating stations (HMIs) work as clients. The clients connect to the servers and load the SCADA project from there. Everything that is done on the clients gets saved to the servers. No data is stored on the clients themselves. Normal users cannot access the servers but can only operate the machine through the clients. With the redundancy, this makes the SCADA system very robust and secure.

The depiction in Figure 5 is a simplified version of the real SCADA used in the core process, but the main components can be perceived. Unlike in the general SCADA depiction this architecture does not necessarily have a control center. The heart of the system are redundant virtualized SCADA servers where the application is running. The servers are in an isolated and firewall protected production network. Only IT administrators can access the servers in accordance with the principle of least privilege, which means that the least possible amount of user access is granted to each user group. Therefore, no user input is needed on the server side for the system to operate.

In the system architecture along with the redundant SCADA application servers is at least a data center server. The data center has a relational database where important production information is stored. The data center acts as a data storage for historical data, and batch data and audit trail are also stored in the data center. Audit trail is an electronic record that enables traceability for every user action. For example, every time when a user logs in to the system or starts a batch an entry is made to the audit trail that shows who did what and precisely at what time.

The SCADA system has other servers hosting different services in its architecture as well. They are not independently depicted in Figure 5 but in real configuration, there are other servers in the architecture. One of such services could be a server where recipes are created and then sent to the SCADA application. Other example of additional services could be a server where production data can be analyzed. Such service can be used to search trends in the production data and visualize the data for example. All the servers and their services are located in the same network and can be accessed remotely by IT administrators. Virtual operating stations can also be in the same network. From such virtual station one can view the SCADA project and configure it if needed.

There are critical components of the SCADA system also on the shop floor. PLCs control the machines and are connected to SCADA system as well. PLCs and the SCADA application communicate with each other and share critical information. PLCs are connected to numerous actuators and sensors that the PLCs control. They communicate in their own isolated machine network. However, the PLCs can also communicate to the production network since the PLCs are connected to the SCADA system. Automation components communicate in the machine network, whereas other software and the SCADA application communicate in the production network. Components in the machine network and production network are isolated but all part of the same system on a whole.

Each production machine has an operating station. From the operating stations operators can control certain functions of the SCADA system. The operator can view the state in which the machine is and change the state depending on the operator's user level. For example, maintenance and engineers have different user rights compared to an operator. From the operating stations, operators can start and stop the batch and follow the machines operation. The operator can also check alarms and audit trail and sign electronic signatures when the process requires one.

Modern production processes utilize machine vision. Machine vision Machine vision utilizes combination of hardware and software. Machine vision systems require cameras, PCs and software that is running on the camera PCs. The cameras and the camera software communicate in the machine network to the PLC.  Machine vision is used to observe desired phases of a process. It can observe the thickness of the core coil for example. Certain parameter ranges are set for the camera system and if the observed product's parameter is out of the boundaries the camera system sends information to the PLC to reject that product.

On the ISA95 automation pyramid a SCADA system is located at level two. However, the SCADA used in the core manufacturing process includes elements from other levels as well. PLCs are traditionally on the control level, level one, but in this SCADA, PLCs are an integral part of the same system. Of course, the PLCs still operate the same way PLCs usually do on the control level but in this architecture, they are in the same network as the servers and use ethernet communication between them. Therefore, the interface between the two levels gets vague, and putting the SCADA system purely on one ISA95 level is impossible.

## 4.3 Data historian in the core manufacturing process

In the core manufacturing process data historian and the SCADA system are two completely separate systems. However, the two systems are going to be integrated, regardless of this thesis. Since the two systems have a lot of overlapping capabilities, this thesis tries to find the best use cases for each system. The goal is to have the two systems create as much as value for different functions as possible.

The main purpose for the data historian is to collect production data. The data can be retrieved with multiple different ways. Either directly from the production machines and their databases or from another system. The data can be collected directly from the PLCs using the OPC UA protocol for example. On the other hand, the data can be collected from local databases on machines or from a centralized data archive. This makes it hard to place a data historian unambiguously into the ISA95 standard framework. As important as collecting the data from production processes is visualizing the data. One of the biggest benefits for different users is the data historian’s capability to visualize data. Not only can the data historian visualize the production data as is, but it can also perform computing. The production data can be analyzed and calculated into whatever trend or formula is interesting or otherwise required.

Figure 6 depicts a hypothetical architecture for the data historian. It gives a general example, how the system architecture could be built. How exactly the data historian would look like on the core manufacturing department, cannot be indicated because the system has not been integrated. However, there are some elements that persist in the architecture despite the configuration and use case. The data historian architecture consists of different layers in different network partitions. These layers can also be thought as the ISA95 standard levels. The architecture starts from the production machine, PLC, level and goes all the way to the end user in the business network. In between the end points there are the essential components for the data historian to operate.

The network partitions do not represent the ISA95 levels unambiguously, but they can be indicative. The partitions can be thought to usually have such systems and services as their corresponding ISA95 levels. This is one reason, why the data historian is hard to place on the ISA95 automation pyramid. The data historian can have interfaces and connections from the bottom of the pyramid all the way to the top. Another reason is that a data historian is traditionally seen as part of a SCADA system [14,16]. Therefore, the ISA95 model may not be the best way to present data historian systems.

