The Impact of Unified Namespace in Industry 4.0

Ádám Péter

Samuel Werner

LUND
UNIVERSITY

Department of Automatic Control

MSc Thesis
TFRT-6254
ISSN 0280-5316

Department of Automatic Control
Lund University
Box 118
SE-221 00 LUND
Sweden

Printed in Sweden by Tryckeriet i E-huset
Lund 2024

Abstract

Digitalization is part of the ongoing transformation for industries that are going through the fourth industrial revolution, Industry 4.0. The Fourth Industrial Revolution introduces significant advancements in digital technology, fundamentally transforming manufacturing environments through interconnected and intelligent systems. Organizations' willingness to adopt and integrate various AI solutions and other digital technologies presents a challenge for integrating, using, and storing data across various operational and information technology systems. Traditional data management concepts and solutions are not built for these new technologies and therefore, rarely able to fully capitalize on the potential of data.

This master's thesis investigates how Unified Namespace (UNS) can serve as a central hub for data communication, thereby addressing the challenges organizations face due to fragmented data systems, while supporting the implementation of advanced analytics and artificial intelligence. This report focuses on implementation of Unified Namespace systems, while identifying potential benefits, drawbacks, and barriers with the technology. It also considers alternative methods and compares them to Unified Namespace architectures.

The study adopts a comprehensive methodology that includes case studies of manufacturing companies, interviews with industry leaders, and simulations. These elements are complemented by an extensive review of existing literature, providing a deep dive into the theoretical foundations of digitalization in industry, UNS and its practical implications.

Key findings in this report include Unified Namespace significantly streamlining data management by acting as a centralized repository, enabling

real-time data sharing at all levels of the organization. Integrating Unified Namespace enables significant time savings, improving operational efficiency, while facilitating higher accuracy decision-making, and at the same time reduces complexity of data systems. The main barriers presented are the initial complexity of transitioning, the upfront investment and potential resistance to change within the organization.

In conclusion, Unified Namespace, despite its challenges, prevails as a concept that can become critical in the transition to Industry 4.0. Not only does it enable seamless communication but is a necessary step for enabling multiple AI solutions and more. This report concludes that gradually transforming communication protocols to Message Queueing Telemetry Transport (MQTT), using the Schultz method, can be a way to decrease initial investment costs, while transitioning to the more effective protocol. Another insight is the importance of using a platform that is able to communicate with all of the business-critical systems.

This thesis provides a foundation for future research to explore Unified Namespace and its long-term impacts on the manufacturing industry. It also serves as a practical guide for industrial entities going through a digital transformation and wanting to implement data management systems.

Acknowledgements

This master's thesis was written during the spring semester of 2024. The master's thesis marks the completion of the author's studies. We, the authors Ádám Péter and Samuel Werner, have studied MSc in Mechanical Engineering, specializing in Product realization. The thesis is written with a pervading collaboration between the authors from beginning to end. Hence, both have contributed equally to all sections in the report.

Firstly, we would like to direct a warm thank you to our supervisors at Novotek, Nils Lindqvist and Jens Molin, for giving us the opportunity to write our master's thesis in collaboration with you. Your knowledge and experiences guided us towards an extremely interesting subject and kept it interesting along the way. Thank you for providing us with insights in the beginning of the work and deepening our understanding through interesting discussions. We would also like to direct a special thanks to everybody working at Novotek for making us feel part of the company and taking the time to help us when we needed.

Secondly, we would like to thank Professor Charlotta Johnsson, our supervisor at Lund University, Faculty of Engineering, for the outstanding support throughout the study. Your experience with working master's thesis projects helped us with valuable guidance and feedback throughout the project. We would also like to thank you for the constant positivity that made our master's thesis a joy to write.

Thirdly, we express our gratitude towards our examinator Dr. Ulf Jeppsson, for enhancing our work with your insightful feedback, constructive criticism and valuable suggestions.

Furthermore, we would like to thank the interviewees who have shared their thoughts and experiences with us during the interviews in the study. We appreciate you taking your time and are grateful for your contributions. Special thanks to SKF and Jan Ek for letting us visit their headquarters.

Lastly, we would like to thank Rasmus Kalvenes and Oskar Joelsson for taking their time and performing an opposition to our work. Thank you for your perceptive comments.

Authors

Ádám Péter
Samuel Werner

Charlotta Johnsson, Professor in Automatic Control & Dean at Campus Helsingborg, Lund University. Chair for ISO TC184/SC5.
Ulf Jeppsson, Senior lecturer and Head of division of Industrial Electrical Engineering and Automation at the Faculty of Engineering (LTH), Lund University.

Nils Lindqvist, Digital Innovation Manager at Novotek Sverige AB.
Jens Molin, Managing Director / Country Manager Sweden at Novotek Sverige AB.

Key Words:
Unified Namespace, Industry 4.0, Industrial Internet of Things, Industrial Data Architecture

Abbreviations

AI - Artificial Intelligence
API - Application Programming Interface
AWS - Amazon Web Services
CPS - Cyber Physical System
DCS - Distributed Control System
ERP - Enterprise Resource Planning
GA - Generic Algorithm
HMI - Human Machine Interface
IIoT - Industrial Internet of Things
IP - Internet Protocol
ISA-95 - International Society of Automation standard for developing an automated interface between enterprise and control systems
IoT - Internet of Things
IT - Information Technology
KPI - Key Performance Indicator
LLM – Large Language Model
MES - Manufacturing Execution System
ML – Machine Learning
MQTT - Message Queuing Telemetry Transport
OEE - Overall Equipment Effectiveness
OPC-UA - Open Platform Communications – Unified Architecture
OT - Operational Technology
PLC - Programmable Logic Controller
R&D - Research and Development
RBE - Report by Exception
RL - Reinforcement Learning
SCADA - Supervisory Control and Data Acquisition
SOA - Service Oriented Architecture
SQL – Structured Query Language
TCP - Transmission Control Protocol
TLS - Transport Layer Security
UDTs - User-Defined Types
UNS - Unified Namespace
UOM - Unit of Measurement

Contents

1 Introduction

Since the invention of the steam engine kickstarted the First Industrial Revolution, the ever-evolving landscape of the manufacturing industry has been characterized by relatively slow changing trends, with major paradigm shifts in each subsequent industrial revolution. The most recent shift toward digital integration in manufacturing known as Industry 4.0, is characterized by smart machinery, interconnected systems, and advanced data analytics, forcing industrial businesses to adapt, or be left behind. (IBM, 2024)

1.1 Background

The Third Industrial Revolution, also known as the Digital Revolution, defines the shift from mechanical and analog electronic machines to digitally programmable machines. In the latter half of the 20th century, the perfection of mass production allowed humanity to experience an exponential growth in global living standards and a population boom. This Digital Revolution spawned fields focusing on digital logic, elevating the importance of microprocessors, metal-oxide-semiconductor field-effect transistors (MOSFETs) and integrated circuit chips. Factories became increasingly digitalized with the introduction of programmable logic controllers (PLCs), allowing for limited automation and the collection of data (IBM, 2024). The management of these, the discipline of operational technology (OT) has been rapidly changing how people and companies operate and interact with their surroundings, eliminating the need for the most repetitive tasks to be performed by humans, while aiding engineers and senior management in streamlining production lines.

The Fourth Industrial Revolution, also referred to as Industry 4.0, (IBM, 2024) is the ongoing digital transformation of industry through the integration of smart machines, leveraging the Internet of Things (IoT) and big data to drive further automation and efficiency gains. In the manufacturing industry, a substantial transformation is occurring through the rapid adoption of digitalization. Companies are not merely using digital technologies to support existing operations but are fundamentally transforming their business models by integrating data-driven intelligence into their operations. Despite these advances, a critical challenge persists; to fully capitalize on the potential of data, it must be integrated from disparate business units into a unified system, such as a data warehouse or data lake. This centralization is essential for harnessing data to fuel innovation and strategic decision-making within the organization. (Manditereza, 2023)

This strategy is especially problematic in manufacturing due to the wide variety of Operational Technology (OT) and Information Technology (IT) systems with different meanings and definitions of data. That is, as more data is collected and stored centrally, the inconsistencies make it increasingly difficult to make sense of the data, which is detrimental to advanced analytics that organizations need for intelligent decision-making. (Manditereza, HiveMQ, 2023) Using shop floor data outside the factory walls is becoming ever more common, with the line between OT and IT made blurry from a data analytics standpoint. Integrating these traditionally incompatible systems presents its challenges; leading to inefficiencies, data redundancies and missed opportunities for optimization.

The concept of Unified Namespace addresses these challenges by providing a framework for integrating diverse data into a united and interoperable environment. Acting as a central hub and structural framework, UNS allows real-time data access across the organization, from shop floor to senior management. Facilitating enhanced decision-making, while reducing time spent on searching and analyzing data, implementing UNS aids in improving overall enterprise efficiency. This thesis work explores the theoretical and practical aspects of UNS, seeking to understand its impact on modern industrial practices and its potential to allow businesses to become more data-driven.

1.2 Purpose and Research Questions

The primary objective of this master's thesis is to provide insight into Unified Namespace and establish a foundation for implementing a Unified Namespace for industrial companies.

The goals are to provide manufacturing companies with an introduction to Unified Namespace and a foundation for building their own. It aims to provide industrial companies with a direction for what a Unified Namespace can be, while trying to find a common ground and definition for what it is. Moreover, it will also attempt to determine what value an implemented Unified Namespace will deliver to industrial companies and why it may be beneficial to implement. The thesis will also try to find inefficiencies in the current infrastructure and compare the concept with existing different solutions, to enhance results.

The authors hope that the thesis will contribute to academia by increasing the knowledge about digitalization, especially with regards taken to Industry 4.0 and Unified Namespace, while identifying knowledge gaps associated with Unified Namespace and data management of industrial companies, driving new research areas. Moreover, the authors also hope that the thesis will spread the concept of Unified Namespace and establish characterizations of Unified Namespace, while simultaneously giving rise to new ideas and concepts to handle data transformation. Furthermore, the thesis could provide insights into the barriers and opportunities of digitalization for manufacturing companies, enabling applications of artificial intelligence and advanced digital systems, while avoiding some common pitfalls.

Based on the background and the stated purpose of the thesis it is structured around four research questions, presented in Table 1 below

Table 1. The research questions of the thesis.

RQ1	Why is Unified Namespace relevant for manufacturing companies and what problems does it solve?
RQ2	Are there any barriers, challenges, or drawbacks to implementing and using Unified Namespace?
RQ3	How can/should a Unified Namespace be implemented in a manufacturing company?
RQ4	How does Unified Namespace differ from existing infrastructure?

1.3 Delimitations

Focus areas of this report are to investigate implementation of UNS and how it affects a business and its operations. The report will not implement a UNS in an organization and will not build multiple UNS's to exemplify different implementations of it. This is mainly due to other focus areas and cost associated with acquiring software. Simulations will not use enough parameters and nodes to be able to establish if one single communication protocol is desirable from a latency standpoint, as it is not the focus area of the report. The purpose of the simulations is to establish if UNS is functional and if its structure can benefit business operations and data management.

Furthermore, because of the research method, the researchers limited the number of interviews to five and the manufacturing companies were all located in Sweden. All concepts presented by companies and literature could have been delved into, but restriction was necessary as there is a time limit and lack of relevance for the report's objectives. However, drawbacks and benefits of the different methods are presented and discussed in the report.

1.4 Structure of the Report

The structure of the report is presented in Table 2 below:

Table 2. The structure of the report.

Chapter	Description
1 Introduction	The background and the purpose of this master's thesis project are presented, alongside delimitations of the report.
2 Method	The research method of the thesis is described in detail, which guides how the data is collected to answer research questions. Furthermore, the research process is presented, and credibility of the study is discussed.
3 Theory	In this chapter, the findings from the literature review are presented in several topics. These topics generate a comprehensive understanding of data management, DataOps, Industry 4.0, industrial data systems and Unified Namespace.
4 Research Results	The fourth chapter presents results from interviews with Novotek and external companies, along with learnings. Firstly, learnings of UNS are presented, based on interviews with Novotek and literature. Secondly, the simulation with alternative methods for building a UNS, and lastly interviews with companies are presented.
5 Discussions	In this chapter, an analysis and discussion of the results gathered in the fourth chapter are conducted. They offer some observations from interviews and simulations, while discussing some limitations of the research.
6 Conclusion	Lastly, the conclusion relates to the research questions and summarizes the actual result. It also suggests further research areas in the context of UNS.

2. Methodology

2.1 Research Method and Strategy

Most research methods follow well-designed procedures and are considered either basic, applied, or developmental. Basic research is conducted to test or arrive at a theory, meaning the main objective is to establish general principles without any practical applications to its findings. When theory is applied to test, evaluate, and solve problems, it is referred to as applied research. Developmental research is, not entirely unlike applied research, a systematic work that develops learnings from applied research, to produce additional knowledge, and directing or improving existing products or processes. (Foegeding, 2020)

Research methods are generally either quantitative, qualitative, or mixed research design. Quantitative research typically tries to measure variables numerically, with standardized instruments to establish relationships among variables. The goal of it is to confirm assumptions, confirm theory, test hypotheses, and determine cause and effect relationships. The process starts with deciding one or more hypotheses, which will be tested after data is collected. Data is collected from experiments, close-ended interview questions, and countable and numbered observations. Furthermore, the data collected is independent and collected from a sample that makes it as generic and objective as possible. On the contrary, qualitative research emphasizes words and meanings rather than numerical analysis and collection of data. It develops interpretive narratives from data to attempt to capture complexity and examine poorly understood issues. Qualitative research also tends to be somewhat subjective and include open-ended questions, observations of behavior and literal reviews with similar research questions. The goal of this type of research is to identify common themes or ideas, understand language used and interpret texts. For both quantitative and qualitative research there are several different methods. (Zou & Xu, 2023) (Elsevier, 2024) Lastly, mixed research design can be seen as a mix of the two perspectives, where researchers see the previous frameworks as complements rather than rivals. (Zou & Xu, 2023)

It is essential to decide what research method to use in the report, since it will determine many, if not all, of the data collection methods. It will mainly decide what type of data to collect, how it is collected and how it should be analyzed. (Elsevier, 2024) This report will be structured according to the qualitative method and can be considered applied research. This decision is due to the research questions formulated in chapter 1.2 and the lack of prior research. This report aims to find patterns to UNS, how it should be implemented and establish what alternatives that exist. These goals are indicative of a qualitative research method. Within the subgroup of qualitative research methods, case study is chosen, as it best correlates with the deep knowledge needed and the multiple types of data sources used, mentioned in chapter 2.2. A case study is typically used to relate value to an organization, entity or method, which is the main research question in this report. Yin (2009) suggests, in his book, that if research questions are phrased as how and why, a case study is among the top choices for research methods. He also adds that what-questions are applicable as long as they are exploratory standpoints, meaning the primary objective is to discover, understand and generate ideas. Data collection methods typically used for a case study are interviews, written documentation and observations. As this research aims to establish a foundation for Unified Namespace, a case study is particularly fitting. (Yin, 2009)

2.1.1 Case Companies

In order to comprehensively assess the applicability and impact of Unified Namespace (UNS) across different industrial contexts, this study incorporates a multi-case approach. This approach allows for exploring nuances of UNS implementation in diverse operational environments and identifies common themes, while it answers research questions with enhanced reliability. (Yin, 2009)

Including multiple companies in the case gives the report greater accuracy, while capturing a broad spectrum of industrial digitalization, activities and maturity. Companies were selected based of the following criteria and can be seen in Table 3:

- Companies are manufacturing companies in different sectors.
- Companies are at a mature stage.
- Companies should currently be working with digitalization.

Table 3. Data about the case companies.

Company Name	**HQ, location**	**Number of Employees**	**Revenue 2023 [Billon EUR]**	**Number of production sites**	**Locations in number of countries**
Alfa Laval	Lund	20 300	5.54	>40	160
SKF	Gothenburg	42 602	9.12	108	130
Stena Recycling	Gothenburg	3 500	1.40	178*	8
International Packaging business**	-	-	-	-	-

** Stena recycling locations in Europe*

*** As the company wishes to remain anonymous, data is left out*

Out of these companies, half were customers of Novotek. The International Packaging business, and the interviewee, wishes to remain anonymous throughout this report and therefore information that may be revealing for either is withheld. This can be seen in Table 3.

To further enhance the result, software companies are included in the report as they have a different perspective on UNS than manufacturers. The companies can be seen in Table 4.

Table 4. Data about the case software company.

Company Name	**HQ, location**	**Number of Employees**	**Revenue 2023 [Billon EUR]**	**Launched in number of countries**
HighByte Inc.	USA, Maine	33	0.0038	18

The companies are further introduced and discussed in chapter 4.

2.2 Research Process

Through collaboration and interviews with Novotek AB about industrial automation and Industry 4.0, the topic of Unified Namespace was presented. Research into the subject revealed limited prior research and thus the research questions were formulated to establish a basis for further research, while identifying common themes and ideas. Through literature reviews and interviews it was possible to build up an extensive knowledge base into the subject matter and formulate interview questions. This knowledge, coupled with support from Novotek, allowed for simulations on a UNS structure, with reflection and analysis of the data. From multiple interviews and simulations, it was possible to draw conclusions and make suggestions for further research for UNS. The research process can be seen in Figure 1, that illustrates it with a flowchart.

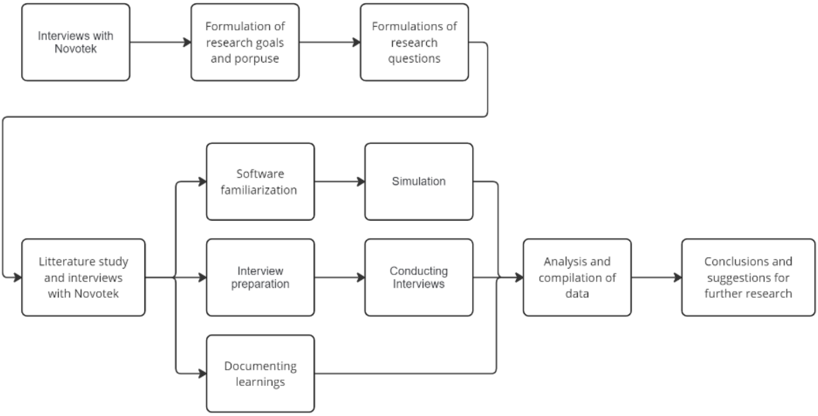


Figure 1. Flowchart for the research process.

2.3 Data Collection

2.3.1 Interviews

Interviews were conducted to firstly get an understanding, and to collect information and data of relevant processes and activities. The interviews were initially constructed with open-ended questions, with the aim of increasing understanding and initiating the development of research questions. With increasing knowledge, the questions became more close-ended, i.e. more specific, which provided an extensive understanding of the subject. This stems from there being three structures of interviews, structured, semi-structured and unstructured. A structured interview involves a step-by-step process in which the researcher asks predetermined questions. This interview process is also known as the standardized interview and has the aim of collecting information and data relating to the relationship and concepts of different concepts. Unlike a structured interview, the unstructured interview does not rely on predetermined questions, but instead relies on spontaneity. The interviewer would generally use open-ended questions, that are in line with the subject and context, to guide the interview. This type of interview is more conversational and to find out detailed information, follow-up questions must be asked, with the aim of finding out about a person’s individual experiences and knowledge. Lastly, the semi-structured interview is a combination of structured and unstructured. (Höst, Regnell, & Runeson, 2006)

This report uses all types of interviews. The initial interviews used an unstructured format where it was important to develop an understanding of the subject matter. These interviews were conducted in Novotek’s offices. As knowledge grew, questions became more specific and resembled the semi-structured interview framework, however, there was little to no preparation behind them. Interviews with companies were also semi-structured as it was important to prepare with some questions, but certain questions could not be formulated before the encounter with the company. The interview with John Harrington from HighByte was structured so it was possible to formulate questions before the interview.

2.2.2 Written Documentation and Material

Literature reviews are conducted to provide deepening of knowledge and to gain new or larger perspectives. They are a crucial component of a thesis as they initially can be used to provide understanding and a general picture of the subject. As the work progresses it is used to deepen understanding and find gaps in existing frameworks and make sure the work is original. (Höst, Regnell, & Runeson, 2006)

The literature review was performed in Google Scholar, Lund University Library, Wiley and more. Non-bias material such as scientific reports, books and some journals were priorities. However, to accumulate extensive information on the relatively new subject it was necessary to use white pages and articles written by companies and some articles from relevant people. These companies include HiveMQ, Novotek, companies that were interviewed, and a public Figure within the community called Walker Reynolds.

2.3.3 Simulations

A broad definition of simulations is to manipulate data generation processes or a data analysis to determine how a certain element of the analysis is affected. Simulations are used with increasing interest among researchers and companies alike and are used in various fields, among which manufacturing systems are one of the most widely used. System in a simulation describes a collection of entities that act together towards a goal or logical end and can be performed on a physical system or a model of a system and can have an either analytical solution or just be a simulation. Meaning, if the system is too complex it may not be possible to find a numerical solution. Furthermore, a model can be either static or dynamic, deterministic, or stochastic, continuous or discrete. (Law, 2015)

In this report the simulation is performed to gather information on how an implementation could be performed and to find potential advantages or disadvantages with UNS, the results of which are too complex to describe as numerical values. Therefore, the simulation is a typical simulation performed in a virtual space, i.e. a model. The model itself is dynamic, deterministic and continuous. This is due to the time being important in the system and the products mattering for the actions performed by the machine. The system is also too complex to find numerical values for, and the purpose is merely to examine if the concept is a viable option.

2.4 Credibility of Study

Ensuring credibility and unbiased opinions is of imperative importance and as there is a collaboration between Novotek and the writers, some techniques and methods have been applied to ensure credibility.

2.4.1 Triangulation

Triangulation involves using multiple data sources, theories and methods to verify data and conclusions. This is performed to strengthen validity of the case study, as multiple sources are used. (Yin, 2009) In this thesis it is firstly achieved by comparing multiple sources of documented literature and discussing findings with the supervisor at LTH and the local office of Novotek. The comparison of sources provided areas where literature differ from each other, while discussions with supervisors and Novotek employees provided feedback and insights, developing new approach angles. In addition to this the multi-case approach provides its own triangulation since it provides a platform for triangulation of Novotek's answered questions. Multiple interviews provide a triangulation, in and of itself, as inconsistencies are noticeable, and discussions can be had if they are sector related or company specific.

2.4.2 Reliability

Reliability, in the context of verifying reliable data sources, is about fact-checking sources and making sure that they are trustworthy and credible. Reliability is crucial in a report and ensures that a report can be consistently reproduced and, in the end, come to the same conclusions under the same conditions. To ensure that a report is reliable means that the authors are critical when data is collected, evaluating bias, sources' purpose, and the context of the information. Typically, sources published in an academic environment, and a relevant field, are of high reliability. Otherwise, sources with transparency, as in usage of citation, providing sufficient detail and so on, are preferred and triangulated/cross-checked. (Höst, Regnell, & Runeson, 2006)

In this report, reliability has been a focus area within credibility, as UNS is a new field within data management which leads to few academic sources being available, since little previous research has been done. Consequently, some sources in this report are from companies producing software intended for UNS. These sources have undergone scrutiny, have been cross-checked with other sources, and have also been discussed with experts from Novotek. The authors acknowledge that Novotek may have biases towards UNS, and therefore, facts accumulated from Novotek have also been cross-checked and triangulated.

3. Theory

3.1 Layers of Traditional Automation and ISA-95

Traditionally, since the birth of the Digital Revolution, also called the Third Industrial Revolution, the flow of data throughout industrial companies has been similar. It has relied on one-to-one connections, characterized by layers, where layers mostly communicate with the layer above or below, see Figure 2. This leveled approach is mentioned in ISA-95, which describes the standard for implementing an automated interface between control systems and enterprises. The main objectives of ISA-95 are to standardize consistent terminology and information models to clarify communications between industrial players and create a functional hierarchy defining process in a manufacturing organization. (Brandl & Johnsson, 2021) Here is a brief overview of how the integration of automation, and the traditional flow of data within an industrial company can generally be described.

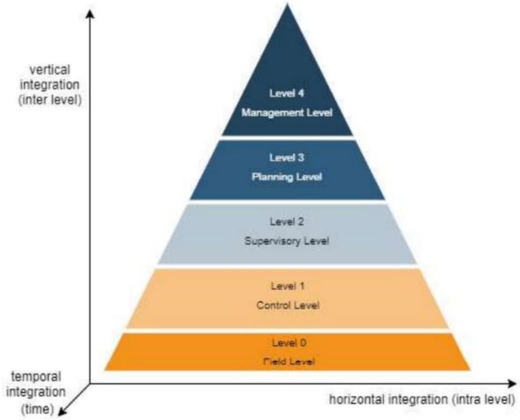


Figure 2. The automation pyramid. (David, et al., 2019)

According to a joint study from Tampere University (TUNI) and the Norwegian University of Science and Technology (NTNU), integration in automation can be described along three dimensions, namely horizontal, vertical, and temporal or longitudinal directions. Vertical integration takes place between different levels (inter-level), whereas horizontal integration refers to that within each level, also called intra-level. An interesting observation of the vertical and horizontal integration is that they can symbolize data becoming more aggregated further up in the pyramid. Temporal integration, however, takes the time variable into account, in other words takes place along the lifecycle of the plant. (David, et al., 2019)

Level 0 – *Field level*, or the shop floor equipment. At the bottom of the flow of data is the information from the production floor. This layer consists of a variety of sensors and devices, such as measuring instruments and actuators. Essentially, anything that measures variables connected to the production floor can be sorted into this layer.

Level 1 – *Control level*. This layer is responsible for the control and manipulation of production processes. Programmable logic controllers, or PLCs, are traditionally used to automate production processes. In some more modern systems, in situations that require control of more complex processes, PLCs can be replaced by a distributed control system, or DCS. Variables from the layer below, from the shop floor equipment are fed into the control system, that then becomes the brain of said equipment. For instance, if the automation engineer wishes to keep a certain temperature in a process, there will most likely be a temperature sensor sending information to the control level, that will decide whether to activate the actuator for the heater or AC.

Level 2 – *Supervisory level.* Supervisory Control and Data Acquisition (SCADA) systems. These systems function as a connection between the equipment that needs to be controlled (sensors, actuators etc.) and the monitoring interface where data can be analyzed and be used for reports. (Scada International, 2024) SCADA systems are able to communicate with PLCs, distributed control systems (DCSs), and Human Machine Interfaces (HMIs) and act as a bridge between these. Human operators usually access the system through the HMI, as it includes a graphical user interface that is hopefully more intuitive to work with. SCADA is what allows companies to manage their on-site processes remotely, as it allows them to access data without being physically on-site.

Level 3 – *The planning level*. This layer is responsible for monitoring the entire manufacturing process, from raw materials to finished goods. To achieve this, a manufacturing execution system (MES) will be implemented. This is software that will use a combination of data from layers below and human operator input in order to track variables such as inventory, as well as machine and waste metrics. Data to monitor manufacturing operations and production activities will be available in the MES.

Level 4 – *The management decision making level*. This is where the company decision makers access information and statistics to create long-term strategies and control the entire company. Together with data from the MES, enterprise resource planning (ERP) system creates operational clarity that allows stakeholders to make informed decisions. The ERP focuses on things such as material use, plant schedules, coordination of orders and shipping.

Commonly, each level in an industrial system uses domain specific software, designed by different software and hardware suppliers, each employing different *communication protocols*, which can be likened to programming languages. This naturally creates barriers for dataflow and inhibits data exchange. Traditionally organizations used point-to-point integration, meaning that level 1 only communicated with levels 2 and 0. Consequently, if the ERP system requires data from the control level, it must go through all the intermediate levels, slowing down the process of data retrieval. The data must go through, sometimes multiple, language transformations to finally reach ERP, a process that consumes significant bandwidth, meaning the rate data transfer occurs at a part of a network. Along with the data flow process, variables can also get lost, as one selects a handful of them to save bandwidth. Additionally, every level requires some dedicated engineering to facilitate data exchange and translate the data to the correct language, referred to as communication protocol in industrial context. Dedicated solutions can pose a potential risk since they can become dependent on one or a few people familiar with all the connections. The company would in a worst-case scenario have to remake the program as the people responsible leave the company, which could become expensive, while it also makes future integrations, updates or other improvements more difficult. It is important to clarify that this does not always have to be the case, as it is possible to connect a SCADA system to an ERP for instance. However, as they naturally use different timescales (milliseconds, seconds for SCADA or days, weeks for ERP), there is a compatibility issue between them. (Manditereza, 2022)

As displayed in Figure 2, the traditional architecture allows for tri-lateral integration. This thesis focuses primarily on vertical integration with regards taken to the life cycle of an industrial plant (the temporal integration), focusing on the data flow between each automation level through time. However, temporal integration will also be an important factor in finding sustainable industrial automation solutions in a rapidly changing environment. A study for the International Association for Research in Income and Wealth (Erumban, 2018) estimated that industrial machinery lifetime, on average, varies from 26 years to 34 years. Meaning, machinery getting commissioned today is expected to function in around three decades, but also that machines from decades ago are still frequently used. Meanwhile, the trend for the average cost for one sensor used in Internet of Things (IoT) applications, described in section 3.2.1, has been on a steady decline, making them more accessible for companies. This lowers the entry costs for companies tracking their production even more thoroughly. Numbers show that these sensors have on average decreased from 1.3 U.S. dollars to 0.38 U.S. dollars, nominally, between 2004 and 2020 (Statista, 2024). At the same time, companies have been pouring enormous resources into Internet of Things (IoT) implementation and research to stay on top of their competition. This is to say that newer machinery, already built with IoT applications in mind can have a difficult time communicating with old school, simpler machines with less sensors and trackable variables. If this trend continues with the advancement of manufacturing technology, companies ought to form their long-term strategies and investments with future IoT compatibility in mind. Bringing the topic back to the importance of temporal integration within automation, the data architecture allowing for IoT necessitates development that allows for communication between today still yet used machinery that was commissioned decades ago, with machinery that is about to get commissioned today. This, while at the same time leaving space for the future, improved machinery that might function in a way that is not even yet invented, to connect effectively to the same Internet of Things.

ISA-95 aims to address this by establishing a standard of data transfer, and to some extent, establish different name groups at different levels. Through ISA-95, it becomes easier to migrate data between different levels and through standards it also becomes easier to maintain, upgrade and solutions become less personalized. However, it fails to establish standards for individual machines, which means that the different integration points, even though standardized in a broad manner, vary. One example of this is if machine variables are called something different in the MES than in the SCADA system, which would not necessarily cause issues for a small company working with limited machinery. At scale it can become a huge barrier for further development, as this replicates for many or all machines. Furthermore, one large and modern machine could have several hundred trackable attributes, or tags, that without proper structure, within the organization, all collected data could become unusable. (Wally, Huemer & Mazak, 2017) (Lindqvist, 2024)

3.1.1 Industry 4.0

Since the First Industrial Revolution, technology has been a driver for change within the industrial complex, and sometimes it makes leaps that define a new revolution. The First Industrial Revolution occurred as the industry mechanicalized, the second as electrical energy was implemented in production and the third when industries digitalized. (Lasi, et al., 2014) During the 21st century, humanity entered a new era of industrial revolution, Industry 4.0, or the Fourth Revolution, which is accelerating exponentially. Drivers for this revolution are, like previous industrial revolutions, technology and more specifically Internet, and “smart” technologies. Yet again industries are being fundamentally reshaped, this time through digitalization of the entire industry and consumer markets, impacting not only the production but the entire value chain. (Ghobakhloo, 2020) While Industry 4.0 refers to a wide range of concepts, the fundamentals are as follows:

- *Smart factories:* The term was first used in 2010 by German legislators and is central to Industry 4.0. Smart factories have multiple definitions, but commonly contain definitions and references to it being a Cyber-Physical System (CPS) that integrates the network/Internet for processing and analyzing external and internal data, to achieve an autonomous or flexible production. However, the main goal of the smart factory is to understand complex processes of manufacturing and increase systematic processes. It aims to vertically integrate both physical assets and IT-systems, to achieve flexible production and autonomous/intelligent production. A smart factory typically has four intelligent features: (Shi, et al., 2020)
    1. Smart factories have sensors that are aware of their surroundings and allow machines to self-adjust, learn and

    maintain their environment. Decisions are made based on changes in the environment.

2. A smart factory also has interoperability and real time control of the Internet. Interoperability is an integrated system where machines from different manufactures that use different communication protocols are able to communicate with each other. Real time control allows for adjustments and changes to happen instantly and data collection and analysis too. These metrics are controlled over the Internet and typically use the Internet of Things (IoT).
3. Smart factories are also highly integrated with robot vision and systems Artificial intelligence (AI). AI enables the smart factory to analyze and make decisions and can in theory be in control over the entire production process. AI will lead to, among other things, increased efficiency, enhanced predictive maintenance and greater flexibility. It can enable analysis through data from sensors but needs robot vision, imaging hardware and software to coordinate in a flexible environment.
4. Lastly, smart factories process Virtual Reality (VR), which can, among other things, facilitate the integration between machine and human. (Shi, et al., 2020) (Lasi, et al., 2014)

- *Cyber-Physical System, CPS:* By using modern sensors, network and technologies it is possible to efficiently and accurately integrate physical assets in a cyber environment, to the point that they merge. From an automation perspective CPS is about specialized systems where the computing and communications are based on physical objects and structures. CPS is, as previously mentioned, connected to smart factories and has a physical and a virtual layer. Technologies tied to CPS within Industry 4.0 include Internet-of-Things, Big Data and other smart technologies. (Alguliyev, Imamverdiyev & Sukhostat, 2018) (Lasi, et al., 2014)

- *Self-organization:* Machines being able to adapt to various situations is central to Industry 4.0 and can be seen as a level of autonomy for machines. Smart machines are able to make decisions by themselves by having access to data, thus decision making in Industry 4.0 aims to become decentralized. This will also provide an added layer of flexibility and the line or machine becomes more reconfigurable than previously. (Wang, et al., 2016) (Lasi, et al., 2014)
- *Corporate social responsibility, CSR:* A key objective of Industry 4.0 is CSR and the advancement in technology creates many opportunities for improvement in, for example work environment and potential reduction in pollution, among other things. A key difference between Industry 4.0 and previous industries is that factories should be designed to follow human needs, rather than the other way around. The added level of flexibility also contributes to the idea of a decrease in the amount of equipment production needs and factories could potentially become smaller as a result. A direct consequence of this would be increased resource efficiency and decreased energy usage, as factories become more specialized. (Lasi, et al., 2014)

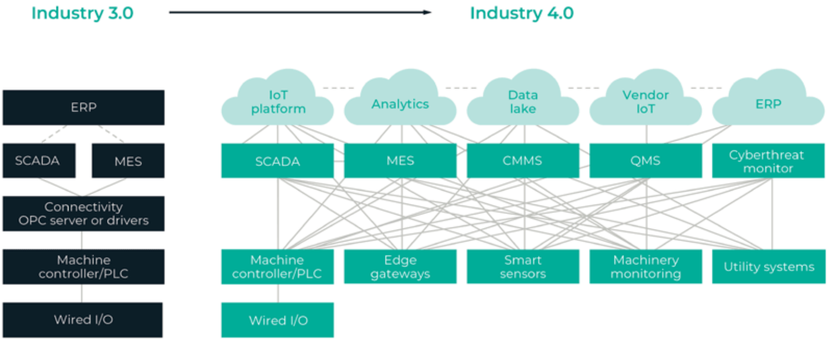


Figure 3: Differences between connections of physical and virtual architecture in Industry 3 and Industry 4.0. (Harrington, 2021)

In Industry 4.0, the data architecture for industrial data is predicted to become more complex, presented in Figure 3. Connections between edge devices and applications multiply, and are less likely to follow the strict layered approach. Interconnectedness in the smart factory demands more connections, adding to its architectural complexity. During and as a consequence of the industrial transition, the number of sensors and amount data from the production exponentially increases. Traditionally, all data travels through all levels of the organization limiting scalability, increasing latency, while complicating finding and analysis of relevant data. To exemplify, in case of two separate systems of a higher-level request the same data from the PLC, the one-to-one data architecture will poll twice from the PLC, one time for each query. In combination with the previously mentioned issues, this will use a lot of bandwidth. Increased complexity of the data architecture is also a contributor for increased demand for complex maintenance and development solutions. (Manditereza, 2022)

To summarize, Industry 4.0 enables more resource efficient practices throughout the value chain. It describes many, mainly IT driven, systems that will have a tremendous positive impact on flexibility for factories and production, which can be seen as one of the goals of Industry 4.0. (Lasi, Fettke, Kemper, Feld, & Hoffmann, 2014) The revolution is driven by data use cases, enhancing the need for connecting the operational technologies (OT) with the different IT-systems that an organization uses, while supporting real-time communications, i.e. changes in operations, production and more. Through tremendous data collection, the goal is essentially to integrate the production with communications systems that an organization uses, allowing enterprises to become data-driven. Integration of these systems will provide a platform for decentralized development, higher utilization, and autonomous systems, but can be complex to obtain as connections become increasingly complex. Data is becoming the main parameter of a producing organization, the backbone of Industry 4.0. and therefore, vital to collect and utilize. Industry 4.0 will enable continuous production, lower costs, faster launch to markets, more overall sustainability and much more. (Shi, et al., 2020)

3.1.2 Industrial Artificial Intelligence

Artificial Intelligence (AI) and business intelligence are becoming an ever-increasingly important cornerstone for the industrial complex. AI can be defined as "*The intelligence depicted by machines based on mathematical algorithms and statistical analysis from data.*" (Peres, et al, 2020) (Zohaib, et al., 2023) AI is often wrongly used synonymously with the term machine learning, but includes a wider set of specifications like sensing and reasoning. Machine learning is a subfield within AI and is a set of methods that can detect patterns from old data and use these learnings to predict future outcomes.