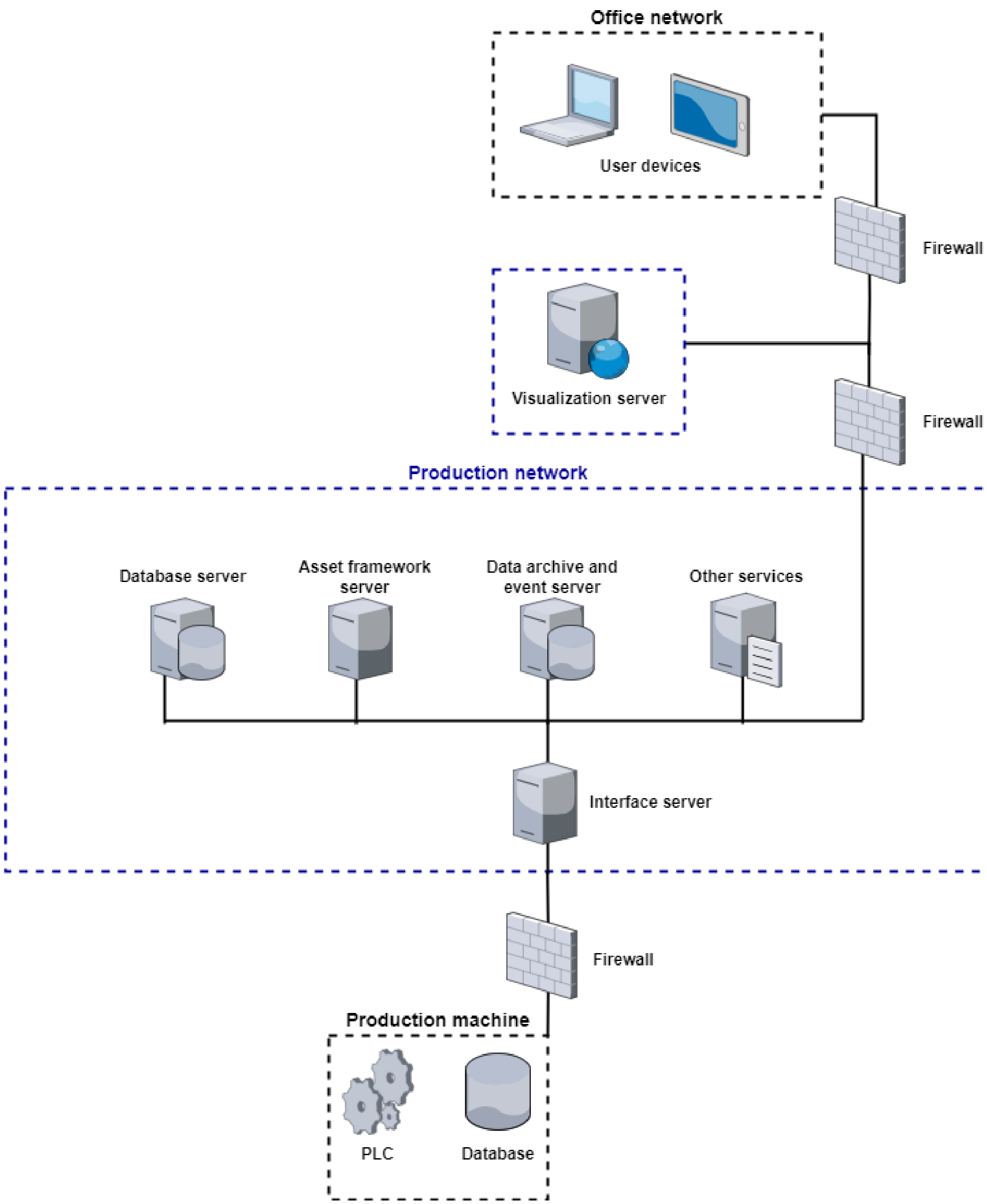


***Figure 6.*** *An example of a data historian system architecture.*

The first task for the data historian is to collect data from different sources. The source can be a PLC using OPC UA connection, or a relational database for example. The database is usually an SQL database (structured query language), meaning the data can be accessed with structured, standardized, queries. From the machine the data is transferred to an interface server using the preferred communication method. The data is formatted accordingly on the interface server.

There can be more than one interface server to improve fault-tolerance and to share the load if there is a large quantity of data sources. The data is then transferred further from the interface server, usually to a data archive. The interface server is in the same network partition, and ISA95 standard level, as most of the other data historian components. Between every level and network partition is a firewall to block all unwanted communication. Only the necessary communication protocols and their corresponding ports are allowed.

Other servers in the same level as the interface server include at least a database server, an asset framework server and a data archive. Other services are also on the same level. These services can include different kind of tools to view events and alarms and other notifications for example. The asset framework server and the data archive are the key components of the data historian system. The data collected can either be relational or time series data for instance. The data can be collected first to the data archive server. From the data archive the data is sent to an asset framework server where the data will be structured and organized. The data can also be requested straight from the data archive where it is not yet structured. The asset framework standardizes the data even if the type of the data varies depending on the source. The data can be everything from measurement data of a production machine to environmental conditions data. The job of the asset framework is to harmonize and contextualize the data.

On the next network partition that is closer to the office network and thus, upper in the ISA95 pyramid, is a visualization server. The visualization server has the tools to visualize data and create displays for end users. Apart from collecting and archiving data, visualization is perhaps the most important aspect of the data historian. The data historian provides a powerful visualization tool for the end users. With the visualization tool the users can follow the performance of a machine, shift or a batch for instance. The performance can be plotted on a graph to help the users understand the data. Analyzing those graphs can help the users, whether it is an operator, an engineer or a technician, notice anomalies in the data and thus, optimize the performance based on that information.

With the visualization a display that shows an overview of the performance of a whole department can be created. With a quick glance, the user will get a good understanding of what the current situation in the department is. This is a huge advantage when creating data driven processes and modernizing manufacturing. An example of what a department overview in the data historian visualization tool could look like is depicted in Figure 7.

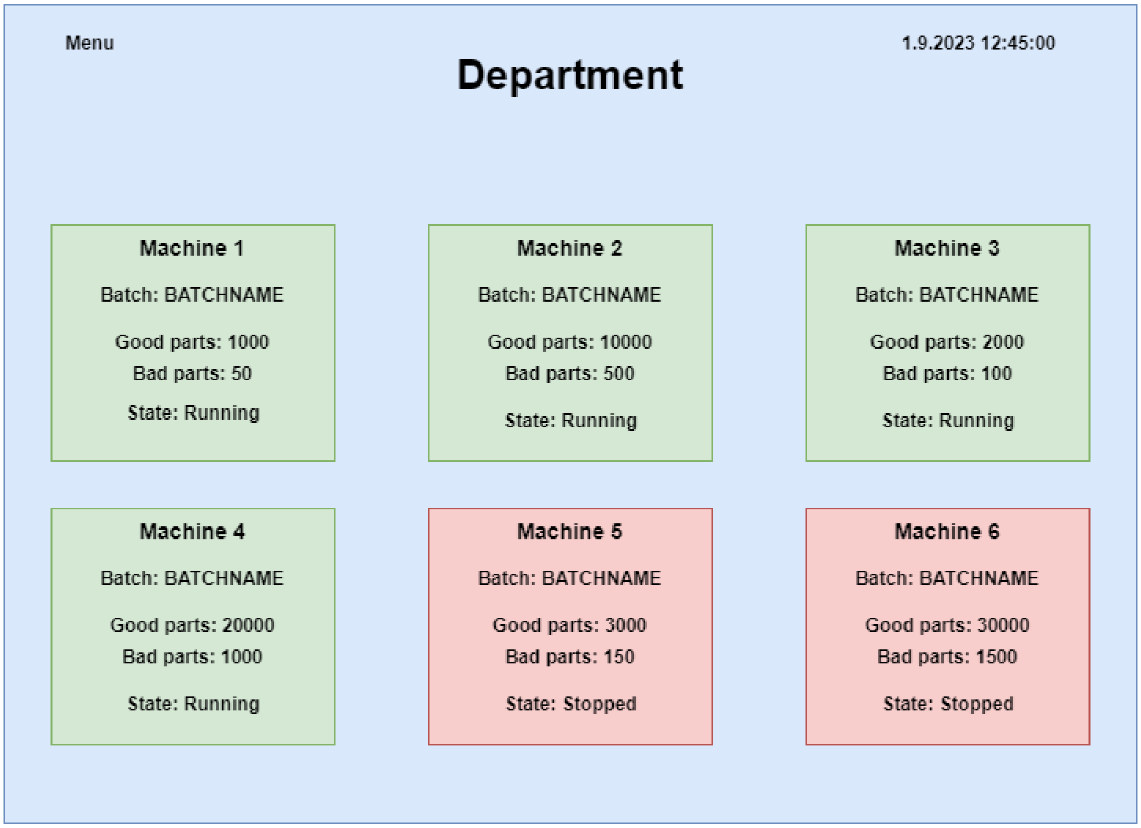


***Figure 7.*** *An example of a data historian overview display of a manufacturing department.*

In the overview users can see multiple machines at the same time as shown in Figure 7. The overview would give the user information about whether a production machine is running or not and how it operates. It can also show the batch that is currently ran on the machine, and how many parts has been produced in that batch and how many of them are rejected bad parts. The user can then choose a machine, click the icon, and get more details about the performance of that machine. This kind of overview display is only one example of the visualization opportunities. Basically, the user can choose to visualize anything that is possible to squeeze out of the available data. With calculations the data can be revised to countless forms and visualized thereafter.