Industrial Artificial Intelligence, in the context of industrial applications specifically, should not be confused with the widely adopted Large Language Models, or LLMs, like ChatGPT or Gemini, as the fundamental functions of these can differ greatly. Industrial AI encompasses machine learning, robotics and other fields to mainly develop industry, optimize flow, as well as planning and diagnostics, and is therefore seen as a crucial element of Industry 4.0. Industrial AI can be categorized in six levels, where 0 represents the lack of AI and 5 is a fully autonomous system: (Peres, et al, 2020) (Zohaib, et al., 2023)

Level 0: Has no autonomy and a human has full control, which can be exemplified by a system where a robot does predefined, hard coded tasks.

Level 1: The first level of autonomy is similar to Level 0 but a machine can make suggestions on goal-oriented improvements. However, the operator still has full control over all decisions, so this level is categorized as an assisted level of autonomy.

Level 2: At this level the human still has predominant level of control and responsibility, but the machine can make improvements on its own within predefined boundaries. This level is called partial autonomy.

Level 3: Is known as the delimited level. In this level the AI system has control over larger boundaries, i.e. it is still partially programmed by humans, can adjust behaviors, make - and implement plans. The AI system warns when problems occur, but still need human validation for the solution of the problem. It is now essential to equip the robots with sensors and it can collaborate with entities in the environment.

Level 4: The system is now adaptable and functions autonomously within even larger areas of system boundaries. Self-optimization is enabled, and the humans have now taken on a supervision role, where intervention only is necessary in emergencies. The system is even able to solve some emergencies by itself, if necessary.

Level 5: The final level of Industrial AI is categorized as full autonomy and through collaboration with other systems it can adapt to disturbances or fluctuations. However, the system boundaries are still defined by humans but is rather a limitation than a necessity. In case of an emergency the system is able to deal with it and put itself in a secure mode, which boils down to humans not being a part of the system. (Peres, et al, 2020)

Industrial AI and applications such as ChatGPT by OpenAI, based on LLMs, both fall under Artificial Intelligence yet differ significantly in use case and underlying mechanisms. Industrial AI models usually require specialized machine learning algorithms, due to the different equipment, processes, and products in a business and between industries. In order to tune models, one needs an immense amount of data, that for use cases, such as text generation (essay writing, summaries, answering e-mails), coding or translation, is widely available through the Internet. This is a key reason for why LLMs have demonstrated remarkable capabilities in natural language processing tasks, but struggle in domain-specific tasks, which is what Industrial AI is trademarked as. The availability of industrial data is very limited, as businesses generally are secretive with their processes and historically lack the infrastructure for useful data-handling and tracking, even for their own operations. (Zohaib, et al., 2023) Incorporating another company's data would not necessarily benefit the algorithm as it could generate unreliable results, as optimization algorithms are usually tailored to that specific company. Industrial AI development is further obstructed due to company secrecy, as organizations tend to keep their competitive advantages a trade secret, which will lead to slower development of Industrial AIs overall.

A parallel between "classic" LLMs and Industrial AIs is the need for good quality, and large quantities of input data in order to work optimally, although Industrial AIs will need to perform with comparatively less input data. Industrial data needs to be consistent, clear and have context, as one cannot expect useful output without persistent input. Furthermore, pre-trained LLMs still have a lot of imperfections and deficiencies and cannot be trusted completely to perform in important areas. Developers are still trying to solve problems like hallucinations, value non-alignments, the weak ability to specialize, and the black box effect (Chen, et al., 2024). Industrial AI applications based on LLM are different compared to traditional deployment and development use cases, and as a consequence of its uncertainty and instability, as of now, is simply not reliable enough to be trusted with industrial processes.

Despite high expectations and the industries pursuit of AI enabled solutions, it has yet to be widely adopted in industries. (Peres, et al., 2020) Regardless, Industrial AI has shaped some industry sectors with the emergence of predictive maintenance, supply chain optimalization, and more recently digital twins, and is prone to gain even more traction. (Zohaib, et al., 2023) For machine learning, the best-case scenario is to give the Industrial AI access to all available data, but this data rarely exists contextualized. Often, when data is collected in Industry 3 and current Industry 4.0 solutions there are predefined metrics that are prioritized for human consumption and calculations. During the transfer between different execution layers, important parameters get picked out by that specific program or the engineer, excluding a large amount of data that could be useful further upstream. This filtration used to be necessary to simplify and optimize data streams, given the large amount of data new systems and sensors produce. Today, it is more important than ever to maximize the value from the available data. AI requires all available data to make decisions, as the parameters used to make those decisions may not be obvious to humans. One of the most time-consuming activities is to decide what data to collect. However, AI can assist with that process today. It can uncover connections within datasets that an engineer might dismiss. Furthermore, it is important for the data to have context, i.e. have some sort of structure to optimize output of AI models, for it to reach reasonable conclusions. (Zohaib, et al., 2023)

There are a couple of possible solutions that may be adopted according to studies in the field, however it is important to emphasize the lack of a one-fit-suits-all solution. Reinforcement Learning and Generic Algorithms can automate feature engineering and aid in hyperparameter optimization, as these functions focus on creating models that find the best path to explain the correlation between independent sets of data (Zohaib, et al., 2023). In the same study, Automated Machine Learning and Neural Architecture Search are proposed ways to dynamically pick out the best ML model from a dataset, while emphasizing the need for IoT-based OT monitoring systems to assist the training of industrial ML models. Various types of Industrial AIs are gaining ground, but there is no way to crown a winner just yet. Most of the models are still evolving, and picking out the best one will depend on a variety of factors, including the type of industry, available data, acceptable model bias, as well as the ultimate business case and value creation of the implementation.

3.2 Industrial DataOps in Industry 4.0

DataOps, a term often used referring to data operations, is a data management practice for building, deploying, and managing data applications the same way as for software products. (Harrington, 2021) Industrial DataOps is a branch of this, focusing on software solutions for industrial companies as they transform, adopting Industry 4.0 in their manufacturing. With the ability to closely monitor every part of a company’s workflow, comes the ever higher need to make sense of the collected data in order to avoid drowning in unusable information, which is one issue companies can suddenly find themselves in. Industrial DataOps is a term describing solutions that contextualize and standardize the dataflow. It is what helps the business decision makers and engineers understand all the information.

Data in industrial environments can prove to be inconsistent across different machinery, lacking context such as units of measurement and what the expected tolerances are (Harrington, 2021). This can be manageable when operating a small number of machines, however on a larger scale it can become a real issue. A maintenance engineer reading a number dubbed “K2:19” with a value of 63 perhaps might understand the context and if the value is acceptable or not, but long-term the lack of appropriate nomenclature can rapidly get out of hand.

The way OT (Operational Technology) and IT (Information Technology) systems communicate to one another can be natively different. OT, or industrial networks, mostly communicate through protocols such as, OPC-UA, MQTT or Modbus, depending on the age, model and function of the device. These include operational hardware and software, such as PLCs, SCADA and DCS systems. IT systems on the other hand usually communicate through APIs (Application Programming Interface), or other bespoke integration protocols, traditionally. Edge devices refer to the devices at the endpoint of a data network, in an industrial application these could be the sensors or actuators in a production line. There exists a process of convergence between OT and IT, with common practices and communication technology available for both. Industrial DataOps solutions must therefore integrate between devices from the OT layers conforming to industry standards, and simultaneously to the most modern business applications on the IT side. (Harrington, 2021)

In today's era of big data, vast datasets are generated daily from sources such as industrial processes, cloud applications, and the digitization of historical records. However, the costs associated with preserving such large-scale data center infrastructure for data storage are prohibitive. Surveys indicate that only about ten percent of big data research organizations can sustain their data generation activities financially, leaving them to drown in unusable data. (Chen, 2018) A significant issue within the data science community is the lack of efficient and effective data access solutions that facilitate the sharing, analyzing, and interpreting of big data resources. Data analysis is only as beneficial as the quality of input data, enhancing the importance of structurization, contextualization and accessibility of generated data.

3.2.1 The (Industrial) Internet of Things

The term IoT was first used by Kevin Ashton in 1999, which described a system in which the physical world could be connected to the Internet by sensors. The basic idea behind it is turning common objects into connected devices allowing them to be controlled remotely through the Internet. While his description relied on radio frequency identification, it today uses mainly Internet Protocols (IP) to communicate with sensor enabled devices. In industries, the Internet was used for wireless technology before Ashton mentioned it, through different machine-to-machine industrial solutions. However, these early solutions were often ad hoc solutions, i.e. purpose-built and closed network solutions. (Rose, Eldridge, & Chapin, 2015). The key difference between the IoT and the Industrial Internet of Things (IIoT) is that the IoT can be classified as machine-to-user interaction, while IIoT is about connected industrial assets to the information systems. The smart factory is, through the large amount of data collected, supposed to respond to changes quickly and dynamically, thus affecting the entire industrial value chain. (Sisinni, et al., 2018)

While the basic idea of IoT and IIoT is the same, using low-cost and resource-constrained devices that allow for use of the Internet ecosystem and network scalability, communication requirements vary. An example of IoT is a smart home, where devices like lights, ovens, and window blinds are connected and can be controlled remotely. This system's purpose is to service a human, has ad-hoc connections, and does not require very high accuracy nor security. IIoT, on the other hand, is found in manufacturing plants, made up by shop floor equipment, with high data volumes and strict cyber security. IoT focuses more on user friendliness and flexibility of devices being able to connect. Contrary, current IIoT emphasizes on possible integration and interconnection of production plants, lines and machines. IIoT generally has a higher demand for reliability, connectivity and data volume transferred, especially for control applications because of this. Naturally, because of trade secrets and general cyber security, the need for IIoT to be secure is central. While the two concepts are closely related, there exist major differences between IoT and IIoT, as presented in Table 5**.** (Sisinni, et al., 2018)

Table 5: Differences of IoT and IIoT

	IoT	**IIoT**
Service model	Human-oriented	Machine-oriented
Connectivity	Ad-Hoc	Structured
Data Volume	Depends on application	Always high
Data Accuracy	Not stringent	Critical
Security	Medium to high	Critical

The networks and messaging system that industries use vary, with a wide range of available options described below. The hierarchy and some configurations of different systems can be seen in Figure 4, describing the layers needed for communication. In the background, to make communication possible there needs to be a network layer, physically transporting signals to a certain receiver. All IIoT solutions use some hardware to connect to the Internet, usually an ethernet cable, RS-485, RS-232, CAN or 5G (a wireless solution) for the physical layer. IP defines the rules that allow certain information to arrive at the correct location and is sometimes referred to as the backbone of the Internet. It is on top of the network different transport protocols can exchange messages, which is the mechanism for the communication itself, usually called communication protocols. For these exist specific frameworks, defining the data load so they can be easily parsed by the receiving application. On top of it all are the applications, which make use of the data, store it, analyze it, and present it to operators to make use of. (Sisinni, et al., 2018)

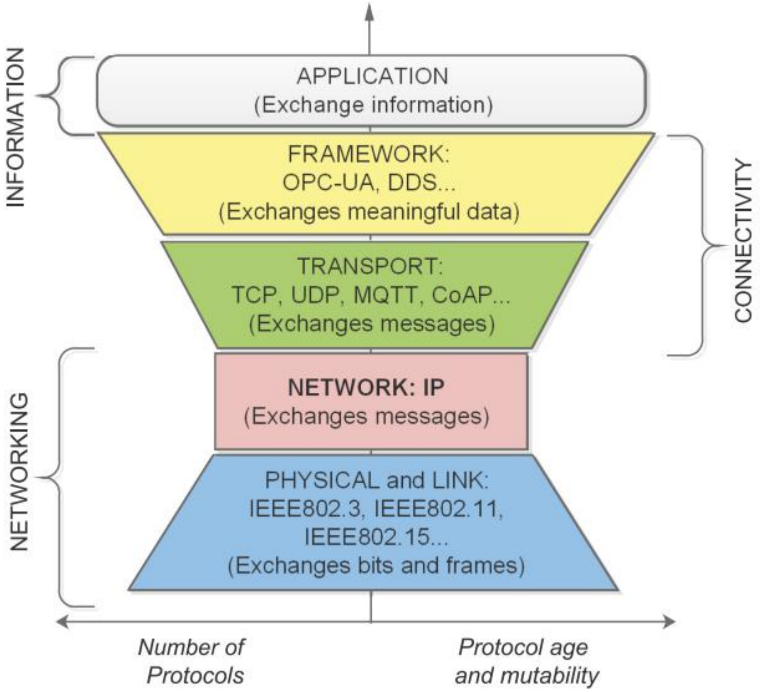


Figure 4: A typical representation of the virtual space of an organization. (Sisinni, et al., 2018).

3.2.2 Open Platform Communications

The OPC standard, developed and maintained by the OPC Foundation, is a series of specifications that define exchange of data in the industrial automation space, among other industries (OPC Foundation, 2024). It describes the interface between Clients and Servers, or Servers and Servers, to monitor events and alarms in real-time, but also access historical data and other applications. It was first released to the public in 1996 with the purpose of reshaping PLC protocols, such as Modbus, into a standardized interface converting generic-OPC read/write messages to device-specific request and vice-versa. This would allow SCADA and HMI systems to easily communicate with the PLC protocols, via the OPC standard. The standard got a widespread adoption across industries such as manufacturing, energy and building automation, to name a few.

Pre-OPC, for the communication between a PLC and an HMI application, a specific driver was used as the “middleman” for communication. More often than not, PLCs from different brands could be used on the same factory floor, and in this case, different PLC drivers had to be used for each model. For instance, as shown in Figure 5, in case a PLC of brand A and another one from brand B needed to communicate with an HMI system, they both would require their own driver. (Yngvesson, 2024)

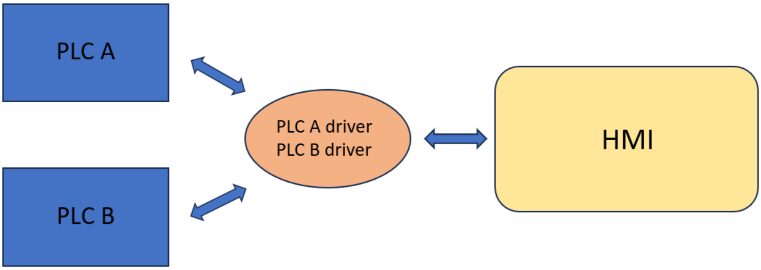


Figure 5: Communication between two PLCs and an HMI before OPC was implemented.

With the implementation of the OPC standard, the PLCs of different brands could share the same driver, easing compatibility issues. This means that businesses could now more easily connect PLCs developed by different brands and required fewer ad-hoc solutions just to translate messages, to name a few. An illustration of the communication using OPC can be seen in Figure 6. However, a requirement for this is for both the OPC server and the OPC Client to be based on the Windows operating system, which still limited the ecosystem to some extent. (OPC Foundation, 2024)

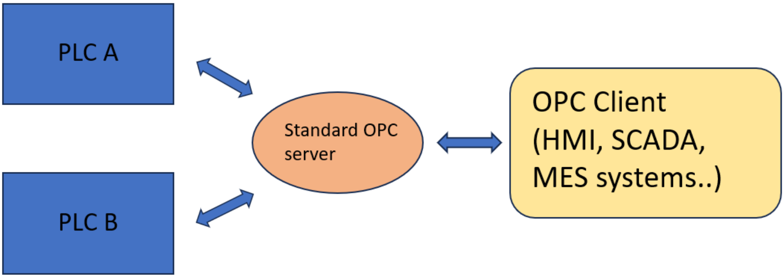


Figure 6: Communication between two PLCs and an OPC Client after OPC was implemented.

With the evolution of manufacturing architectures, security and extensive data modeling became ever more important. To tackle this, the OPC Foundation released Open Platform Communication – Unified Architecture, or OPC-UA for short, to provide a more scalable and extensible open-platform architecture. OPC-UA was designed to be platform independent, meaning it could no longer only run on Windows, but also other operating systems, such as Linux or Android, and on different hardware platforms from embedded micro-controllers to cloud applications. The update would also allow for more complex information models, sending details and properties of devices along with the values, giving a context to the transferred data. This updated framework also included better message encryption along with new functions, such as allowing devices to read and write data based on access-permissions. Moreover, the framework supports both client-server and publish-subscribe data transfer mechanisms, giving users flexibility. Client-server communication follows a so-called service-oriented architecture model (SOA), where a service provider receives a request, processes it and sends back a response. Publish-subscribe mechanisms are an alternative to this, as it is more scalable for networks with a large number of devices (OPC Foundation, 2024). In this case, devices publish data through a broker that other subscribers can access and can publish back, much like how some of today’s social media applications work.

In an OPC-UA network, the files that can be transferred must completely correspond to the ISA-95 standard, and today, OPC-UA is commonly used in various IIoT applications.

3.2.3 The Future De Facto Standard in Publish/Subscribe type IIoT Messaging

There are several methods to implement communication between devices that make up the Internet of Things. Message Queueing Telemetry Transport, abbreviated MQTT, is however becoming the de facto standard used for the system architecture in most publish/subscribe based IIoT applications. Today, it is uncommon to be used in higher-level. It is a lightweight publish/subscribe messaging protocol, ideal for connecting remote devices using minimal network bandwidth and basic messages to optimize for a small footprint. (MQTT, 2024) It was initially implemented in the oil industry to remotely monitor oil pipelines but has since spread to other industries, such as manufacturing, automotive and telecommunications. Besides MQTT, OPC-UA is also often used in publish/subscribe IIoT applications.

A MQTT network consists of one or several clients and at least one broker. Clients are allowed to publish and subscribe to brokers and communicate bi-directionally, through the broadcast of small messages. Clients can be everything from mobile devices, backend systems to small microcontrollers, and the network can therefore be scaled to millions of devices fairly easily. A representation of how it functions can be seen in Figure 7, where two clients are subscribed to the same sensor through one broker. (MQTT, 2024)

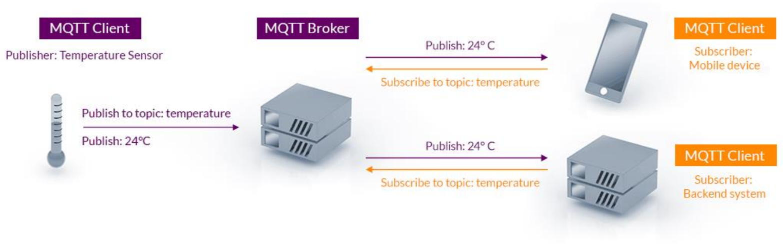


Figure 7: A representation of MQTT communication from a sensor to two clients. (MQTT, 2024)

Shop floor equipment, or IT type applications that can publish data to topics. These topics get amplified using an MQTT broker and made available to other devices, or nodes, via the broker. Subscribers to these topics, referred to as listener nodes, can in turn access this data and process it. For instance, messages published by a CNC-machine regarding different values and events get sent to a broker, through which a plant historian stores and archives the information. Through the same broker, an MES application can at the same time track things such as machine status, production rate and tool use, meant to be used for predictive maintenance, production planning and general analytics. The publisher node can therefore send all of its available data through the same broker (or several brokers), with the upside of letting the subscriber applications decide what data they will be subscribing to. All of this sets the ground for a truly data-driven manufacturing system, easing access to, as well as finding and using data. Using the MQTT broker as middleware allows for the centralization of all data in an efficient manner, meaning applications no longer have to poll for each value that they are interested in, where many times the needed data point travels through several layers and applications before arriving. This reduces bandwidth and allows for less points of possible failure. (Inductive automation, 2024)

Another reason for the popularity of MQTT in IoT applications is the reliability it offers. One typical problem is that networks, not connected by physical wires, can face is the lack of persistency in the connection, for instance when a car drives through a tunnel or in case weather conditions interfere with the satellite connection. In practice, it is possible to connect IoT devices over cellular network as well, however these tend to be less consistent and reliable. MQTT uses a message delivery protocol, called service levels, which essentially allows the broadcast to continue until a confirmation for the arrival of the message is registered. There are three service levels declared in the protocol, which can be defined in the message body: *0 - at most once, 1 – at least once and 2 – exactly once*. This ensures that a message is received when connection is reestablished, else will be sent continuously until confirmation of receival gets registered. (MQTT, 2024)

It is also possible to encrypt messages through devices, using Transport Layer Security (TLS), and end-to-end cryptographic protocol widely used in communications through the Internet. It is used to encrypt the transmission, so hackers and eavesdroppers cannot access sensitive information. (Internet Society, 2024)

MQTT can be extended by a specification that defines how data is sent and received, called Sparkplug B. The specification was designed by the Eclipse foundation, for clients at the edge of a network to communicate with applications, such as historians, analytics programs, and SCADA systems. Sparkplug B is an open-source protocol, and it essentially defines how the *topic* and *payload* of messages must be structured and sent to an MQTT Sparkplug enabled broker, ensuring compatibility for the over-time changing machinery at a plant. (Koprov, Ramachandran, Lee, Cohen, & Starly, 2022) Some important aspects of the Sparkplug B architecture are:

Publish/Subscribe: The protocol uses a publish/subscribe approach naturally, instead of a poll/response method like most common communication protocols.
Standardized payload definition: Data messages are standardized and can be decoded the same way by all network participants.
Automatic discovery: Devices can automatically sense what data will be sent by other participants in the Sparkplug network, both on- and offline devices.

Report by Exception (RBE): Sparkplug saves network bandwidth by only reporting device status that has changed, avoiding the same messages being broadcast unnecessarily.
Death-Birth Certificates: Birth certificates define what a device can and will send. Death certificates encapsulate information sent to the network in case a device (perhaps suddenly) disconnects, using Last Will and Testament messages.
Resolute connections: All nodes in the network are on by default using persistent Transmission Control Protocol, TCP connections. (Rout, 2024)

In its messages, Sparkplug uses a structure of topics for the message payload. This is what allows an easy-to-follow structure when publishing messages for listener nodes to subscribe to. A generic topic can be constructed in the following manner:

Namespace/group_id/message_type/edge_node_id/[device_id]

In this generic payload, a namespace can be "*spAv1.0*" or "*spBv1.0*" depending on the implemented specification, as there exists a Sparkplug A, which was designed first and got improved upon, birthing Sparkplug B. For this reason, if someone today is talking about Sparkplug, they usually refer to the Sparkplug B protocol as a consequence of its greater capabilities and higher adoption level. It is also possible to transfer data using pure MQTT, depending on the use case. *Group_id* might be a string, and *message_type* is what decides how the message payload will be handled, for example a birth, a death or a critical application state message. Edge nodes are the devices at one *edge* of the network, essentially the nodes that do not act as a middleman. These might be devices that convert proprietary or analog signals from old, legacy machinery that are not Sparkplug enabled as well. Otherwise, Sparkplug enabled devices can also publish directly to the *device_id* topic, as there is no need for a translator in between the data types. (Koprov, Ramachandran, Lee, Cohen, & Starly, 2022)

One downside of Sparkplug B is the in-built topic element limit of 3, meaning there is a limit on the hierarchy tree's branches. This becomes a problem when implementing an ISA-95 hierarchy, as it sets a limit on the depth the tree can have. For instance, using Sparkplug B it would only be possible to do *enterprise*, *site* and *area* levels. This hinders a complete UNS implementation using only Sparkplug B and no other protocol, suggesting it should be used as a complement rather than the go-to protocol. (Yngvesson, 2024) (Larochelle, 2022)

Another communication protocol (transport layer of Figure 4) that is commonly used in industrial devices is Modbus, released by Modicon in 1979. Modbus is widely used through several industries to transmit signals from devices and control instruments back to the main controller, for example a pressure monitor that sends data to a computer. There exist several versions of Modbus that evolved for different use cases, however it is most often used for connecting a supervisory computer with a remote terminal unit in SCADA systems. (Schneider Electric, 2024) One large reason for its popularity is its openness, meaning manufacturers can incorporate it in their systems without paying royalties to the mother company.

To summarize, MQTT is a natural publish/subscribe, lightweight and diverse communication protocol. Sparkplug B is based on MQTT but is instead a specification with added properties, that somewhat can influence its use cases. It is possible to run pure MQTT, or Sparkplug, but there needs to be a middleware between the two if they are run in the same ecosystem. Both are used in IIoT applications, where numerous devices generate large amounts of data and are dependent on scalability.

3.3 The Unified Namespace

The term Unified namespace was popularized by Walker Reynolds, that created the first UNS project in 2005, and who is an avid supporter and spokesperson for the technology. The exact definition of UNS is still not completely agreed upon, but UNS is a system or methodology where all entities, such as files, data or resources etc. are organized in a centralized repository of structured data. It describes the structure, events and other data points of a manufacturing company in a centralized way to form a single source of truth, where all network stakeholders are able to easily find the data they need, giving a snapshot of the organization. The definition out of an automation perspective focuses on data points, devices and applications being organized based on attributes like location, function and more. (Jaylin, 2024) Compared to Reynolds' definition this a narrow perspective, as he claims that UNS is an architecture, not simply a nomenclature, that meets five requirements:

1. "UNS is the single source of truth."
2. "UNS is the foundation of future digital transformation."
3. "UNS is the hub that connects all smart devices and IT infrastructure."
4. "UNS is the sematic hierarchy of structure of business and data and events."
5. "UNS is where the current state of the business lives, enabling a real-life snapshot of the business." (Jaylin, 2024)

A UNS works by utilizing real-time data from the devices at the edge of the network and acts as a central hub or broker for all the exchanged information between these. It has the ability to collect data from different IIoT protocols, translate and add context to them and feed it as input data to systems that would use a different data format by nature. Every device gets treated as a node in the network and is allowed to both publish and subscribe to information through this centralized Unified Namespace.

The need for a new data architecture like the UNS can be derived from the shortcomings of the traditional automation pyramid, which was the result of the data management evolution that companies utilized pre-Industry 4.0. As elaborated upon in the previous sections, if a data engineer wants to make PLC parameters available in the cloud using traditional architecture, data must typically travel through the SCADA system, then to the MES, before finally reaching the cloud, relying on one-to-one connections with a poll/response method. While it is possible to connect a PLC or SCADA directly to the cloud using specific gateways, this approach requires separate implementation for each device, complicating the implementation, effectiveness, and scalability. With the increasing numbers of data sources in modern data-driven industrial companies comes the need to implement a transport protocol that is effective in adoption, later use, while being easily scalable. (Reynolds, 2023)

Many poll/response commands tend to devour bandwidth, as every data point is likely to be sent through several connections. This can in turn be handled by larger servers, which increase hardware costs, or selective polling of data, which might result in important data points being lost in the process. Another downside of having an excessive number of one-to-one connections is that in practice, the implementation will mostly be done over a longer time period, sometimes throughout decades. Therefore, the understanding and serviceability of one-to-one data architectures often relies on the human-to-human knowledge transfer over time, which increases dependency upon a certain group of people. All of this highlights the need for easier, standardized nomenclature and common data architecture. (Reynolds, 2023)

3.3.1 Data Modeling in Unified Namespace

As a response to the increasing data in the 1960s, data modeling was introduced to manage data by providing a way for specifying structures of data in a filesystem. “A data model is a set of models that can be used to describe the structure and operations of a database”. Meaning, if data is added to a database, it is necessary to have a data model to define how that data is added to said database. Data is traditionally organized in a statistical and standardized way based on the structure of data, constraints and other parameters. Structure of data refers to how the data is structured, i.e. if it is a tree structure, network model, hierarchal database and so on, while constraints is about setting up rules for data. Data modeling can be resembled by somebody saving a file on the computer and putting it in a folder, but everything happens automatically. (Navathe, 1992)

Data modelling is central to UNS since different subsystems, ERP, MES, etc., must understand the data exchanged. The data model is typically integrated into the Edge and is either a standardized operation, mainly aimed towards the ERP, or a custom-made for local analytics. Since UNS does not have a standard, the data model can be designed however an organization feels fits. However, there are some guidelines the data model should incorporate for the UNS to work well:

- Firstly, the data model must include *data version control/property*, which allows tracking of changes and compatibility with different versions. This is usually done through codification, meaning a version gets its own unique ID or serial number. Adding *timestamps* can be a way of creating unique IDs, but also makes it easier to perform analysis based on time series. Therefore, the general advice is to add a timestamp to the model.
- Similarly, every object in a UNS should have a unique ID, i.e. *object ID*. This will increase usability and identification for the user and optimize UNS tracking and identification.
- To be independent of external path context, it could also be necessary to add the topic structure to the data model, which is often done by embedding the *MQTT topic path*. The MQTT topic path can be seen as the topic structure of MQTT.
- It is also useful to add *sensor data*, like temperature or pressure etc., that will enable easier analysis and decision-making for the users.
- *Status information* is a vital parameter and should be added to the data model. This parameter defines if the asset is running, stopped, has maintenance notifications and/or error codes.
- To make the experience more pleasant for the user, functions like *descriptive metadata, operational parameters* and *location information* can be added. (Manditereza, 2023)

It is important to note that the UNS is not a master data model, but according to the previous definitions in section 3.2.

3.3.2 Common Implementations of Unified Namespace

While following the same basic principles, the Unified Namespace can be implemented in different ways depending on the use case or application, as a UNS architecture refers to how the network is structured, as opposed to the actual methods that are used to implement it. The actual communication protocol is mostly important for optimalization reasons; commonly however, the Unified Namespace is constructed using an MQTT broker as the intermediary. However, it is important to distinguish the difference between the actual Unified Namespace and the broker itself, as they are not the same thing. UNS is about the design approach to create contextual data flow for edge systems to exchange information, whereas the MQTT broker solely is the technology that makes it possible in an efficient and scalable manner using the pub/sub pattern. (Reynolds, 2023)

There already exist standards that specify how businesses can structure their data in a consequent and logical way. The go-to industrial standard for this guideline of organizing data is ISA-95, which defines how companies should structure their terminology to improve communication between industrial actors. Figure 8 is an illustration of the ISA95 standard and shows a folder like structure with enterprise acting as the main function, site as the next and so on.

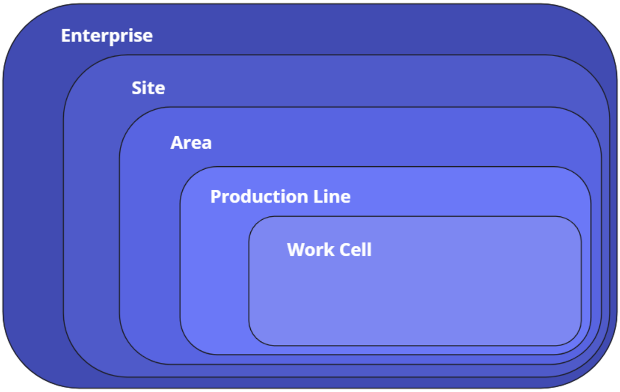


Figure 8: Structure according to ISA-95.

It is usually preferable to follow the same formatting as per defined in ISA-95 when constructing a UNS, however it is not a condition for successful implementation. Depending on the industry where the company is active, this terminology might not be suitable. In that case, UNS implementation allows for creativity and custom fitting for specific business purposes, as the main purpose is the usefulness. (Yngvesson, 2024)

Figure 9 below is an example of how a Unified Namespace can be implemented in a company making chocolate candy bars. In this example, the topics convention structure was constructed following the classical ISA-95 model, divided into sites, areas, production lines and so on. Devices, like a mixer or an oven, publish information to the namespace through the broker. In the case of older, non IIoT ready machinery (or MQTT ready in most UNS cases) it is possible to retrofit an IoT ready apparatus as a gateway to still access the rest of the network, made possible by point-to-point integration. In the Figure below, it is through the broker that all the applications communicate and are able to get access to the same information, easing handling for users and developers alike.

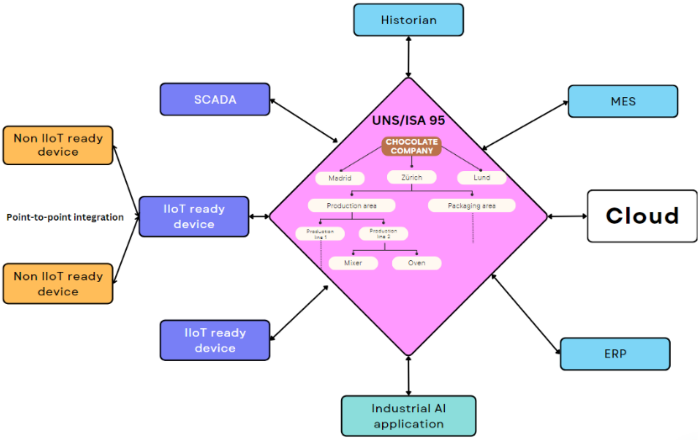


Figure 9. Example of a UNS implementation for a chocolate company with three production sites.

In theory, it is possible to have several Unified Namespaces implemented on different levels and connected to an enterprise level UNS. Technically speaking, nothing stands in the way of every production line having their own UNS, connecting to another *higher-level* namespace on a site or enterprise level. This is the basis of Schultz method. (Manditereza, 2022) As showcased in Figure 10 below, it may also be beneficial to have one system on each site that later connects to one namespace on the enterprise level in a similar manner.

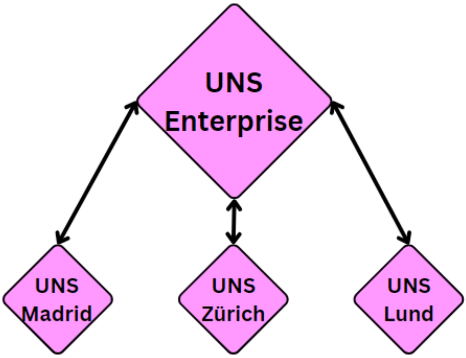


Figure 10. Example of the same chocolate company building UNS according to Schultz method.