## 4.4 Manufacturing execution system and other systems

The SCADA system is not only integrated to the data historian but also to other systems. These systems are not covered with the same emphasis because this thesis focuses on the data historian and SCADA particularly. However, there are many other important systems used in the core manufacturing process. One of these systems is manufacturing execution system (MES), which is prominent on the ISA95 automation pyramid as well. MES usually represents the level three on the pyramid. MES system can be described as a software system that works as a layer between

ERP and process control systems [23]. MES system controls, tracks and documents the manufacturing processes all the way to finished products.

Other systems in the target department include machine specific software that are not part of the SCADA system. The production machines in the core department are big and complex machines and can have many sub systems in them. Thus, one production machine can have multiple different software. The main connecting factor is that they are in the same production network. Different kind of scales with high accuracy are also used in the department to measure the weight of various substances. Scales and other small devices also have their own software in them. They are not part of the SCADA system, the data historian or the MES system but they can be connected to them via different interfaces.

The MES system is not integrated to the core manufacturing process at the time of writing this thesis. However, in the future MES will play a key role in the department. In the future the MES system will also be integrated to the SCADA system. However, their relation is not examined further in this thesis. The purpose of the MES system is to manage equipment, design and execute batch records, keep record of the batches, handle material flows and control and optimize processes. The MES system is used to manage and document the entire manufacturing process in real time. The MES system is also used to connect the manufacturing processes to higher factory wide systems such as ERP systems.

# 5. BACKGROUND SURVEY OF THE DEPARTMENT NEEDS

In this thesis, finding the roles for the SCADA and the data historian system is the objective. The approach to find the roles is to take a look at the user requirements of each system, and thoroughly analyze them. A survey was done as a part of this thesis to find out, what the key persons in the target department would want to know from the manufacturing process. The answers of the survey will reveal the genuine needs of the department. Based on those results, the user requirement specifications are analyzed and from the specifications the best system to address a need is tried to find out. This chapter introduces the survey questions and their corresponding answers. The answers are then analyzed. In the following chapters the requirements are thoroughly analyzed and then reflected upon the analysis of the survey answers.

## 5.1 Survey questions

The target department in this thesis is the core manufacturing department. The goal of this thesis is to provide valuable information to the department and its personnel. In the best case, this thesis will help find the optimal roles for the SCADA and data historian system. This would further help the department become data driven. In order to ensure the department's perspective, a survey was conducted for people that work within the core manufacturing department. The people that the survey was made for represent different roles in the department. The roles include operator, production engineer, data engineer, supervisor and department manager. The survey is not supposed to be a research method but work as a supportive background info instead. Thus, the answers are not transcribed in this thesis. They are, however, analyzed and summarized. The findings from the user requirement specifications are then reflected upon the answers of the survey. In doing so, it is ensured that this thesis servers the department and its needs the best.

The survey consists of five questions. The questions were set to be as open as possible. The intention was to give the respondent as much as freedom as possible to think the answers and not guide them in any particular direction. This way, the idea was to get valuable information about the real needs, and about the biggest challenges and hindrances the people working on the target department face.

The questions are the following:

1. What information would you like to have to support your work at the core department that the machines and systems do not currently provide?
2. Do you feel that some helpful information/data regarding the production processes is not available?
3. Do you feel that there is information/data available that is not being utilized?
4. Could the available data be visualized better?
5. Are there already too many computerized systems or could additional systems bring more value to the core department?

## 5.2 Analysis of the answers

The respondents' answers are analyzed in this chapter. Every question is divided into their own subchapter. Not every respondent's every answer is examined separately, but the answers are summarized, and the key points are raised instead. However, if there are contradictions between the answers to the same question, the contradictions are then examined. The respondents work in different roles in the department. The role is addressed with every answer, but otherwise the responses are not individualized.

### 5.2.1 Question one

The first question was, what information would you like to have to support your work ate the core department that the machines and systems do not currently provide. On the first question a data engineer wished that there would be an easy and an accessible way to get a cross section of the current situation in the department from the production data. The data engineer wanted to have this data exported to their own computers and then analyze the data as they will.

Another data engineer wanted to have more parameter data available overall. The respondent also wished the possibility to access the production data in real time remotely. The people that are not working on the shop floor may want to analyze the data from their own workstations. The data engineer also wanted to have more capabilities to examine trends of the production data. The respondent wanted these trends to be easily accessible, regardless of where they work at.

The data engineer also wished that there would be an automated system to spot and notify about significant deviations in the trends.

A production engineer wanted to have more information about the environmental conditions of the department for the machine operators on the shop floor. If a door is left open for example, which affects the conditions, the operators will not get information about that. This is not strictly related to the manufacturing processes itself but is crucial information for the department, nonetheless. A supervisor, on the other hand, thought that all the necessary information is available. The supervisor noted that the problem is that it is yet unknown what is relevant information and what is not.

An operator on the department thought that the SCADA user interface is counter intuitive and too complex. The operator wanted the user interface and the usability to be improved so the information would be more easily accessible. The operator also said that the alarm notifications are not informative enough and they are hard to find. The operator also agreed that they wanted to have more historical information. Historical information about the batches that have been produced, and what has happened during that batch. The respondent also wanted to have historical information about a single production machine – how it has performed on a certain timeframe. The operator, like the production engineer, also wanted to have more information about the environmental conditions on the department. On the other hand, the respondent also admitted that the operators are not interested in analyzing the data thoroughly.

A department manager said that the new department is producing vastly more production data compared to the old department. In the new department it is possible to analyze the variances in the processes. The department manager stated that every solution that increases measurements and data gathered from the production lines and the products improves the processes overall. The respondent said that the more data the production machines and their systems generate the better. Then the data needs to be automatically processed and later decide in detail, what to do with the initially processed data. The department manager also pointed out that it is not clear what is relevant data and what is not on the new core manufacturing department.

### 5.2.2 Question two

The second question was, do you feel that some helpful information/data regarding the production processes is not available. On the second question a data engineer raised that the way the production is tracked is maybe not done the best way. Currently, the production is tracked based on the number of manufactured core coils.

On the other hand, a production engineer thought that there are a lot of sensors that collect enough data already. Another data engineer agreed on this and thought that the existing systems already gather measurement data and there are enough parameters. The data engineer pointed out that this is a huge improvement compared to the old department, where not enough data was collected. A supervisor agreed that the machines and their systems gather enough data already. Although, the supervisor noted that all historical data that affects the production processes is of interest. Data that is irrelevant for the production processes is sometimes felt like it is an irritant.

Although the core manufacturing department is sophisticated and highly automated, there are some manual work stages. An operator pointed out that the time consumed by these manual work stages is not monitored. There can be a lot of variances in the manual stages depending on whose doing it. This causes the time for a batch to be produced change significantly. The operator also wanted to have more information when a certain set of materials is used because there are also a lot of variances in the seemingly same material. Although the substance and the manufacturer are the same there can be variances in the material qualities. The respondent wanted to have more historical information regarding the material used in a certain batch.

A department manager thought that enough data is gathered on the new department, but it is not clear yet, which is relevant data, and which is not. This is because the manufacturing processes are somewhat new on the new department. The department manager also said that there are some manual inputs on the SCADA system that are not gathered. These kinds of inputs could be useful to determine the time it takes for certain processes.

### 5.2.3 Question three

The third question was, do you feel that there is information/data available that is not being utilized. On the third question the respondents were unanimous that there is plenty of data available, but it is not being utilized properly. A data engineer said that a lot of data engineering is needed for the available data. The data engineer wanted to have the tools to be able to refine the raw machine data to useful indicators. These indicators can be then analyzed by the department managers and data engineers in order to support decision making.

Another data engineer wanted to have more historical data available in the form of trends. These trends could help notice changes in the long run. These long-term changes can be hard to notice by only examining one batch at a time. On this question the department manager raised the manual inputs again like in the previous question. The times calculated from the inputs could help track the performance.