Dividing namespaces into smaller units within a large business can significantly reduce complexity. Imagine a scenario where a company has three sites: one in Madrid with Spanish as the preferred language, another in Zürich where German is predominantly spoken, and a third in Lund where Swedish is used. Every site has ten production lines, with one-thousand signals in each line coming from PLCs, sensors and actuators, which in modern manufacturing systems is a conservative Figure. While all three sites use the metric system, the enterprise HQ is based in the USA and uses the imperial measurement system and only uses English as the primary form of communication, as well as general communication within the company mostly happening in English. In practice, attempting to implement one common UNS for these entities has the risk of ending up being unusable, as coordinating a unanimous platform can face severe pushback based on cultural differences. Moreover, automation engineers and maintenance personnel tasked with navigating this infinite jungle of signals might be overly challenged, impeding everyday work instead of easing it all. This highlights one implicit downside of the Unified Namespace, as collecting all data will inherently raise the complexity of the system.

As the main use case for this architecture is usability, it provides flexibility for implementation. As shown in the example, it might be more effective to establish an individual UNS hub at each site and another at the enterprise level. Some multi-national corporations have multiple plants and globally specialized in different aspects of the business, for instance manufacturing, recycling or logistics.

Another potential implementation of UNS is to allow suppliers and even customers, that an organization has close collaboration with, to link to the UNS. This scenario is illustrated in Figure 11, where a supplier is linked to the Chocolate Company UNS. Such integration can enhance the supplier’s and/or buyer’s innovativeness, its supply chain, and thus decrease risk and cost. The overall increase in performance will also benefit other parameters like sustainability. However, it requires a high degree of complex management, while increasing the risks associated with security factors, as well as potentially limiting future flexibility. (Ragatz, Handfield & Scannell, 1997) Therefore, it is important to implement such a solution with companies with a long history of collaboration and trust. It is also necessary to limit access to sensitive data and information.

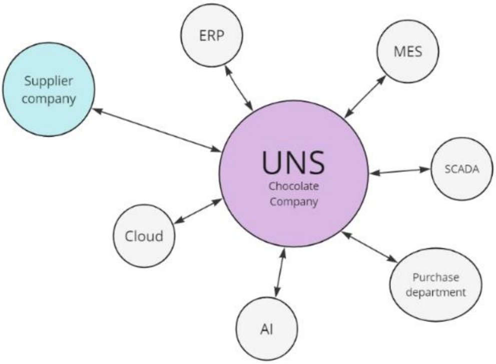


Figure 11. Example of a chocolate company building UNS allowing suppliers to connect.

Although implementing a UNS through an MQTT broker is the most common way, it is not the only one. Companies often use data-lakes to collect OT-data, which often tend to be cloud-based. In this case, edge devices publish data directly to the Cloud, without an intermediate to standardize data. A significant challenge is that different devices tend to publish data in a different structure, and many times without proper context. To structure and contextualize this large variation of OT-data, one can use a Unified Namespace, facilitating easier analysis, especially of data streams arriving from several plants. To summarize, the UNS architecture allows for flexibility, recognizing the different structures of businesses and not trying to implement a one size fits all solution. (Reynolds, 2023)

3.3.3 The Goal of Unified Namespace

The goal of UNS is to simplify device communication and data management. It achieves this through contextualizing and transforming data from machines and systems so that it becomes (more) usable. As the data gains context, it becomes easier to analyze, and since data is available in a centralized repository, it should become easier to find specific data points. Instead of solving problems by creating an ad-hoc structure, the UNS is a common infrastructure. (Reynolds, 2023) The main benefit of common data management is increased efficiency, both in terms of publishing the data, but also for development of future projects.

The implementation of a Unified Namespace for the handling of industrial data can standardize data nomenclature for the AI to make sense of, while also providing with the context. It centralizes data flows, allowing solely one data stream to be setup between the broker and the AI application, by being able to choose which stream it subscribes live data from. (Jaylin, 2024) Most likely, it will also require another stream from a database or a historian for access to historical data. This dataset provides the optimal AI integration, is only partly possible to collect with Industry 3 solutions (see section 3.1) but with the help of Unified Namespace (UNS) it should be possible to unlock the full potential of AI.

4. Research results

4.1 Implementation of Unified Namespace

4.1.1 Why is Unified Namespace Relevant?

As previously mentioned, UNS uses a common infrastructure for solving problems. A common structure will mainly internally simplify integration of new machinery, factories, debugging and adapting what data to publish. This is because as new machinery is purchased it is often possible to specify protocols beforehand and other potential drivers. The common structure will have a significant impact if an organization hires companies to develop a solution, as the contractor will immediately be familiar with the layout. Training staff and transferring them will also become cheaper and easier to do.

As UNS mostly uses sub/pub protocol, it uses less bandwidth than equivalent methods, thus preserving the PLC in the machines. This allows the company to increase data extraction with fewer connections, while increasing the lifetime capacity of the PLC. Furthermore, the lower bandwidth also allows for more scalability than similar solutions, as there is more available network capacity. Sub/pub protocols in combination with the repository, allow virtually infinite users to access the data without overloading the system. This affects the overall management of the company, becoming more efficient and easier for members of all levels of organizations to access data. Due to everybody having access to the information they need, in real-time, any potential developments or changes are reported to relevant people immediately. (Lindqvist, 2024) The fact that the data is transmitted and received in real-time inherently has many benefits in and of itself. In an Industry 3 solution the ERP would usually receive and transmit data on a monthly, quarterly, and sometimes even yearly basis. By having real-time relevant data available immediately (combined with sensible historical data), the organization will be able to respond faster to changing conditions, like factory breakdowns or material shortages. Decision making is thus improved with a higher level of accuracy. Machine operators and maintenance personnel also receive quality information in real-time, allowing them an overview without physically being present at each location.

Industrial data is central to the latest industrial revolution, as the interconnectedness of devices both generate and utilize great amounts of data. Analogies with data and oil have been made, arguing for an entirely new market segment emerging from data acquisition. However, if data lacks context, it also lacks value, as neither specialist person nor algorithm can make sense of it. It is like having oil but not refining it before using, causing more harm than good. Data without context can lead to misinterpretation and confusion not only for AI and DataOps systems, but for human consumption as well. Conclusions derived from data without context can also potentially lead to correlation without any causation. Meaning, it is possible to analyze the data and get results, but they have no effect on the process. Though contextualization of data issues relating to misinterpretation is minimized, AI and DataOps become more efficient. (Zohaib, o.a., 2023)

Once data is structured, there are many advantages, not only relating to analysis of it (with or without AI), but also in regard to the storage of data. Certain data is regulated and has to be stored for x number of years, depending on the industry. This is especially true for the airplane industry, where production and manufacturing data has to be stored for several years, in a manner that can easily be backtracked. If data has context, it is easy for manufacturers to filter through and thus store the corresponding data at the correct time. This saves capacity, as data is not stored unnecessarily and savings in terms of servers and cloud storage space can be made. Similarly, it might be useful to store all data from the PLCs, as its unknown parameters might prove useful for future applications. Structuring data after generation is extremely time consuming, and by possessing that same data from the start it is possible to escape a long start-up period. (Yngvesson, 2024)

A company that uses UNS and contextualizes all available data can create a new market segment, where that data is used to train ML-models and similar applications. By selling data for developing techniques, IT companies can produce a model adding value in the segment. This product will highly benefit the selling companies if they are granted access, as it is based on their data and use cases. Some organizations already see value in this market, and currently give data for certain operations away for free for access to a model trained on that data, used to optimize their production. Another possible application is a new service model where companies offer to train ML-models for one company and license it out for others. (Yngvesson, 2024)

A survey from 2018 shows that almost 80% of companies use some sort of Electronic Performance Monitoring, where both negative and positive behaviors are measured. Companies can usually prevent theft and other negative behavior, while employees performing well are rewarded. (Tomczak, Lanzo, & Aguinis, 2018) A potential application of UNS is monitoring purposes, as the data is published and received in real-time. Although not the main idea behind UNS, it is hard not to think about it because application would be both easy to do and useful. By creating tags on event-based data tied to, for example, an operator, it would be possible to see how long reaction time to alarms and operations are. To exemplify, an operator of a chocolate packaging machine is tasked with changing the paper role for wrapping every 45 minutes. A manager can easily obtain how long it takes that operator on average to react to an alarm that it is time to change the role and how long it takes to change the role.

To expand beyond the monitoring of the employees, the result of these tags can be used to identify operational bottlenecks instead. Data collected can in turn be used as a basis of argumentation for automating, or not automating specific parts of a production process. The parameter used for comparison between human operations and machine is often time based but can easily be modeled and calculated to represent financial incentives. Furthermore, tags can be created to identify other bottlenecks that are not tied to employees, find new improvements and investment potential.

Companies are actively working with traceability, and it has become a vital risk management tool for most industries, food and automotive among them. Traceability is also required in certain industries by law but can also be used as a tool to avoid legal troubles. Traceability is about being able to identify every batch, or product, tool and object to identify any or all information about it. (Schuitemaker & Xu, 2020). Through high traceability, companies aim to avoid large recalls and identify common failures with its production. The more specific traceability the system allows, the less products have to be recalled and higher customer satisfaction can be achieved, while optimizing production at the same time. Commonly, traceability is achieved by giving a batch, product or process a unique ID. The goal is to have a unique identifier for each product and every part that encounters that product. Even though there are clear benefits with traceability in manufacturing, it is rarely implemented beyond the batch level, mostly due to the large amount of data that a highly specific system would produce. (Lindqvist, 2024) Further complicating matters, industries often struggle with data standardization, undergo numerous data transformations, and manage complex, extensive supply chains, all of which require integration into specified traceability solutions. (Yngvesson, 2024) Implementing traceability at the batch level within a production line or factory is relatively straightforward. However, beyond that, data collection and integration become significant challenges. With UNS, it is possible to achieve a holistic perspective. Through data modeling it is possible to create unique identifiers for every product, leading to efficient data integration and improved scalability. Since suppliers and customers are able to connect to an organization’s UNS, it also simplifies traceability of products along the value chain, while trying to establish a standardized way of naming products and operations.

4.1.2 Potential Drawbacks of Unified Namespace

As mentioned, most companies use a mixture of new and old machinery. According to the study done in the Netherlands, the average age of industrial machinery varies between 26 to 34 years (Erumban, 2018), and it is not uncommon for certain sectors to have machines that are over 60 years old. Machinery over 60 years usually miss PLCs, and instead relies on older technology like relay logic control but can also have one or more PLCs retrofitted. Of course, there exists mechanical machinery that lacks the need for programmable control. However, most machines that are 40-60 years old either came with a PLC preinstalled, as they became more widely used in the late 80s, or had one retrofitted, especially if they are complex machines. (Lindqvist, 2024) Nevertheless, these machines use old technologies. Older PLCs lacked standardization, and different models and manufacturers used different communication protocols, programming languages and languages depending on region. Extracting data from these machines and PLCs can thus be time-consuming and costly. It is possible to use an intermediate broker such as KepServerEX, that translates protocols to more commonly used ones in modern industry, like OPC UA or MQTT. There, however, exist a number of legacy machines that were never connected to a common network, and doing so for the first time will inherently increase costs for companies implementing UNS. Management is likely to ask for proof (perhaps rightfully so) why and how this would increase output or efficiency for the machines that performed fine without any network connection before.

When integrating a machine into the UNS, it is still necessary to have somebody with knowledge of the PLC’s specifications. A desirable feature of the UNS is that it can convert and translate data from one unit to another, for example Fahrenheit to Celsius. If a unit is changed in the PLC, the UNS data model is not smart enough to correct the transformation and will be publishing wrong data. Therefore, the UNS will not be able to remove and replace the programmers that have the know-how of the current systems. However, it hopefully limits dependability on other specialists. (Yngvesson, 2024)

Even though UNS uses lightweight protocols, such as MQTT, it may require large servers to process data from a large factory or multiple smaller ones. This is even more accurate if all data is extracted from all PLCs within a large production, and if that data also requires some sort of processing. In case all data is stored in historians and clouds, the storage solutions need to be extensive, implying premium costs. Therefore, if the organization wishes to extract all available data with context, it can become more expensive to implement a UNS solution.

Adding to investment complexity, it is not entirely apparent what measurable KPI parameter improve because of UNS. UNS does solve problems related to dataflow and enhances the way a business operates. However, it is hard to single out and measure how much more productive an employee becomes because of UNS, since there are multiple factors affecting it. UNS will not improve the efficiency of a production line or site by itself but will make the process of doing so easier and more efficient. (Yngvesson, 2024) (Lindqvist, 2024)

Surveillance and monitoring of staff is tricky as they are associated with multiple negative reactions like privacy invasion, decrease in job satisfaction, increase in stress and more. (Tomczak, Lanzo, & Aguinis, 2018) Increase in negative reactions will likely decrease performance of employees and increase employee turnover, which has a negative impact on the company's bottom line. Increased monitoring will not only affect the mood and emotions of employees but can affect innovativeness, as employees can become hesitant to take risks, fearing repercussions. All of this can lead to a decrease in competitiveness. Ethically and morally, it is questionable to always monitor staff and use those parameters as ground for evolution, especially since it is well documented that it causes stress and feelings of distrust. To exemplify, if a company only uses KPIs to evaluate employees, it can be at the expense of product quality. A company aiming to always produce a certain number of units will give incentives for employees to produce the same numbers of units, regardless of quality, taking larger safety risks and so on. Therefore, long term costs of only using efficiency KPIs is likely higher than creating an environment where employees instead care for machines and the products produced.

One of the mains advantages of UNS is that a standard is applied to storage of data, and it is therefore advantageous to collect all data from operations. This is due to uncertainty in which parameters new applications need to operate optimally, and what parameters ML use for analyzing a machine or process. However, there is a significant risk that some data extracted from operations will never be used, leading to unnecessary storage costs and ongoing maintenance. Additionally, leveraging this data effectively necessitates active analysis by the organization. While the ML will give better estimates with more data, the impact of extra parameters is questionable, considering the cost associated with storing the extra data and management of it. There is a breaking point where extraction of data becomes more expensive than the value that algorithms create. AI can, as previously mentioned, help with creating a repositor of data that it predicts will be useful, reducing unnecessary data accumulation and costs.

A major barrier for the new market segment, where data becomes an asset, is that companies are not willing to share data, since it may compromise their market position. Left are companies that are not in a competitive segment, like state-owned authorities etc., but even these are not likely to share data for concerns of safety. It is not possible for developers to develop an optimization tool with highly limited access to data. Furthermore, it is hard to make a model generic, since ensuring that a solution from one application works for another one of a different brand or model, is difficult. Ensuring that a program is generic is further complicated if a company only provides part of a process, as a consequence of secrecy. All of this stems from industrial processes being highly specific, ad-hoc and complex.

4.2 A Practical Example of a Unified Namespace Implementation

Novotek AB is a company active in several European countries, focusing on industrial digitalization, automatization and analytics. Two of their representatives, Nils Lindqvist and Jens Molin, currently working as acting Digital Innovation Manager and Managing Director in Sweden, respectively, have aided in the realization of this report, providing both knowledge and resources. Novotek AB, among other things, helps companies implement Unified Namespace(s) in order to streamline their operations in the long-term. In the following example, a possible way of implementing a UNS will be presented, which will be based on a real-life case, using a mixture of software that Novotek AB might or might not use. However, it is important to clarify that this is simply one way of doing it, and it does not mean that other ways of implementation are bad or wrong. Unified Namespaces can be implemented using a variation of software, messaging protocols and technical solutions.

4.2.1 Software Introduction

Novotek AB is partner to an American firm, HighByte Inc., which develops an industrial data architecture and tracking software called HighByte Intelligence Hub. This software is generally used to access, model and transform plant floor data for analysis, acting as a bridge between machinery and other systems. The program can establish connections with a number of different clients using various protocols, such as MQTT, OPC UA, REST messaging protocols, PostgreSQL or Microsoft SQL databases, or AWS IoT SiteWise and Azure IoT Hub to name a few. It can create conditions for aggregates, deadbands together with custom condition in order to transform raw data into a desired format and create models to contextualize and standardize industrial data. (HighByte, 2024) Novotek AB, among other things, works with the implementation of the Intelligence Hub for their customers, including everything from deployment to service and support later in the process. It offers a convenient embedded MQTT broker that supports communication with both external brokers and can interpret messages both in JSON and Sparkplug formats. (HighByte, 2024) Moreover, the embedded broker can connect with browser-based (HTTP-centric) applications, interfacing clients with either MQTT, TCP or WebSocket, mqtt.ws (MQTT over WebSocket), secured with Transport Layer Security. Considering the flexibility and accessibility of the Intelligence Hub, it will be used during this project to demonstrate the implementation of a Unified Namespace.

Another software that Novotek AB uses is KepServerEX. It is used as an OPC-communication platform that uses as a middleware between shop floor machinery, and other, higher-level applications using a collection of drivers. The software is certified by the OPC Foundation and contains drivers to most existing PLCs in order to translate all data into a common OPC format. Using KepServerEX it is possible to create a demo server, as one can create virtual devices from the leading automation manufacturers through, for example, OPC UA, which will be demonstrated later. The software is generally used to handle the separate connections and data flows from edge devices, collecting them all in one place to be handled, and most likely fed forward to a database, a MES or an ERP. It is possible to incorporate a REST or an MQTT agent into the software to translate between protocols, however the natural communication will mostly take place through some OPC protocol. While there are overlaps between Intelligence Hub and KepServerEX, one can say that KepServerEX specializes in the low-level communication within the production site, while the other software functions on the higher-level connections. (PTC, 2024)

A broadly used software for handling supervisory control and data acquisition (SCADA) is Ignition SCADA by Inductive Automation. It offers a browser-based toolset, allowing for process control, HMI functions, as well as data analysis and tracking within the platform. The program has several overlapping functions with the aforementioned software, such as the use of common communication protocols (MQTT, OPC-UA, SQL), the use of tags as machine and event attributes, as well as fairly simple display of data.

During the demonstration, Ignition SCADA will be mostly used as a tool for visualization of simulated tags and data flows. In traditional data architectures following the automation pyramid presented in Figure 1, programmable logic controllers are considered to connect the edge devices from level 0 to 1. SCADA systems are placed at level 2, the data acquisition level, and act as the middleware communicating with control devices and planning systems. (Inductive automation, 2024)

For creating a database where generated data could be stored, Microsoft SQL was used. As the name suggests, the software is developed by Microsoft, a large player within industrial software development historically. For access to historical data, it is a requirement to save it somewhere first, as most signals are only sent once by default. To the SQL database, it was possible to connect yet another commonly used analytical software, namely Microsoft PowerBi, supplied by the same developer.

4.2.2 Simulation of a Virtual Production Site and Implementation of a UNS

The demonstration is implemented virtually on Dell laptops, acquired from Novotek, running the Windows 11 operating system. The dataflow can be seen in Figure 12 and will start with an Excel document that contains parameters and corresponding values of three fabricated machines. Data is put into KepServerEX, transformed into OPC-UA protocol and fed into HighByte’s Intelligence Hub. Within HighByte’s Intelligence Hub, the data can be transformed into various messaging protocols and is hence transferred to Ignition SCADA via SparkplugB, Microsoft SQL via SQL, while creating MQTT UNS topic structures for other possible software. The database created in Microsoft SQL is also connected to Microsoft PowerBi. For the purpose of this report, even though it is possible to connect to additional platforms with previously mentioned messaging protocols, this representation is deemed ample to exemplify a UNS implementation.

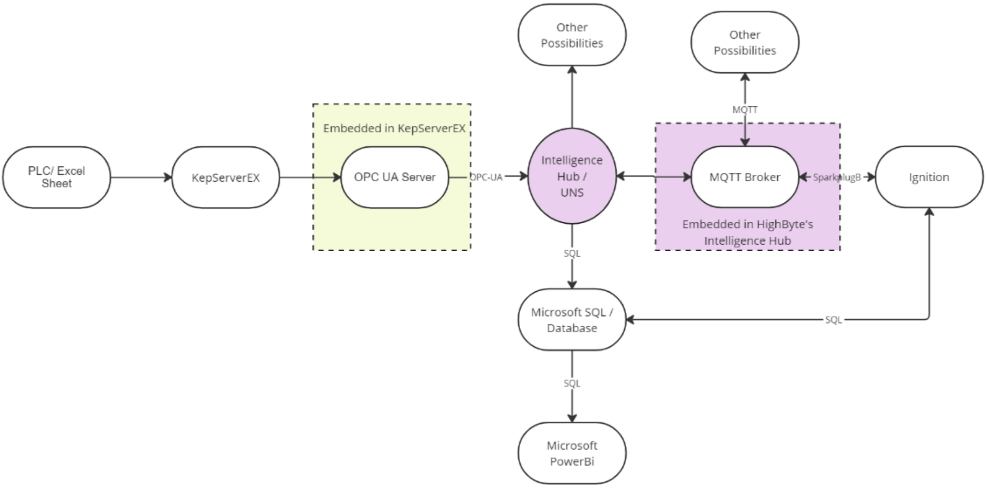


Figure 12 Representation of UNS Implementation.

The demonstration was initiated by creating values that could represent a process or a machine line within a factory. These values were based on common sensor values extracted from PLCs, like surrounding temperature, speed, vibrations and so on. Every "PLC" contained around 10-20 variables and each variable had 1614 data points, which amounted to 26 minutes and 54 seconds of running time, as a value is polled every second. The simulation could have been both longer and shorter without affecting the results. After all values are polled, the program automatically restarts. The three different machines created represent a packaging line, meaning if one of the machines breaks down the others have to stop as well. The line, *dubbed Line 1*, consists of a pellets filler, a packaging machine that closes the boxes and a machine that wraps the boxes in plastic film, if it is requested. Figure 13 is an illustration of the packaging line. Both the pellets machine and the packaging machine are used for each box, whilst the wrapper is used for a random number of boxes dependent on speed. Variety is used to indicate that the specifications of the orders vary depending on different products.

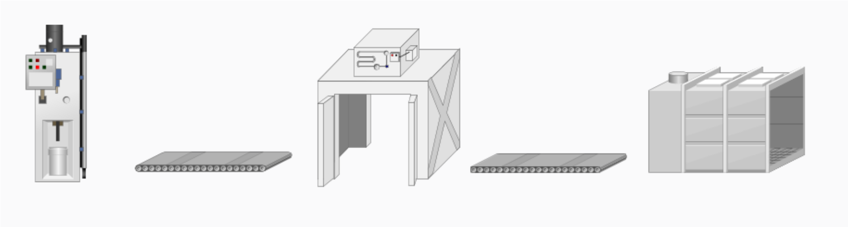
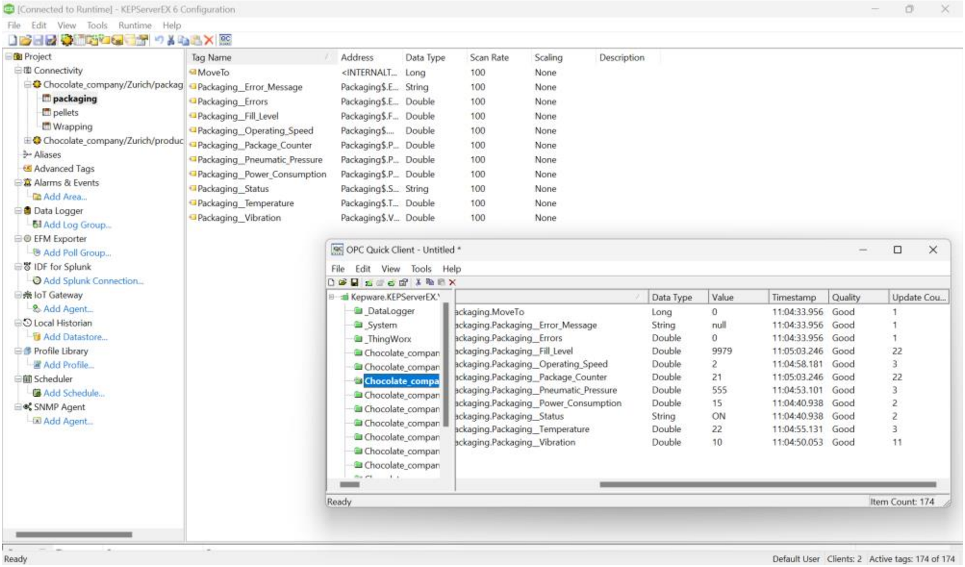


Figure 13, Representation of Line 1.

After the machines are configurated in the .csv file, the data is extracted onto KepServerEX. Within the KepServerEX program and the connectivity tab, a channel is created and named according to the ISA-95 standard. The Advanced Simulator is chosen as a driver that, among other functions, enables connection to the Excel file while also deciding the polling rate. As previously mentioned, the polling rate was set to 1000ms. Three virtual machines are created and named accordingly in the channel, *packaging, pellets* and *wrapping*. The tags are generated automatically from the Excel file with correct data types, corresponding names and so on, which enables a straightforward transformation. Tags can be seen as an intermediary variable between the PLC, HMI and data processing and define a parameter or attribute of a machine, such as temperature or error messages, and are essential for data processing, translation and understanding. In Figure 14, the KepServerEX interface can be seen, where the three machines are visible in the channel named according to ISA-95 standard. Within the packaging machine 11 tags can also be seen, however, the first one is an internal tag and therefore lacks significance. Embedded in KepServerEX is an OPC-UA server, which translates protocols and data into OPC-UA protocol. The pop-up screen “OPC Quick Client”, seen in the Figure, is used to quickly visualize the tags and machines. Throughout the implementation stage it was used to review changes and verify that everything worked correctly.



Tag Name	Address	Data Type	Scan Rate	Scaling	Description
MoveTo	<INTERNALT...	Long	100	None	
Packaging__Error_Message	Packaging$.E...	String	100	None	
Packaging__Errors	Packaging$.E...	Double	100	None	
Packaging__Fill_Level	Packaging$.F...	Double	100	None	
Packaging__Operating_Speed	Packaging$....	Double	100	None	
Packaging__Package_Counter	Packaging$.P...	Double	100	None	
Packaging__Pneumatic_Pressure	Packaging$.P...	Double	100	None	
Packaging__Power_Consumption	Packaging$.P...	Double	100	None	
Packaging__Status	Packaging$.S...	String	100	None	
Packaging__Temperature	Packaging$.T...	Double	100	None	
Packaging__Vibration	Packaging$.V...	Double	100	None	

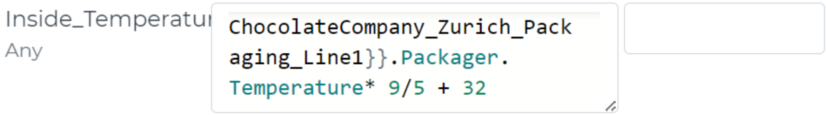
	Data Type	Value	Timestamp	Quality	Update Cou...
ackaging.MoveTo	Long	0	11:04:33.956	Good	1
ackaging.Packaging__Error_Message	String	null	11:04:33.956	Good	1
ackaging.Packaging__Errors	Double	0	11:04:33.956	Good	1
ackaging.Packaging__Fill_Level	Double	9979	11:05:03.246	Good	22
ackaging.Packaging__Operating_Speed	Double	2	11:04:58.181	Good	3
ackaging.Packaging__Package_Counter	Double	21	11:05:03.246	Good	22
ackaging.Packaging__Pneumatic_Pressure	Double	555	11:04:53.101	Good	3
ackaging.Packaging__Power_Consumption	Double	15	11:04:40.938	Good	2
ackaging.Packaging__Status	String	ON	11:04:40.938	Good	2
ackaging.Packaging__Temperature	Double	22	11:04:55.131	Good	3
ackaging.Packaging__Vibration	Double	10	11:04:50.053	Good	11

Figure 14: Snapshot from KepServerEX showing tags for the packaging machine and the OPC Client as a pop-up.

The OPC-UA client is created in KepServerEX and is connected to HighByte's Intelligence Hub. The connection between Intelligence Hub and KepServerEX is established through the OPC-UA server, which is configured to use a specific port. A port can be seen as a designated channel within the network that allows for specific types of communication flow. Each port has a unique number, in the case of Kepware it is 49320 by default. As an OPC-UA network connection is established, it is also visible within the Intelligence Hub. In the case of this implementation, KepServerEX's output is used as input variable intended for other applications to receive. Connections to other applications, like Ignition Scada, will represent outputs from the view of Intelligence Hub. Within HighByte's Intelligence Hub the data is modeled, i.e. relevant tags are chosen, the variables are formatted if the old values are not suitable, and they are given context.

This can be exemplified by the tag names derived from KepServerEX. The tag name contains the Excel sheets' name in addition to the parameters name and as this information is not relevant, it is removed with HighByte’s modeling tool.

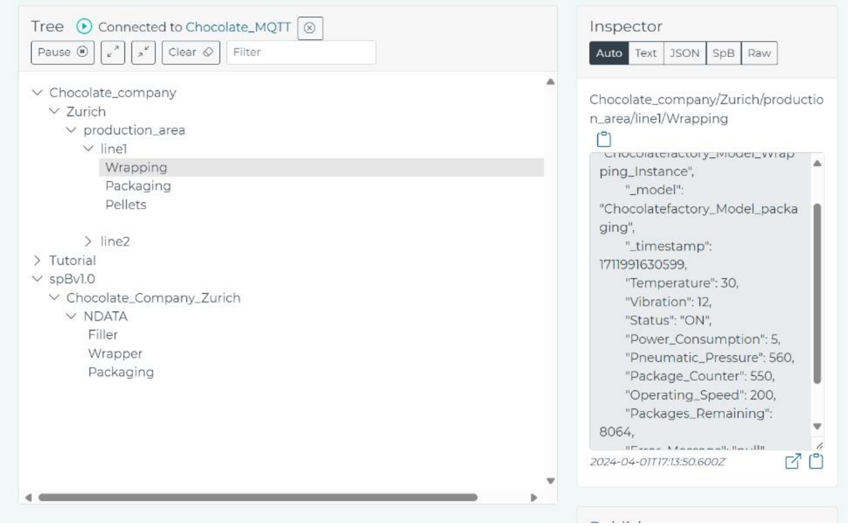
Likewise, one variable is formatted from Celsius to Fahrenheit as UOM, which is executed in the modeling tool, which can be seen in Figure 15. In this demonstration, a model is created for each machine and can be used for all outputs to get a uniform data structure. If one aims to achieve a more uniform data structure, it can be achieved by using the same model for all machines. By doing so, all machines will have the same attributes in the subscribed software. Here, it is also possible to put a default value, which would appear as *null* otherwise.



```
ChocolateCompany_Zurich_Pack
aging_Line1}}.Packager.
Temperature* 9/5 + 32
```

Figure 15. Formatting a temperature variable from the Celsius scale to Fahrenheit within Intelligence Hub’s modeling tool.

Most importantly, all the data is transformed into a UNS structure within Intelligence Hub, as it functions as a real-time central repository for other applications to connect and transforms data into pure MQTT protocol through its embedded MQTT broker. The structured data can be seen in UNS Client embedded in HighByte’s Intelligence Hub in Figure 16. The clear advantage of using pure MQTT compared to Sparkplug B is seen in Figure 16, which is the hierarchical depth achieved with MQTT (tree above), as Sparkplug B is limited to three levels (tree below).



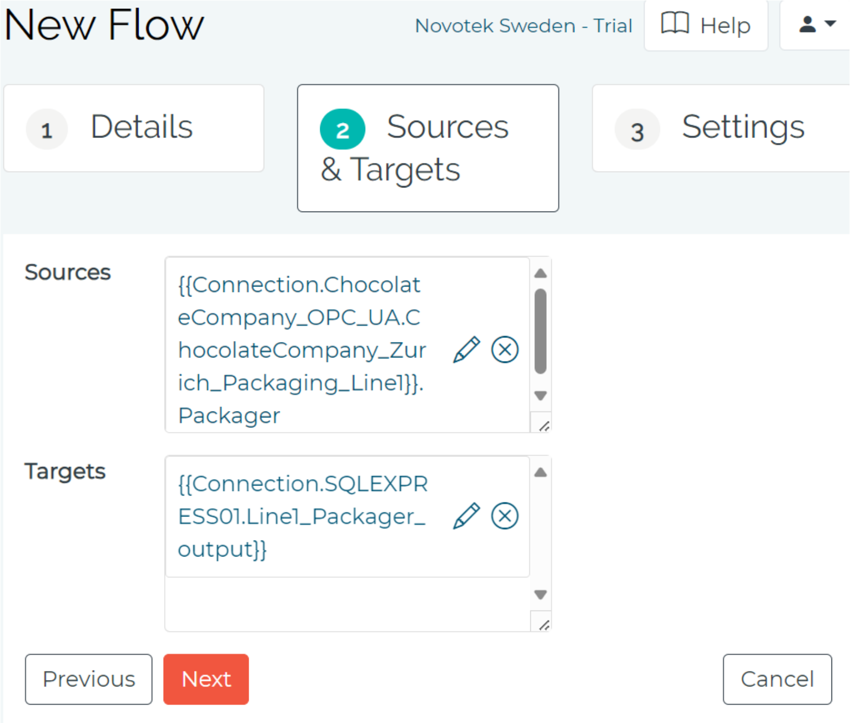
Chocolate_company/Zurich/production_area/line1/Wrapping

```
ping_Instance",
      "_model":
"Chocolatefactory_Model_packa
ging",
      "_timestamp":
1711991630599,
      "Temperature": 30,
      "Vibration": 12,
      "Status": "ON",
      "Power_Consumption": 5,
      "Pneumatic_Pressure": 560,
      "Package_Counter": 550,
      "Operating_Speed": 200,
      "Packages_Remaining":
8064,
```

2024-04-01T17:13:50.600Z

Figure 16. The UNS client using pure MQTT in the top, while using Sparkplug B at the bottom.

One outgoing connection of the UNS client is Ignition SCADA through a local port. As the communication to Ignition SCADA is limited to Sparkplug B for the sake of this work, the data is transformed to Sparkplug B within HighByte's Intelligence Hub. This is achieved by building a flow that has the OPC-UA data as an input and Sparkplug B as an output. A flow can be described as the transformation between different protocols and programs, connecting an input and an output, and is an internal function of Intelligence Hub. Designing a flow is straightforward, as it is a drag and drop feature where input variables originate from the source data, the desired transformation happens internally, and transforms into the desired output form. The inputs and outputs are created within the connections and are modified to include desired parameters and context. In Figure 17, a flow is being created. It sources the OPC-UA connection and targets a model that is connected to the SQL output, i.e. it is going to transmit in SQL.