A production engineer wanted to have more information about the characteristics of the polymer coming to the department that is mixed with the medicinal substance. The characteristics have major effect on the mixing process. An operator again agreed that the information that is collected is hard to find or use. Navigation of the SCADA system is thought to be difficult. The operator also wanted to have summary button of all the functionalities. A supervisor thought that the data is sometimes hard to get out of the system as well. Excessive data may only complicate the processes and make the work too complex for the operators. The supervisor said that the systems must be simple enough for the operators to determine whether everything is nominal or not. Complexity of the systems and abundance of information makes this harder for the operators.

### 5.2.4 Question four

The fourth question was, could the available data be visualized better. A data engineer wished for more ways to visualize the data, even with the current tools available. For example, the current tools cannot be used to make histograms. The data engineer also wanted to have more monitors that could be used to display statistical process control (SPC) data. A supervisor emphasized that critical information about the production processes needs to be visualized well. Everything that is irrelevant should not be visualized at all. The supervisor pointed out that it is still hard to determine what is relevant information and what is not.

Another data engineer agreed that the SCADA system and its addons can already visualize the production data sufficiently enough. Although, the data engineer also said that SPC data could be useful. The SPC data could show the upper specification limits (USL) and lower specification limits (LSL) that determine whether the product is acceptable or not. A department manager said that the department has not yet thought of what the best system would be to visualize the data.

A production engineer thought that the existing systems can already visualize the data decently enough. The raw data can be collected and can then be visualized with other system. An operator raised that the time taken for a cleaning and mixing function are not visualized clearly enough. The operator also wanted to have trends about the material losses of a batch. Analyzing this information from the historical data can help find the root causes for the issues. The respondent also wanted to have historical batch data where colors are used to separate different products since the core can be used for multiple different final products.

### 5.2.5 Question five

The fifth question was, are there already too many computerized systems or could additional systems bring more value to the core department. On the fifth question a data engineer noted that the number of systems is not an intrinsic value itself, but the use cases and the accessibility of the data are more important instead. The data from different systems must be integrated and refined to be used by the end-users, for example the machine operators. Another data engineer thought that there are already enough systems. The data engineer emphasized that the personnel must first learn to use the existing systems.

A production engineer also thought that there are already enough systems and that additional systems will not bring more value. An operator agreed with the production engineer's opinion that there are already enough systems and additional systems would only complicate their daily work. The operator pointed out that it is frustrating to perform different tasks, that are necessary during the production, from multiple different places. They need to walk away from the production machine to carry out a specific task and then return to the machine. A supervisor agreed on this and said that to perform the production processes, multiple systems are already needed. The systems are dependent on each other, and the whole production stops if there is an issue with one of the systems. The supervisor said that it is frustrating that the production stops because of matters that do not directly affect the production. The respondent emphasized the fluent integration of the systems. The systems should be beneficial for the production processes and for the operators and make their daily work easier.

A department manager noted that there are enough systems as well. The department manager pointed out that there are enough planned or already implemented systems, but the more important thing is to integrate them properly. When the different systems are integrated well, the analyses from the data are comprehensive and depict different phenomena well. Combining the data will give enough information to understand, which phenomenon affect which process. The department manager thought that if more systems are taken into use, machine learning and automated data analysis software could be useful to point out causalities.

## 5.3 Summary of the survey

The answers of the survey highlighted a large contrast in the answers. Depending on the respondents' work and position on the department. As a rule of thumb, the closer the respondent's work is to the production machine and daily operations, the less interested they are in the data – or at least analyzing the data. On the other hand, respondents who work with data topics daily are naturally more interested in analyzing the data and refining the data. Respondents closer to the shop floor were surprisingly reluctant about new computerized systems, albeit the new systems are intended to support their daily work. The existing systems are already thought to be too complex and not user friendly. Bringing new systems means it takes time to learn how to use them and different work stages may require using multiple systems, instead of doing everything in one place.

The need for historical data was repeatedly pointed out by the respondents. Many respondents wanted to have historical data in order to follow trends about different phenomena. Trends from the production data could help detect anomalies and increase understanding of the processes on a whole. If there have been some problems on a certain batch or multiple batches, it would be easy to notice the differences looking at trends. This way the root cause for the problems could also be more easily found. The visualization of the trends also plays a big role in analyzing the data. The trends could point out when certain limits have been crossed. In addition to the need for trends, almost everyone could point out a need that is not currently fulfilled overall. For example, the environmental conditions data was raised couple of times. Also, the overall performance of the production lines was interesting for almost all respondents for instance. This is where the new systems, mainly the data historian, could help the department. And for the scope of this thesis, it is a critical finding.

A big challenge for the department and this thesis alike, is that the department is not operational yet. The production machines are still being commissioned and the focus on the department is to get everything running. There has not been a lot of time to focus on new systems and the capabilities they might bring. The personnel have not had the time to learn to use the current systems fluently enough. This has a negative impact on how they feel about the current systems, let alone new systems. However, there are distinctive needs and targets for development where new systems could bring more value and fulfill those needs.

## 5.4 Determining the needs

The intention of the SCADA system and data historian system, and other systems alike, is to improve the core department capabilities. Whether it is to improve efficiency or safety or make the work otherwise easier for the users, the computerized systems should play a crucial role in that. The respondents of the survey contemplated the current situation on the core manufacturing department. The department is already actively using the SCADA system and other systems are to be implemented. The respondents had insights about the targets of improvement for the

SCADA system, but they could also point out distinctive gaps that the department has where other systems could help. Based on the analysis of the survey answers, key findings are summarized into five topics that represent the biggest needs of the department. The five topics are the following:

1. **SCADA system usability and accessibility.**
2. **Definition of relevant data.**
3. **Performance of the production machines.**
4. **Historical data easily available.**
5. **Visualization of the historical data.**

The first topic is the usability and accessibility of the SCADA system. This topic was raised by people that mainly use the SCADA system. The SCADA system was thought to be counter intuitive and otherwise hard to use. Critical information is hard to find on the user interface. The SCADA system also does not provide all necessary information e.g., information about the environmental conditions. This ascertains the need for improvement of the SCADA system.

The second topic regards the definition of relevant data for the department. The respondents of the survey pointed out that because of the novelty of the department, not enough thought has been put to decide what is relevant data and what is not. For the first time for this manufacturing process, a huge amount of data is gathered. The decision must be made what kind of data is relevant and can support the decisions about the production processes.

The third topic is the performance of the production lines and their machines and systems. Information about the performance of the production machines and their processes is vital for the employees working on the core manufacturing department. The respondents of the survey thought that they are not getting enough information about the production processes. It is hard to improve the performance of the production machines and processes if the systems used do not provide that information.

The fourth topic is the availability of historical data. Almost all the respondents wanted to have historical data about the production processes. This information would be used to follow trends in the production. Noticing and troubleshooting problems would also become easier with historical data available. Anomalies in trends would reveal when and what kind of problems there have been in the production.

The fifth topic concerns the visualization of data. The SCADA system can visualize data to a certain extent, but the user interface and the usability has been stated to be far from idea. Moreover, the SCADA itself cannot visualize all the needed information, especially historical data. Along with visualizing the historical data, statistical process control data could also be visualized to make the everyday work easier for the employees, and thus improve operations performance.

# 6. USER REQUIREMENT EVALUATION

This thesis tries to determine roles for the SCADA and the data historian systems in the core department. This problem was decided to approach by evaluating the user requirement specifications of each system. In this chapter the user requirement specifications of both systems are thoroughly scrutinized. The goal is to find which system would fulfill the needs presented in the previous chapter by examining the user requirement specifications of each system. First an overview and the structure of each system's user requirement specification is introduced. All the requirements are examined and then a summary is made from the requirements to represent different categories in the user requirement specifications. The user requirements are compared with each other, and similarities and differences are analyzed. The findings from the requirement are then reflected upon the needs raised by the answers of the survey. The challenge is, that the requirements themselves may not provide enough information in the first place. The requirements may also not tell enough about the capabilities of a system. This challenge is taken into account while doing this research.

## 6.1 User requirement specification overview

The user requirement specifications are made according to the processes presented in chapter 2. To make sure that a computerized system complies with regulatory requirements and fits for its intended use, it must be validated. Important part of the validation process is drafting the requirements and creating the user requirement specification based on the requirements. The user organization of a system drafts a set of requirements that the system must fulfill. The supplier of the system must then accept those requirements and give a reason for not fulfilling a specific need. Later on, the system is verified against the specifications to ensure that the requirements are met. Not every user requirement specification (URS) is made the same way nor follow the same structure. Depending, whether the system is a machine, a process or a computerized system for example, the requirements might be totally different.

In this case, both user requirements are made for computerized systems and determined by the same company. However, there are major differences. One of the key differences is that the URS for the SCADA is made only for the target department. All the requirements are linked to the core manufacturing process. For the data historian, on the other hand, the requirements are defined to cover all departments without going to details of each department's processes. Therefore, the requirements for the SCADA system are more specific and detailed compared to the requirements of the data historian. The user requirement specifications are also done by different people at different times.

Both systems are computerized systems meaning they both require similar validation. Validation requires lots of documentation in different phases of the validation process. URS is one of the first documents made in the validation process. The purpose of the validation is to ensure the system's fitness for its intended use and its GxP compliance. Both the SCADA system and the data historian are in GxP environment and handle GxP relevant data and thus must comply to GxP regulations. Therefore, the requirements should ensure that the system takes patient safety, product quality and data integrity into account. The requirements should also reflect the business process, in this case the core manufacturing process. However, the URS for the data historian is made for the target company in general, and not for any specific department.

## 6.2 SCADA user requirement specification structure

The SCADA user requirement specification is a document drafted to define the requirements of a SCADA system for the core manufacturing department. It is a vast document following a general structure of an URS document in the target company. The URS is divided into different segments. First the scope and purpose of the system is defined, and the SCADA product is introduced. In this case the SCADA system is for the core manufacturing department covering all the machines and production lines. Following the scope and purpose are requirements from different aspects. These requirements include requirements about the functionality of the system for example. Functional requirements give an overview of features that are required from the system. In addition to functional requirements there are data, security and technical requirements besides many other requirements.

The URS should be as accurate and specific as possible. Based on the requirements, the supplier must deliver a system that fulfills those requirements. The more thought is put to the requirements the better the final product should satisfy the needs. The supplier must thoroughly go through the requirements and give a response to every single requirement whether the supplier can fulfill that requirement or not. If the supplier cannot fulfill certain requirements, it must comment and give reason why it cannot fulfill the requirement. Then the customer and supplier have to mutually agree how to mitigate the requirements that are not met. In the case of the SCADA URS of the target company the supplier had multiple requirements that supplier or the product could not fulfill.

However, this is not untypical and often the requirements can be mitigated or there is a workaround for the specific requirement.

The purpose of the SCADA system is defined in the overview of the URS. It is stated in the overview that the SCADA system is a control system of distributed computers, graphical user interfaces (GUI), peripheral devices such as PLCs and data communications all networked together. The SCADA system uses server-client model in which applications and databases are on the servers in server room and operation stations that work as clients are on the shop floor. The information is shared between the clients and servers in isolated production network. The server-client architecture provides high performance and reliability for operation and data storage. The overview also introduces the SCADA product and its addons and depicts the architecture of the system and its relations to other systems. Other technical documentation such as software and hardware design specifications go into further detail of the specific SCADA product.

## 6.3 Data historian user requirement specification structure

The user requirement specification is drafted the same way to the data historian as for the SCADA system. The user organization defines the requirements, and the supplier of the system will address those requirements. The supplier can again decide not to fulfill a specific requirement, but it must be given a valid reason. Compared to the SCADA URS, there are substantially less requirements that the supplier and its product cannot fulfill. The structure of the URS is similar to the URS of the SCADA system. However, the scope and purpose of the data historian and the emphasis on each section of the URS are different compared to the SCADA system.

Where the requirements for the SCADA system were defined specifically to the core manufacturing department, the requirements for the data historian are general and are not restricted to only one production department or use case. Therefore, the requirements are not as detailed as the requirements of the SCADA system. The structure of the URS follows the general target company guidelines, but much depends on who drafts the requirements and what type is the system the requirements are defined for. Compared to the SCADA system, the data historian URS is shorter, and the requirements are defined on more abstract level. Archiving and handling of data is vital to the data historian and thus, the emphasis in the requirements is on the data requirements, whereas on the SCADA the emphasis is on the functional requirements.

In the data historian URS overview, it is said that the purpose of the data historian system is to collect and archive data. Both GMP and non-GMP data are in the scope of the data historian. The data historian is also used to contextualize and visualize the gathered data. The data historian also unifies data handling under one system instead of many diverse systems. In the overview is also a depiction of the general functionality of the system. In the depiction the data historian system has many layers, all the way from a production machine to an end user using visualization tools of the system.

## 6.4 Critical requirements

The user requirement specifications of the SCADA and data historian systems are thoroughly examined in part of this thesis. The goal is to find the most relevant requirements from the specification to capture the best use cases. The challenge with finding the requirements is that the user requirement specifications are quite different with one another, although they follow the same document structure. It is good to acknowledge that the requirements alone will not reveal, which system is better for a specific need. However, to narrow the scope of this thesis it will focus solely on the user requirements. The source for the requirements is the user requirement specifications drafted by the target company, but the contents of the documents cannot be disclosed.

For the analysis of the requirements, six categories are chosen for the key requirements. The key requirements represent a summary of all the requirements in each category. The first category is functional requirements. Functional requirements define the basic functionalities wanted from the system. Second category is data requirements. Data requirements deal with handling the data of each system. The third category is access rights and security requirements. Access rights and security requirements determine who can use the system and what kind of security measures must be in place for the system. The fourth category is electronic records. Electronic records ensure that everything related to the manufacturing process can be traced. The fifth category is interface requirements. Interface in this context means interfaces with other systems and equipment. The sixth category is technical and performance requirements. This category represents the capabilities needed from the hardware and software components. The most relevant and critical points from all the requirements from the previously mentioned categories are tried to choose. The summarized requirements of each category are presented in Table 1.