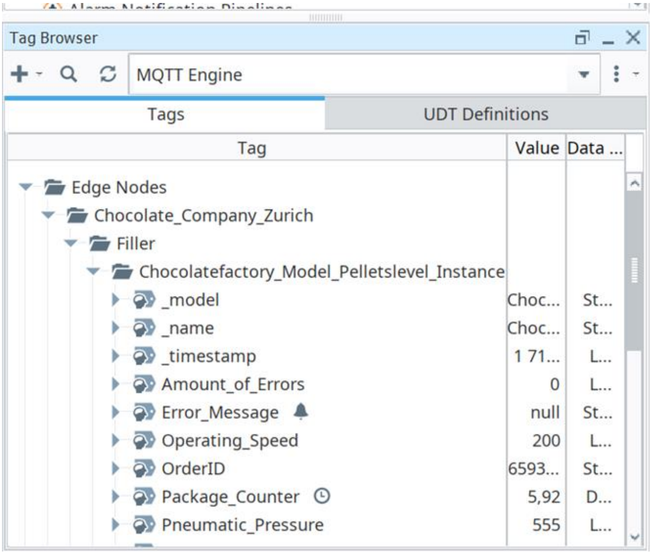


Sources: {{Connection.ChocolateCompany_OPC_UA.ChocolateCompany_Zurich_Packaging_Line1}}.Packager

Targets: {{Connection.SQLEXPRESS01.Line1_Packager_output}}

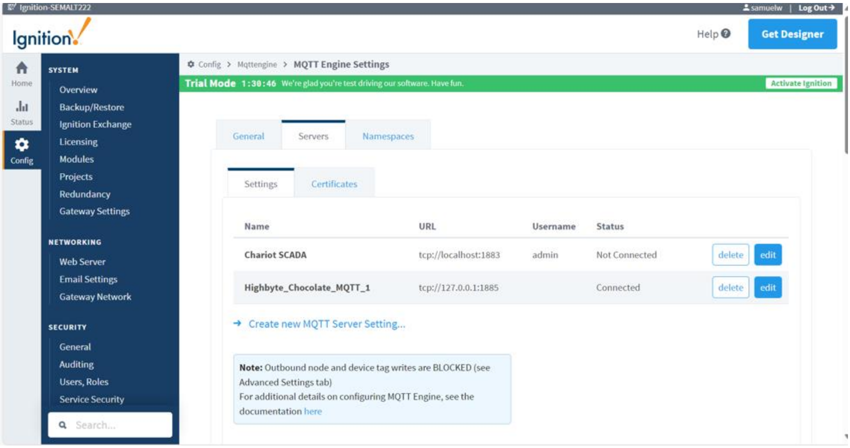
Figure 17. Creating a flow for the packaging machine to the database.

For the Ignition SCADA connection, the input is a model created for a machine, for example the packaging machine, and the output is extracted through the Sparkplug B connection. The output connection has the same port number as the corresponding input within Ignition SCADA input, which in this case is 1885. Within the flow, the polling frequency is also established. Likewise, a connection must be built within Ignition SCADA and as its native communication protocol is OPC-UA, a plug-in is downloaded which allows for MQTT communications. In Figure 18 the tags extracted from Intelligence Hub can be seen in the Ignition SCADA interface, while Figure 19 shows the plug-in that facilitates Sparkplug B communication.



Tag	Value	Data ...
Edge Nodes		
Chocolate_Company_Zurich		
Filler		
Chocolatefactory_Model_Pelletslevel_Instance		
_model	Choc...	St...
_name	Choc...	St...
_timestamp	1 71...	L...
Amount_of_Errors	0	L...
Error_Message	null	St...
Operating_Speed	200	L...
OrderID	6593...	St...
Package_Counter	5,92	D...
Pneumatic_Pressure	555	L...

Figure 18. Tags for the pellets machine in real-time in Ignition.



Trial Mode 1:30:46 We're glad you're test driving our software. Have fun.

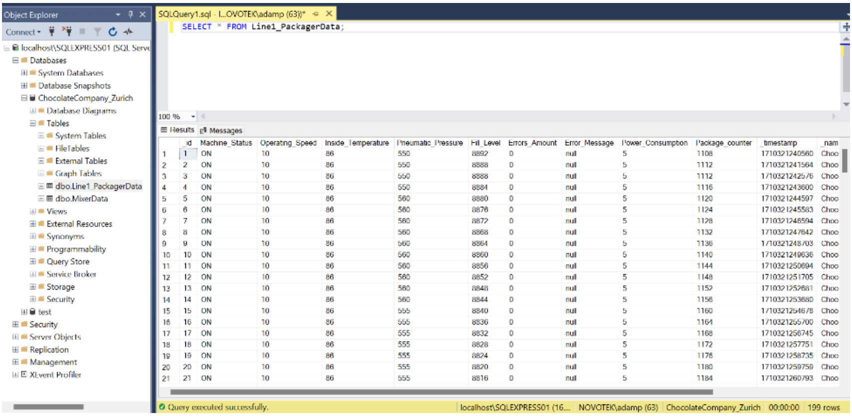
Name	URL	Username	Status
Chariot SCADA	tcp://localhost:1883	admin	Not Connected
Highbyte_Chocolate_MQTT_1	tcp://127.0.0.1:1885		Connected

Note: Outbound node and device tag writes are BLOCKED (see Advanced Settings tab)
For additional details on configuring MQTT Engine, see the documentation here

Figure 19. MQTT connection for Intelligence Hub into Ignition SCADA.

The tags are used to build an interface that, among other things, can be used to track machine performance, acknowledge and address alarms, and also see orders. In Appendix A, screenshots from the platform and its design can be seen. As it is not central for this report, the design of the platform is rather simplistic to showcase and highlight functionality of the UNS structure, rather than a functional platform. For real world scenarios it is suitable, and likely requires building more complex platforms.

Connected to Ignition SCADA is also an SQL database. An SQL database is often used as storage by industrial companies generating an enormous amount of data through their everyday processes, ranging from production in- and output, to internal processes and sales information. Within the database, the data is stored within tables. A query returning data from the packaging machine located in Line 1 can be seen in Figure 20, which displays the saved data from the machine. This database is used for functions within Ignition SCADA and PowerBI that require historical data. It is common that companies would use a database as their data storage, and poll it to create charts used in different ERP systems and other analysis. As the MQTT broker only has a real-time application, it is preferable to connect a database or a historian to save historical data.



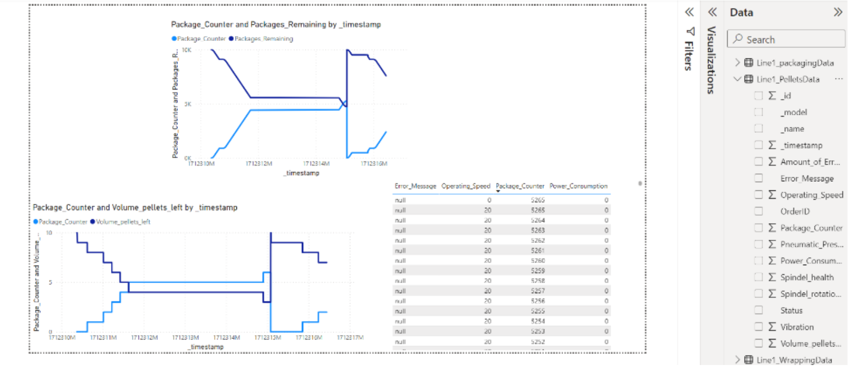
```sql
SELECT * FROM Line1_PackagerData;
```

	_id	Machine_Status	Operating_Speed	Inside_Temperature	Pneumatic_Pressure	Fill_Level	Errors_Amount	Error_Message	Power_Consumption	Package_counter	_timestamp	_nam
1	1	ON	10	86	550	8892	0	null	5	1108	1710321240560	Choo
2	2	ON	10	86	550	8888	0	null	5	1112	1710321241564	Choo
3	3	ON	10	86	550	8888	0	null	5	1112	1710321242576	Choo
4	4	ON	10	86	550	8884	0	null	5	1116	1710321243600	Choo
5	5	ON	10	86	560	8880	0	null	5	1120	1710321244597	Choo
6	6	ON	10	86	560	8876	0	null	5	1124	1710321245583	Choo
7	7	ON	10	86	560	8872	0	null	5	1128	1710321246594	Choo
8	8	ON	10	86	560	8868	0	null	5	1132	1710321247642	Choo
9	9	ON	10	86	560	8864	0	null	5	1136	1710321248703	Choo
10	10	ON	10	86	560	8860	0	null	5	1140	1710321249636	Choo
11	11	ON	10	86	560	8856	0	null	5	1144	1710321250694	Choo
12	12	ON	10	86	560	8852	0	null	5	1148	1710321251705	Choo
13	13	ON	10	86	560	8848	0	null	5	1152	1710321252681	Choo
14	14	ON	10	86	560	8844	0	null	5	1156	1710321253680	Choo
15	15	ON	10	86	555	8840	0	null	5	1160	1710321254678	Choo
16	16	ON	10	86	555	8836	0	null	5	1164	1710321255700	Choo
17	17	ON	10	86	555	8832	0	null	5	1168	1710321256745	Choo
18	18	ON	10	86	555	8828	0	null	5	1172	1710321257751	Choo
19	19	ON	10	86	555	8824	0	null	5	1176	1710321258735	Choo
20	20	ON	10	86	555	8820	0	null	5	1180	1710321259759	Choo
21	21	ON	10	86	555	8816	0	null	5	1184	1710321260793	Choo

Query executed successfully. | localhost\SQLEXPRESS01 (16.... | NOVOTEK\adamp (63) | ChocolateCompany_Zurich | 00:00:00 | 199 rows

Figure 20. Historical data saved in a database, collected from Intelligence Hub.

The same database is also connected to Microsoft PowerBi, which is not facilitated through Intelligence Hub, but is instead, established directly between the two platforms with native SQL communication. As no transformation or added context is needed, it is simpler to use a direct connection as opposed to establishing a connection through the central broker. Nevertheless, the data was previously transformed and modeled within UNS, and if desired, another transformation to Rest API protocol can be created in order to have direct communication between PowerBi and Intelligence Hub. Microsoft PowerBi is a software used for visualization and analysis of organizational data and designed for business intelligence. Below, a sheet can be seen showcasing production data in charts and tables, visualized in PowerBi.



Error_Message	Operating_Speed	Package_Counter	Power_Consumption
null	0	5265	0
null	20	5265	0
null	20	5264	0
null	20	5263	0
null	20	5262	0
null	20	5261	0
null	20	5260	0
null	20	5259	0
null	20	5258	0
null	20	5257	0
null	20	5256	0
null	20	5255	0
null	20	5254	0
null	20	5253	0
null	20	5252	0

Figure 21. A screenshot from Microsoft's PowerBi getting its data from the database.

4.2.3 Alternate Unified Namespace Structures

The previous implementation is not an exclusive structure of building a UNS and there are many common software that can be used to build a UNS. For example, the first "UNS" was built by Walker Reynolds using Data Highway Plus (an older communication protocol) and Excel, where he built the first system that acted as the single source of truth for all operation in a salt mine. (Reynolds, 2023) However, it is worth pointing out that there are several reasons for not building a UNS in Excel, mainly relating to scalability and limited functionality, which will lead to it becoming overloaded and likely many errors. Building a UNS for limited data at one location is straightforward, as essentially it is about collecting all data in a central location in real-time. A modern production produces enormous amounts of data constantly and it is unlikely that this solution would work for more than a couple machines, not to mention a large production or multiple sites. Extraction of data to different platforms is also questionable as many systems use different communication protocols and data transformation is limited, given that it is not designed for industrial applications. Walker Reynolds himself said that they stopped using Excel due to these factors, and explained how his first enterprise level UNS was built using Ignition SCADA. (Reynolds, 2023) Ignition SCADA is an immensely capable software and as it historically has been used for mainly SCADA application, data is transmitted in real-time with high accuracy. Ignition can also be used outside of the SCADA level, as a visualization program for Enterprise level operations, and has a built-in simple database. Even more importantly, data can be given context to and structure within Ignition. Data is given context using User-Defined Types, UDTs, where the user defines tags in a folder structure, enabling an object-oriented approach and creating parameterized data tables (preferably in an ISA-95 structure). UDTs also offer features like changing parameter values if an instance has been established and map nested directions onto MQTT topic structures, meaning it is essentially transformed into MQTT protocols. In Figure 22, an example of a UNS built using different MQTT protocols can be seen. (Inductive automation, 2024)

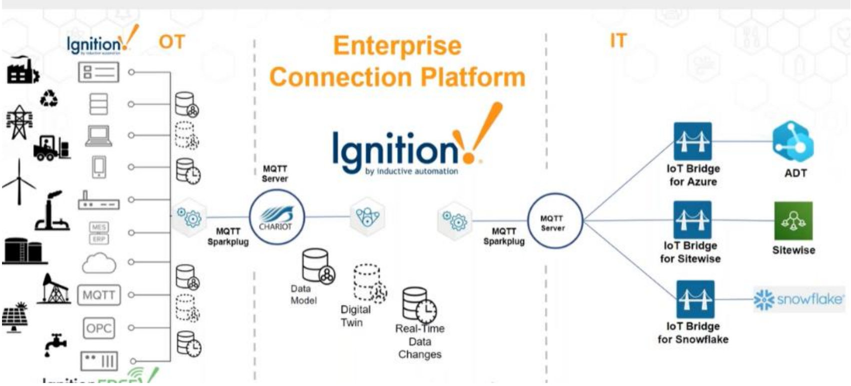


Figure 22. A representation of a UNS built using mainly Ignition. (Inductive automation, 2024)

In the case of Walker Reynolds, his first UNS built in Ignition used OPC-UA protocols, converted from Modbus, and was at the time the largest standalone SCADA system in the world. The system had 11 million tags/data points, 2000 concurrent users, 2 million daily alarms, covered five states and 40.000 connected devices. However, according to Walker Reynolds it was hard to scale, and they reached limitation due to the protocols, originating the switch to MQTT within Ignition. It is worth noting that the UNS was built in 2013 and both the OPC Foundation, but mainly Ignition, has developed their product since then, so the switch may have been avoidable today. Walker Reynolds does, however, believe that MQTT is the ultimate approach to a UNS, as it is more scalable than OPC-UA. (Reynolds, 2023)

Another common set of software used for building UNS is HiveMQ MQTT broker, a platform with a built-in MQTT broker used for distributing messages, and generally a DataOps platform, used mainly for data modeling before the data reaches distribution. Essentially, it is possible to only use HiveMQ, but the distributed data will lack context and will not be normalized, making it hard to read and use. Therefore, another dataOps application is recommended to use, as one of the key-point of UNS is uniform data. Similarly to Walker Reynolds Ignition solution, this setup distributes uniform messages to multiple concurrent users and availability for millions of data points. However, since HiveMQ is a MQTT platform it is more scalable, as the protocol uses less bandwidth, because it is both lightweight and uses sub/pub protocols. Another benefit of MQTT is the topic structure that is enabled through the protocol, allowing for simple location – and tag traceability. Nevertheless, similar structure should be attainable through data modeling and cataloging, making the benefit specifically tied to structure somewhat obsolete.

These alternatives and examples serve as a means of emphasis that there are multiple ways of constructing a UNS and the software used in the demonstration in this report are not the only way of achieving a UNS. Furthermore, it also highlights that a UNS is not tied to a specific program or set of protocols, but rather is a concept that can be achieved in multiple ways.

It also highlights limitations for a UNS is likely due to the software and protocols used for a current system, and that by switching software they may be reduced or eliminated.

4.3 Interviews with Companies

In today’s rapidly evolving industrial landscape, as businesses transition towards using ever more data-driven processes, it has become paramount to actively work with digitalization and data management in order to stay competitive. The following interviews are a result of engaging with company representatives from various industrial companies, active in different business areas, all working somewhat differently with digitalization. The purpose was to inquire about their approaches, philosophies and challenges in digitalization and the use of different data management strategies. These discussions offered invaluable insights into the dynamics of current industrial practices, from people deeply involved in the use and development of the techniques and practices of what lays the groundwork for industrial digitalization, and by extension, the implementation of Unified Namespaces.

These conversations started with the background of the company, the interviewee and their role in the organization. Later, querying about the culture of digitalization within the enterprise, especially among management and decision makers. Analyzing what KPIs the company tracks, how much focus is on data analysis compared to everyday, operational tasks. What software and data management architectures the business uses connecting OT to IT, through what communication protocols the flow of data is made possible, the things they are content with or would like to improve upon. Further diving into the level of automatization within the enterprise as a whole, or sites specifically, analyzing the level of currently implemented digital solutions. By the end of these conversations moving more into the nomenclature, the amount of collectable data, sharing of industrial information between sites or external partners, and the interviewees views and beliefs on future use cases. Finally, presenting Unified Namespace and getting into whether they are working with the concept, or else with something similar with a different definition. Possible positives, negatives and ways to measure its effects, to finish with their professional views on the concept.

Through this comprehensive examination, the purpose is to open a dialogue surrounding digitalization, real-life data management practices and analyze the need and attitude towards Unified Namespaces. Gaining an insight into how real-world companies, often with a long past through different industrial eras, work with these questions is essential to understand if one wants to estimate the future direction the industry is about to take.

4.3.1 Interview with Alfa Laval

Date: 14th of March 2024
Person: Måns Forsberg
Title: Team Manager of Automation & Digitalization within Operations Development Technology Office
Location: Lund/ digital interview

Alfa Laval is a global company with operations and products in mainly heat transfer, fluid handling and separation. The enterprise develops and manufactures a large variety of products, and is present in several sectors such as energy, food and marine industries. Måns Forsberg is manager for a rapidly growing team within Operations Development, working with digitalization of shopfloor operations. The headquarter is based in Lund, Sweden, where the team is working with development and deliveries of common applications and infrastructure to Alfa Laval’s factories. Today the team works with standardizing digitalization for about ten factories, but the goal is to expand further and deploy standard solutions in all of Alfa Laval's sites.

Alfa Laval is a customer-oriented company with a strong background and focus on innovations and product development rather than on standardizing operational structures across factories. However, Forsberg points out that Alfa Laval is now quickly standardizing and modernizing their OT infrastructure and applications in the factories, which in short is the mission of his team. Out of the 10 factories that his team works with today, Forsberg approximates that five or six are now using multiple common OT solutions, and the plan to continue in that direction is clear.

In all of Alfa Laval's factories, like in most factories within the industry, there is a mix of old and new machinery, ranging from highly automated equipment to manually operated machines without control units. According to Forsberg, the mix of differently aged machines and technical status is in general due to existing processes performing well, meaning that technical improvements in the factories are made gradually as issues arise or bottlenecks occur. PLC’s are generally installed where automatic control of machines is required and for more or less all newly installed machines. Due to the high complexity of production, automation in general is still more focused on machine or cell control, rather than on production line or flow control. Forsberg adds though that there are production lines and packaging lines that are fully automated, i.e. lines containing many PLC’s that cooperate.