| | The SCADA system | The data historian |
|---|---|---|
| **Functional requirements** | - Centralized system that supports multiple individual production equipment.<br>- Several batches can be prepared for equipment for production.<br>- Alarms, events and audit trail are collected centrally and can be viewed from any operation station.<br>- Produces OE data that can be used for calculations and reports. | - Collects, stores and organizes archived data from production machines and production systems.<br>- Functionality to notify users about certain events e.g., exceeded thresholds.<br>- Tools to visualize archived data graphically and tabularly.<br>- Supports time and object related data selection.<br>- Supports calculations for the stored and visualized data. |
| **Data requirements** | - Data is stored into relational databases without delay.<br>- Data is protected against unauthorized access and the data cannot be changed or deleted.<br>- Data will never get lost. | - Compression of production data.<br>- Data is protected against unauthorized access and the data cannot be changed or deleted.<br>- Data will never get lost. |
| **Access rights and security requirements** | - Multiple user roles with different permissions.<br>- Nine user roles that correspond the roles on the department.<br>- User roles can be modifiable afterwards.<br>- Anonymous usage not possible.<br>- Logins to the system must be done with unique personal user IDs. | - Different user roles with different permissions.<br>- User roles include read user, key user and administrator.<br>- Separate read and write permissions.<br>- Anonymous usage not possible.<br>- Concurrent logins of the same user impossible. |
| **Electronic records** | - Compliance to GAMP 5 and EU GMP guidance.<br>- For electronic records are considered database data, audit trails, configuration parameters, batch reports, OE data, recipes and events.<br>- Electronic signatures for reviewing batch and recipe creation, approving batch release and recipe approval and for master data changes. | - Compliance to GAMP 5 and EU GMP guidance.<br>- For electronic records are considered process data, batch data, audit trail data and configuration data.<br>- Records must be accurate and retrievable for the lifetime of the system.<br>- Audit trail cannot be modified or deleted. |

| | | |
|---|---|---|
| **Interface requirements** | - OPC UA readiness to connect to other systems.<br>- Interfaces for production equipment.<br>- Interfaces for the SCADA addon products. | - Read-only connectivity to production machines and systems.<br>- Ability to monitor the communication between the data historian and the interfaces.<br>- Support for OPC UA and ODBC connectivity. |
| **Technical and performance requirements** | - Virtualized environment.<br>- System should provide real-time information about the status of a machine.<br>- Redundancy for servers.<br>- Encrypted communication. | - Separate production, quality and development environments.<br>- Software components must not close end of life status.<br>- System must provide acceptable response times. |

***Table 1.*** *Critical requirements summarized from the user requirement specifications of the SCADA and data historian system.*

## 6.5 Requirement observations

Table 1 depicts requirements for the SCADA system and the data historian system. The requirements are tried to summarize from the six requirement categories for each system. The first observation from the requirements is their tendency to be very technical. The requirements are often technological aspects of the system and also highly specific. Therefore, summarizing the requirements was more difficult. Another key observation is the similarities in the semantics of the requirements. Even though the requirements are drafted from different perspective, the formatting of the requirements is almost identical. The SCADA requirements are drafted from the core manufacturing department's perspective, whereas the data historian requirements are drafted for the system to be used anywhere in any use case. Nevertheless, the individual requirements are similar. However, some differences and unique aspects can be found on the requirements.

Functional requirements is the category that tells most about the systems. Functional requirements should describe most about what the system is capable of doing, rather than telling about the details of the system itself. For instance, the SCADA system should be a centralized system that is able to support multiple production machines. The data historian, on the other hand, should be able to collect, store and organize data from production machines and systems. These requirements differ with one another and describe the basic function required of each system. The data historian's functional requirements give quite good picture about the functionalities of the data historian system. The emphasis on the requirements is on visualizing the stored data.

Unfortunately, the functional requirements of the SCADA system do not give as good overall picture of the functionalities. The requirements emphasize the centrality of the system and that it is possible to use the system from individual machines and their operating stations. The SCADA system shares the possibility to make calculations based on the stored data with the data historian system. Clearly, the overview in the SCADA URS gives a better picture what the system is intended for. The overview states that the SCADA system uses a distributed networked computers, communications and user interfaces for supervisory process management. Alas, the overview is not written in the form of requirements and thus does not require anything from the supplier of the system. The intention of this thesis is to focus on the requirements, and from the requirements deduct the feasibility of the systems against the department needs presented in chapter 5.4.

The functional requirements are the most fruitful category from this thesis' standpoint. However, even the functional requirements leave a lot to be desired. The functional requirements do not directly reveal the feasibility of the systems against the department needs. The data historian seems to somewhat answer the topics of availability of historical data and the visualization of it, as presented in chapter 5.4. Yet, the requirements themselves do not tell how they would fulfill that need. The same way, the SCADA requirements do not describe how they would improve the observability of the production performance. This is a clear indicator that examining the user requirements was a bad approach to the research problem.

In the next category, data requirements, the requirements are already more technical than the functional requirements. Although the data historian URS had more data requirements the summarized requirements for both systems are very similar. Most of the data requirements for the data historian were irrelevant for the scope of this thesis. The requirements for both systems do not disclose much about uses for the data. Instead, the requirements concentrate on what technical aspects the data must fulfill. Both systems have nearly identical requirements, making it challenging to draw any meaningful conclusions from them. Although, most the needs presented in chapter 5.4 concern data, this category will not provide any sufficient answers.

Access rights and security requirements is the next category in table 1. This category describes the types of users and their rights for the systems. User rights are important part of the systems since so many functions use the systems for different purposes. Not everyone can have the user rights to do production, and equally, very few should have the rights to edit data in databases. The SCADA system's user rights deal with different kind of roles for the production processes. As for the data historian, not so many user roles are needed. Most users of the data historian can manage with read only viewer rights. All the other summarized requirements are again technical in their nature. This category will not reveal much about the feasibility of the systems regarding the needs either.

Electronic records are essential in pharmaceutical industries due to the highly regulated nature of it. Every action must be traceable. This also applies to the computerized systems. That is why the compliance to regulations is the most important requirement for both systems. Both systems share almost the same definition for electronic records. The other requirements in this category deal with handling the electronic records, and yet again, do not disclose much about the systems' feasibility. However, one of the department needs was definition of relevant data, as presented in chapter 5.4. Therefore, the electronic records category sets the requirements for what data the systems must at least collect and store.

Interface requirements are mostly about connections to other equipment and systems. The SCADA system must be able to have interfaces within levels one and two on the ISA95 automation pyramid. This includes production equipment with PLCs and other systems. On the other hand, the data historian must be able to have interfaces with almost any level on the ISA95 model. Other difference with the systems is that the SCADA is used to control the equipment and systems that it is connected to, whereas the data historian reads and collects data from the access points. Interface requirements are crucial when dealing with system integrations. If the data historian is to be implemented on the core department, the communication between the SCADA and the data historian must be considered carefully. However, on the current needs, this category will not provide significant help.

The last category in table 1 is technical and performance requirements. In this category the technical aspects of the requirements outweigh the insightful qualities. Noteworthy for the data historian in this category is the requirement for test environments. Test environments help when something novel is tried. For instance, a new type of access point is used or a new production equipment or system in general. Using test environments will not interfere with the production either. When the data historian is implemented on the new department the test environments can be safely used before moving into production. For the SCADA system real-time information is required. Real-time information is crucial when controlling the production processes. Some processes may require user input on short notice.

## 6.6 Inferences

When looking solely at the user requirement specifications of both systems, the systems are very much alike. However, this is not the case. The similar structure and similar approach to define user requirements by the target company gives the impression that the systems are alike. The requirements have to also take the pharmaceutical regulations into consideration. This creates certain aspects for the requirements that the systems must always comply to, regardless of what kind of system is in question. The differences could have been easier to analyze if the requirements focused more on the use cases that the systems are intended for, instead of the technical and regulatory aspects.

In chapter 2.4 system user requirements were generally discussed. The chapter introduced a way to hierarchically divide the requirements. On the topmost level were the mission requirements that reflect the needs and objectives of the user in a bigger picture. Stakeholder requirements were on the following level. They were in the context of the mission requirements and set the performance parameters and constraints for the system. Considering the needs presented in chapter 5.4, the mission requirements would be the correct ones to address the needs the best. Functional requirements in the target company's specifications are perhaps closest to the mission requirements. However, even the functional requirements are too specific and technical. The genuine objectives and needs for the systems remain unclear.

One of the most notable, and surprising, observations for this thesis is that the user requirement specifications may not be adequate for deciding what systems to use. The user requirement specifications do not give enough information about all the needs intended for the system. Requirements for each system concentrate on the technical capabilities of the systems itself rather than on the use cases that are intended for the systems. This may also indicate that the requirements are drafted poorly. Especially the technological nature of the requirements steers away from the needs of the department. From the perspective of this thesis, it is a critical finding. Two conclusions of equal value can be drawn. Either the research method was poorly chosen or the requirement specifications themselves leave a lot to be desired. Be that as it may, the department needs cannot be addressed satisfyingly enough by examining the user requirements.

The main observations can be summarized into few themes. The first theme is the poorly chosen research method. The second theme is the quality of the user requirement specifications. The combination of these two themes caused the thesis to achieve its objectives only partially. However, when analyzing the needs presented in chapter 5.4 the roles can be somewhat determined. There are use cases for both systems. The capabilities of the SCADA and the data historian system are presented in chapters 4.2 and 4.3 respectively. Based on these capabilities combined with the user requirements, some conclusions can be made. The needs and their corresponding systems that address the needs are presented in table 2.

| Established need | System that addresses the need |
|---|---|
| SCADA system usability and accessibility. | The SCADA system |
| Definition of relevant data. | The SCADA system, the data historian |
| Performance of the production machines. | The SCADA system, the data historian |
| Historical data easily available. | The data historian |
| Visualization of the historical data. | The SCADA system, The data historian |

***Table 2.*** *Established needs and the systems that address the need.*

The needs presented in chapter 5.4 and table 2 can be reflected upon without taking the requirement specifications into consideration. Use cases can be found for both systems when examining the needs. The first need was the usability and accessibility of the SCADA system. It is obvious that this need can be addressed with the SCADA system. The SCADA system requires training for the users in order to accumulate experiences with the system. With training the use of the system becomes more fluent. The SCADA system can also be refined by developing the software itself based on feedback from the users. Altogether, this need is covered within the context of the SCADA system.

The next need was to define what is relevant data. Since the data historian is not yet implemented, this topic also touches the SCADA system. However, in the future the data historian could be of use when the uses for the data are considered. The SCADA system, along with other systems used in the production equipment, gather unprecedented amounts of data for the target department. This data needs to be organized and the best uses for the data be thought. The SCADA system's role is to gather production data and relay it to the operating stations. Alarms, events, and everything performance related data is critical. In addition to that, a lot of other measurement and sensor data is gathered as well. Such data may not be necessary for the daily operations but with the help of data analytics, it could be of use. All the data is gathered and stored. This data can be transmitted to other systems, such as the data historian. With the tools available in the data historian, the raw data can be refined with calculations into useful information. For example, the wear in components can be tracked from the data and thus, predict maintenances better.

Performance of the production machines was the next topic. Information regarding the performance of the machines is critical for the operating efficiency. This need is relevant for both systems, and both systems can have a role in this. The respondents of the survey felt that the accessibility of the SCADA system is poor, and relevant information is difficult to find. The SCADA system can be developed further, and the training can be increased, like stated previously. Moreover, the data historian can help track the production performance. The data historian can be used to visualize historical data, which will give valuable information about how the lines and machines have performed. The data historian can also show online SPC data and give an overview of the whole production line or one batch for instance. Thus, the capabilities of data historian can have a huge positive impact regarding this need.

Like previously stated, the data historian can be used to collect and visualize the historical data. This was also one of the needs raised based on the survey. Historical data contains information from the past. All unexpected events and faults can be traced from the historical data and thus, troubleshooting becomes easier. When following trends in the historical data, unexpected events and fault situations can be predicted better. Positive trends can be followed as well. If an exceptionally positive period can be found by looking at the historical data, the factors that accounted for the positive performance can be analyzed. Analyzing both positive and negative trends can help the department to continuously improve.

Visualization is strongly linked to the historical data. Visualization of the data was also seen as a big need. The current systems could not visualize information satisfyingly enough. Both the SCADA and the data historian system have capabilities for visualization, but the data historian is virtually designed for that. The visualization tools of the data historian give the users freedom to choose what kind of data they want to visualize, and what kind of calculations they want to make. For instance, an overview of the current situation of the department can be created, as depicted in figure 7.

The established needs and their corresponding systems that address the needs are presented in table 2. In conclusion it can be stated that both systems are needed and both systems have their unique roles. The observations made in this thesis highly emphasizes the need for the data historian, although it has not been implemented on the target department yet. Most of the needs require the contribution of both systems for the need to be addressed. The SCADA system makes the daily operations possible, and the department is not necessarily dependent of the data historian. However, the SCADA system alone will not satisfy all the needs and use cases. In an ideal situation the systems complement each other. In parallel use the systems can improve the overall performance of the target department. The benefits can directly increase the production efficiency and output of the production lines. Other benefits can be increased cost-efficiency achieved by reducing losses and predicting maintenance needs for example. These benefits are all part of industry 4.0, towards which this department is also proceeding. Altogether, the parallel use of the SCADA and the data historian system is beneficial for the core manufacturing department.

# 7. CONCLUSION

In this chapter the observations made in the previous chapters are discussed and concluded. The goals of this thesis are recapped along with the biggest challenges. Further development around the topics presented in this thesis are discussed in this chapter as well.

## 7.1 Goals and challenges

This thesis was done to a pharmaceutical and life science company that has a production facility in Finland. The target company is ramping up a new department where the core of various pharmaceutical products is to be made. The new department takes a huge technological leap compared to the previous similar department. With the new department comes new computerized systems. A new SCADA system is adopted, and a data historian system is planned to be implemented. The purpose of these systems, along with enabling fluent and efficient operation, is to make the new department data-driven and to increase the digital maturity.

In the first part of this thesis the reader was familiarized with pharmaceutical industry in general. Pharmaceutical industry has many unique aspects due to its highly regulated nature. The cornerstone of pharmaceutical industry is patient safety. To ensure patient safety, pharmaceutical companies must ensure that everything is done by following good manufacturing practices. The next part examined automation and data. In this part a general depiction of a SCADA system and a data historian system was given. Their place in the ISA95 automation pyramid was also discussed. The next part introduced the target department of this thesis and its processes and systems. The department's SCADA system was thoroughly scrutinized. A hypothetical architecture, of what the data historian could look like, was also given.

The goal of the first part of the research section of this thesis was to establish the needs of the target department. To figure out the needs, some of the target department employees were surveyed. All the respondents worked on different roles on the department. The survey consisted of five questions. Based on the answers, five key topics were raised to represent the biggest needs of the department. The needs are to be addressed by the new systems. To find out which of the systems would better fulfill the said needs, and how they would do it, the user requirements of both systems were investigated. First the specifications were introduced and examined. From the specifications six most relevant categories were chosen. Then, requirements from these categories were summarized. The summarized requirements were analyzed, and insightful answers were tried to find.

With any project, everything starts from planning. Whether it is a new production equipment or a new system, the same rules apply. Among the first things in planning is addressing the needs and objectives for the coming system, for instance. Requirements are conceived from the needs. The requirements should always reflect the business needs, and they set boundaries for the system's design. In this thesis the requirements are in key role. To help determine the roles for the systems was approached by examining user requirement specifications of each system.

The goal of this thesis was to determine whether the SCADA system or the data historian system would be more feasible for the needs set by the target department. Thus, the aim was to define the roles for each system. The research question about the system roles was difficult to answer. One reason for this is because the feasibility is hard to measure. There are no clear pre-existing indicators for measuring the feasibility. The needs established by surveying the target department acted as a framework for the system roles. How well the systems fulfilled those needs was not unambiguously answered. Hence, the objectives for this thesis were only partially achieved.

The roles could be partially determined by analyzing the capabilities of the systems and combining them with the user requirements. It was shown that there is a clear role for both system and both systems are needed. The data historian complements the SCADA system. Both systems have their own advantages. The observations made in the role analysis highly emphasizes the need for the data historian, although it has not been implemented on the core department yet. Their parallel use can bring considerable benefits for the core department. These benefits steer the department towards the themes present in industry 4.0, along with increasing the overall performance of the department.

Maybe the most beneficial outcome from this thesis, however, were the analysis of the survey answers. First of all, the survey answers established the department needs. In addition to that, the survey revealed, how the employees on the department view the systems and their strengths and targets for development. The attitude towards the systems and the increased availability of data varied significantly depending on what the role of the employee is on the department. Reluctance to change was noticeable among some of the respondents. These observations serve the target company particularly well. Based on these observations, however, from the perspective of this thesis, they could have been more easily measurable. Altogether, the survey was among the biggest accomplishments of this thesis.

## 7.2 Discussion and further development

The topic of digitalization and data-based value creation are key factors for industries to stay competitive. The bigger trend in top of all technological hype is industry 4.0. The themes in this thesis are in the framework of subjects regarding industry 4.0. In fact, the target department is in the middle of technological upheaval. Data-driven decision making, and technological capabilities are discussed frequently. Next big steps are already discussed. For instance, cloud technologies including cloud computing are thought. These technologies can bring huge competitive advantages. However, before all the benefits can be reaped, the fundamentals must be in place. When the fundamentals are refined, there is a lot of room for development regarding industry 4.0 for the target company. In the future, the core department can serve as a forerunner. Now that the department is close to having all the necessary technological means, the possibilities are endless.

The survey revealed that a lot of thought must be put to the key processes in the core department. The relevant data must be sorted out from the masses. Also, the current systems require training before their use is natural. Decisions must be made about how different systems are utilized. Until enough effort is put to these topics, all talk about cloud technologies etc. is irrelevant. Regardless, computerized systems are in the center of the new department and its production equipment. Additional systems themselves may not bring more value. Finding the best use cases, roles, for the systems is important instead. This is as important as it is topical for the target department. Hence, this thesis was conceived from that problem.

From an academic standpoint, the research problem in this thesis was challenging to approach. System user requirements were chosen to narrow the scope. Finding source material for system validation in pharmaceutical framework was not hard. System user requirements are a universal topic in any project. Abundance of source material could also be found about SCADA systems. On the other hand, finding literature about data historians turned out to be more challenging. Examples of such data historian systems that are used in the target company were hard to find. Furthermore, the stumbling stone of this thesis was analyzing and drawing conclusions about the user requirements.

Determining system roles can be an important issue for situations where multiple systems are being implemented, and the systems have overlapping capabilities. Every situation is different and there might not be a single comprehensive method for determining the roles. The user of the systems decides how the roles are defined. User requirements can be used to find out the roles for systems. However, if the needs and use cases are not clear from the beginning, the user requirements alone may not provide satisfying answers. A systematic method for defining system roles, in scenarios like this, could be a subject that is worth researched further.

# REFERENCES


[1] Mahdi Sharifzadeh, Industry 4.0 Vision for the Supply of Energy and Materials, Wiley Telecom, 2022, pp. 313-327

[2] European Medicines Agency, About us, Available at: https://www.ema.europa.eu/en/about-us (Accessed May 2023)

[3] Finnish Medicines Agency, About us, Available at: https://www.fimea.fi/web/en/about_us (Accessed May 2023)

[4] Food and Drug Administration, About FDA, Available at: https://www.fda.gov/about-fda (Accessed May 2023)

[5] European Medicines Agency, Good Manufacturing Practice, Available at: https://www.ema.europa.eu/en/human-regulatory/research-development/compliance/good-manufacturing-practice (Accessed May 2023)

[6] Good Manufacturing Practices and Distribution Practices, Commission Directive 2003/94/EC, Available at: https://health.ec.europa.eu/medicinal-products/good-manufacturing-and-distribution-practices_en (Accessed May 2023)

[7] Good Manufacturing Practices and Distribution Practices, Eudralex Volume 4, Available at: https://health.ec.europa.eu/medicinal-products/good-manufacturing-and-distribution-practices_en (Accessed May 2023)

[8] International Society of Automation, ISA95, Enterprise-Control System Integration, Available at: https://www.isa.org/standards-and-publications/isa-standards/isa-standards-committees/isa95 (Accessed May 2023)

[9] P. Lezotre, International Cooperation, Convergence and Harmonization of Pharmaceutical Regulations, Elsevier, 2014, pp. 1-2

[10] Y. Al-Worafi, Drug Safety in Developing Countries, Elsevier, 2020, pp. 22

[11] The International Society for Pharmaceutical Engineering (ISPE), What is GMP?, Available at: https://ispe.org/initiatives/regulatory-resources/gmp/what-is-gmp (Accessed June 2023)

[12] GAMP5 A Risk-Based Approach to Compliant GxP Computerized Systems Second Edition, International Society for Pharmaceutical Engineering, 2022

[13] D. Buede, The Engineering Design of Systems, John Wiley & Sons, New Jersey, 2009

[14] K. Stouffer, S. Lightman, V. Pilliteri, M. Abrams, A. Hahn, Guide to Industrial Control Systems (ICS) Security, National Institute of Standards and Technology (NIST), 2015, 247 p.

[15] International Society of Automation, Intech, Beyond the Pyramid: Using ISA95 for Industry 4.0 and Smart Manufacturing, Available at: https://www.isa.org/intech-home/2021/october-2021/features/beyond-the-pyramid-using-isa95-for-industry-4-0-an (Accessed August 2023)

[16] G. Yadav, K. Paul, Architecture and security of SCADA systems: A review, Elsevier, 2021, 26 p.

[17] H. Eren, Assessing the Health of Sensors Using Data Historians, Institute of Electrical and Electronics Engineers (IEEE), 2012, 4 p.

[18] General Electric Digital Solutions, Historian, Historian overview, Available at: https://www.ge.com/digital/documentation/historian/version90/c_hgs_historian_overview.html (Accessed September 2023)

[19] Rockwell Automation, FactoryTalk Historian SE Installation and Configuration Guide, Available at: https://literature.rockwellautomation.com/idc/groups/literature/documents/in/hse-in025_-en-e.pdf (Accessed September 2023)

[20] J. Collin, A. Saarelainen, Teollinen internet, Talentum, Helsinki, 2016, 333 p.

[21] The Rules Governing Medicinal Products in the European Union Volume 4 EU Guidelines for Good Manufacturing Practice for Medicinal Products for Human and Veterinary Use, "Manufacture of Sterile Medicinal Products Annex 1", Available at: https://health.ec.europa.eu/system/files/2022-08/20220825_gmp-an1_en_0.pdf (Accessed August 2023)

[22] B. Galloway, G. Hancke, Introduction to Industrial Control Networks, Institute of Electrical and Electronics Engineers (IEEE), 2013, pp. 863-865

[23] SAP, Products, What is a manufacturing execution system (MES), Available at: https://www.sap.com/products/scm/execution-mes/what-is-mes.html (Accessed September 2023)