The most common communication protocols between edge-devices are OPC or OPC-UA, but other communication protocols also occur as a result of the large variety in the machine park. Data management and structuring is partly based on ISA-95 where it has been applicable, but Alfa Laval does not follow it to the letter. Forsberg explains that a full transition to complying with the ISA-95 standard would be a major and complex undertaking. Data is currently extracted from PLC’s using standardized templates to ease the structuration of data, besides the actual extraction of data. The templates can be seen as models that are plugged into the PLC’s, in order to gain uniform data output. Forsberg says that these templates work for some specific PLC brands, meaning that adaptations are needed when factories purchase machinery with PLCs from other vendors. The extracted data is used to calculate KPI’s like Overall Equipment Effectiveness (OEE) and for manufacturing order reporting.

On the question of whether Forsberg thinks that it is worth extracting all data from a PLC, he hesitates. He reasons that there is limited use to extract all the data from a PLC but does believe that there is more valuable data to be extracted for analysis purposes. It is important to remember the primary purpose of a PLC, which is to control a machine, thus a lot of internal variables will only serve as control variables. Therefore, collected data usually requires transformation work to extract useful information, which might take away resources from other operations.

Even though there are lots of ongoing efforts to standardize and modernize the application landscape on the shopfloor, many factories within Alfa Laval still rely on homegrown applications, and lots of manual operations are still reported with pen and paper. This inconsistency stems from Alfa Laval having a wide variety of factories around the globe, many of them once purchased with their own applications on different levels in the organization, with their own technical solutions and machinery. Another reason for the relatively low degree of standardization across factories can be their unique characteristics. In general, sites with more repetitive production are characterized by further advancements and focus on digitalization, while other elements for the variety can be differences in culture, traditions, cost structures and priorities.

Forsberg also points out that a complicating factor in this ongoing standardization is to convince factories to switch from existing applications, that often work well, to standardized solutions. There are elements of risks, costs and change management that need to be accepted, and the transitions can be complex and time-consuming. However, he specifies that it is necessary, even critical, to standardize applications and data flows to reduce long-term costs that arise in operations. Staying with machinery and solutions too long will impose risks from an IT/OT security perspective and be at risk of losing vendor support, both in terms of competences and spare parts. By having a common OT landscape that is up to date, enterprise costs and risks can be minimized in the long term. High dependability on single vendors can nonetheless be a risk, and therefore the company needs to be careful when selecting new common solutions.

Despite current progress within data management, Alfa Laval does not currently have a UNS structure implemented but is moving in that direction, although it is not yet an outspoken aspiration. Forsberg was previously aware of the technology and believes that UNS can be useful for organizations. He mentions that he has previously built similar solutions, although they lacked the actual label of a Unified Namespace, while also having a key set of differences. The implementations back then were typically centered around a large database, acting as a single source of truth, from which connected applications had access to the same exact data in order to uniform data operations. While the implementations differ from the one in this report, it showcases the natural need for a single source of truth to achieve high efficiency within DataOps. However, these solutions were typically constructed for single factories, and constructing a solution for an entire organization like Alfa Laval is undoubtedly more complex. As Alfa Laval further develops digital capabilities, Forsberg sees the need for a standardized way of naming materials, batches, lots, products, machines and so on. He adds that having data available in a centralized repository would genuinely simplify current and future initiatives, such as mapping product and process flows, while at the same time easing implementation of future solution development.

Discussing the implementation of a Unified Namespace is not something that they currently do within the team per se, however he confesses a frustration stemming from the lack of unified nomenclature when handling data flows. The effects of UNS would in his opinion affect time and efficiency of projects and thereby save money in the long run. Alfa Laval already uses non-typical data in the ERP system, that is more closely related to MES to closely track production variables. Forsberg mentions that their production often includes elements of manual work and craftsmanship, which can be hard to capture automatically in applications. He believes that UNS can play an important role in creating a neutral representation of their production and product flows, mainly through standardized data modelling. The greatest benefit of UNS, according to Forsberg, lies in its ability to minimize time spent on development, data collection and searching for data, and thereby reducing the time needed for analysis and identifying opportunities.

Some limitations in Alfa Laval’s ongoing digitalization journey, expressed by Forsberg, are limited access to data from legacy machines, lack of professionals actively working within this area, and (still) a limited overall focus on the digitalization of shopfloor operations. To start implementing one or many UNS’s for Alfa Laval would most likely be a major undertaking, due to the variety of existing solutions, structures and priorities. Despite the challenges, Forsberg believes that Unified Namespaces would be a great potential for Alfa Laval. Standardizing data will eventually facilitate streamlining of operations, improve efficiency, and unlock valuable insights. While full implementation might be a long-term goal, taking some initial steps towards UNS will pave the way for a more digitally connected future for Alfa Laval.

4.3.2 Interview with an International Packaging Business

Date: 20th of March 2024
Person: Anonymous
Title: Director Operational Excellence
Location: Sweden / digital interview

This interview features a senior operational executive at an international packaging company, who preferred to stay anonymous. With several years of industry experience gained both domestically and internationally, he has worked with streamlining business processes to turn around struggling companies. Today, working with business operations on a larger scale inevitably includes engaging with questions of data management, data architecture and company level automation systems. Therefore, to improve internal and external processes, both he and the company are continuously working with developing data management, realizing the value gained by tracking and analyzing their operations.

The company boasts a diverse product portfolio, estimated at around 2,500 items with numerous variations. Despite this complexity, the interviewee believes that the current state of industrial digitalization is advanced, considering the prerequisites of the business. To remain competitive in Western Europe's highly automated industrial landscape, companies must prioritize process control. The next focus point of their enterprise-wide operations development program is to unify the data systems across the entire organization. As of now, different ERP software and PLC suppliers are used depending on the site, and they are in the process of phasing out older programming languages to be replaced by modern Siemens PLCs. His team is currently working on rolling out a common ERP system for sites that currently lack a structured infrastructure, using a machine compatible, open-source application called Aurora. Apart from unifying system architectures between sites, he would like to see real time tracking of production, to at any time see if and why a certain production line is at a standstill. Managing geographically dispersed sites, it would be advantageous to access both past and real time data from sites to better manage everyday operations. Apart from this, there are other advantages of closely tracking production in real time. He feels that some employees would surely share the feeling of achievement when completing a daily goal for the machine, or amount of parts produced, raising productivity. Instead of spending hours searching up and analyzing data, the production leaders could focus on interacting with their employees, to together create a more humane workplace leading to better productivity once again. A company culture where colleagues have a sense of community, are proud to accomplish goals together and go home feeling fulfilled is what all companies should be thriving towards.

Due to their specialized products, industrial data sharing between company sites is uncommon, except for those producing similar items. However, some data exchange does occur on a machine-to-machine basis, in other words consists of best practice information for specific events regarding specific machinery. The interviewee emphasizes that, in his experience, sites with less flexible product portfolios enjoy high data transparency across both human and automation-driven processes. Working with Key Performance Indicators (KPIs) in the forms of benchmarking, best practice and standard operations procedures is what will, in the long term, grant companies their competitive advantage. In his opinion, this will also allow the use of AI applications to contribute to operations, if the right input data is given to it. However, as of now they are working extensively analyzing the top five and bottom five events or processes that worked well, or poorly during weekly meetings with production leaders, which is entirely based on current data flows. Significant improvement remains, as the lack of data visualization necessitates manual, outdated analysis methods. This is where it would come in handy with a centralized program connected to the data streams from machinery, as today there is no existing application implemented where you are able to access all needed data and one would instead have to search in different software depending on what they are looking for. In the future, weekly production reports could even be summarized by AI instead of a person, freeing up even more time for employees on all levels.

In this industry, especially for packaging that is in contact with food, it is required that suppliers and manufacturers have severe traceability systems, to backtrack possible issues and complaints from customers. For the moment, the company owns a large variety of machinery from different manufacturers and eras. If a reclamation is received from a client, the level of back traceability to a specific batch or operator would depend on the actual equipment in question. As a result, an order or batch cannot consistently be traced back to the exact time and space to analyze what went wrong. However, this is possible using machines with a newer Siemens based driver that they aim to implement widely, which would allow the data to be tracked and stored. Rather than only storing data for the purpose of quality assurance, which would be for around a maximum of two years before it reaches the end-customer, it is stored for much longer. Some data collected from PLCs is stored well over five years and this data is used to find similar causations for production issues and to solve problems. This results in major time savings. In brand new machines and machines that have been repurposed and rebuilt, it is possible to single out production errors in batches and sometimes even products. This allows for close monitoring of production, early interference and smaller recalls or scrapping, if necessary. The manager says that machines are very expensive and therefore, rather than purchasing new equipment, it is usually favorable to upgrade existing machines. These upgrades are made to improve performance, but a common upgrade is to convert or install a PLC that provides a uniform data stream, compared to the other, less intelligent machines. He continues by saying that even though parts of the machine park are based on old machinery, they have been upgraded to be in the forefront of manufacturing technology. As a result of these upgrades, maintenance personnel and control engineers will, for example, be able to diagnose machines from home.

The interviewee had little to no knowledge of UNS prior to the interview. After a brief presentation he believes that it is the right direction to be heading towards. There is a genuine need for a centralized data repository, with easy access for users to retrieve desired information. He points out the need for existing software to help with the implementation, along with the proof of need and reasonable return of investment values, in order to get manager on board. In general, the company has relatively short decision-making time considering the large size of the business, where the general managers are quick to implement new ideas. The reason behind the lack of implementation of a modernized digital infrastructure is the lack of human capital and prioritizing other projects, such as refurbishing machinery and other physical infrastructure. Despite this, the company is now focusing more on digital infrastructure, as in order to keep their competitiveness in the coming 10-15 years, they have to implement AI solutions, etc., in their business. Currently, the primary discussions within the company surround robotization, data collection and real time data analysis, and what they can do to implement these to the best of their abilities.

When asked about the possible downsides of a UNS architecture, the answer is data breaches. The uniqueness of the company is precisely what allows them to have a competitive advantage and makes them the best in the world in what they do, he argues. Therefore, it is essential that access to operations data is strictly limited only to actors within the organization, and that it cannot be transferred outside, even with employees leaving the company. This does not only include competitors, but also customers and suppliers that would not be granted access to such a system, as there is simply too much valuable information they can gain in terms of bargaining power, for instance. The only exception would be to allow customers and suppliers to see production planning data, so they can then in turn adjust their output and plan their own production accordingly, which is something that he himself has implemented already in 2003-2004 when working for another company. To allow external actors to gain insight into one’s processes any further than this, however, would be unthinkable for the company, therefore a possible UNS implementation would be strictly internal.

While acknowledging the challenges, the interviewee concludes with a sense of optimism regarding a potential UNS implementation. He envisions a future where a centralized data repository empowers real-time production analysis, leading to faster decision-making, improved employee engagement, and ultimately, increased competitiveness. He believes overcoming the initial hurdles of implementation will pave the way for a more efficient and data-driven future for the company.

4.3.3 Interview with SKF

Date: 23rd of March 2024
Person: Jan Ek
Title: Technical Strategy Lead (Manufacturing)
Location: Gothenburg / physical interview

SKF was founded in 1907 in Gothenburg and are the inventors of the spherical ball bearing. Today, they are market leaders in bearings and a global industrial giant with more than 77 factories worldwide and over 40.000 employees. In the end of 2019 and beginning of 2020 SKF presented their concept and team, future factory, which would lead the transformation towards a data – and AI driven production environment. Since then, it has evolved to be the norm for production within Gothenburg and is nowadays just called manufacturing development.

Jan Ek has worked for SKF for over thirty years and is, together with his team, leading the digital transformation for the enterprise. Over the past six years he and two others have implemented and transformed data management and systems within production and management. Initially the focus of SKF has been on developing the OT-side, with little focus on IT. Despite starting with an OT focus, SKF Gothenburg actively pursues digital transformation through continuous improvement. Jan Ek exemplifies this by saying he previously worked mainly with automation, but automation goes hand in hand with good data practices and a need for developed data management. As the team works according to a continuous improvement philosophy, they have had to restart the project from scratch a couple of times. This is due to various reasons but mainly because of the need for scalability and new learnings, which Ek finds trivial for the final development. The fact that they might need to restart from scratch was clear from the beginning of the project, and with supportive management behind the team it went smoothly every time. By following the philosophy of gradual improvement, they are able to build scalable solutions and automate as much as possible through coding, which ultimately keeps costs down, limits the team size working with it and makes it more agile. Jan Ek describes their journey as a staircase and that they now are on the way, or already passed, the first step which is the visualization of data. Next step is implementing predictive solutions based on the data, i.e. predictive maintenance, among other things. Predictive solutions in combination with AI will likely lead to full autonomy within the production. However, this is something for the future, as they are not fully there yet. Jan Ek says, “When we develop these solutions, and beyond, is where we become truly profitable”.

As of right now, Jan Ek is working with giving users the right tools and data to be able to visualize operations. They have, for example, eliminated the need for a SCADA level in their organization by being able to visualize operations and data in real time. Operators within the production use tablets to control and visualize the machine and are usually responsible for multiple machines simultaneously. Fault codes, production errors and so on can be seen directly from that tablet or screens posted in the production site. Instead of building custom HMIs for each team and purpose, the technical data team connects everything backend and provides users with the toolset of setting up their own interface in a drag-and-drop manner. This allows operators to build their own interface for machines and they can customize them as they like. Outsourcing allows them to focus on the backend systems, while recognizing that each team has the most knowledge of suitable interfaces.

Despite being a global company, SKF factories, according to Jan Ek, operate within their own cell with different machinery, software and methods. In Gothenburg, SKF mainly uses Kepware and Brownfield Connectivity (for Siemens products) to extract data from PLCs. They also use HighByte's Intelligence Hub to model and add context to data and Ditto for a digital twin and real-time visualization. Lastly, SKF uses Grafana for visualization sourced from both real-time and historical data, but also allowing users to create their custom interfaces. SKF also uses various Azure solutions, like cloud and AI. Within the Gothenburg facility the communication protocols from machines consist mainly of Modbus, OPC-UA, REST, HTTP, Web socket and to some extent MQTT. Jan Ek adds that he and SKF avoid using the MQTT Sparkplug B protocol, since he experiences it as too limited for manufacturing. He continues by saying that there are many legacy ad hoc systems from the 90s that they still take into consideration when implementing new data management solutions. Through Intelligence Hub it is possible to model, transform - and connect data flows so that most systems can access the data they need. Jan Ek continues by saying that it is necessary to have KepServerEX connected directly to the PLCs, as this way it is possible to create many instances of the same machine. With Modbus, for instance, it is usually only possible to have one machine instance. Intelligence Hub is necessary as the PLC only needs to send data in one data stream, as opposed to having several connections for every application wanting to access it. Even if a PLC uses OPC-UA, which is considered more scalable than more traditional communication protocols, it can only send approximately 100 parameters before getting overloaded. After data is distributed to Intelligence Hub it is then transferred into multiple systems (in a similar manner as in the implementation example of this report), thus eliminating overloaded PLCs. SKF also uses two or three ERP systems, which Jan Ek would like to see consolidated into one going forward.

SKF factories display a wide range of digitalization maturity due to a lack of standardized procedures across the group. For instance, the Future Factory project was never intended for Gothenburg, since they were already beyond that stage. Some factories within the group still use pen and paper and are not even connected to the network, while others (like the factory in Gothenburg) are highly autonomous. Jan Ek elaborates by saying that there are around thirty factories, of the 77, that are worth connecting to the network. In Gothenburg alone, there are many machines from the 80s or earlier lacking PLC and network connection. Of the approximately 1000 pieces of equipment at the site, only 500 are ever worth connecting to the network, due to a lack of business benefits of the others. A major difference between other factories and the one in Gothenburg is that other factories do not use HighByte's Intelligence Hub to handle data flows. According to Jan Ek, the other systems that SKF has developed lack some features but are more user-friendly, and therefore the user rarely has to code. As there is no benefit of sharing data among factories within the group, data sharing is limited. He elaborates that there is no benefit for the factory in Gothenburg to access data from another factory, as they work with different products. The only benefit would be for remanufacturing of a bearing that is produced elsewhere, and vice versa.

Different types of data have different procedures for how it should be stored within the Gothenburg factory and organization. Due to data already having context (which is achieved with HighByte's Intelligence Hub in Gothenburg) it is possible to filter out data that is undesirable. Jan Ek is also open to the idea of one day selling their production data to other companies to develop AI models or use for analytic purposes. However, the data is not yet of high enough quality for commercialization, i.e. it is not given enough context or standardized adequately. Today, data is loosely structured according to ISA-95 and ISA88, as they do not follow it to the letter, due to the data only being consumed locally and not globally. Within the Intelligence Hub software, SKF standardizes data by using the same (functional)model for different machines, creating instances based on these pre-defined models. This results in the transformed data looking the same for different machines, and from an outside perspective all machines would look the same. Jan Ek exemplifies this with the color coded lightbars on their machines that they use internally. When representing the status of machines, they use color codes such as green, yellow or red to indicate machine status in a uniform manner. This would both be shown graphically, but also within the backend data streams to ease tracking. These color codes could mean somewhat different things as machines are from different decades, for instance yellow meaning that there is some warning that does not yet impair the machine but does not specify what exactly. By leveraging that highly useful data and transforming it to a uniform manner, it is easy to show if a machine is running or not, and everyone across the organization knows how the machine is performing. This is also used in the aforementioned interface for operators, where they use this to oversee their machinery in an intuitive manner. Since SKF has actively been working with structuring data, the obstacles caused by unstructured data have decreased, and finding desired information has become easier. Through their journey they have learned that it is critical to standardize from the core and avoid transformation tables for parameters as much as possible. This is due to the unnecessary added level of complexity. Jan Ek does however mention that there is some technical debt in the work that they did five years ago, although the main issue is still the lack of standardization within different systems.

Today, SKF enables a Large Language Model, or LLM, for analyzing historic data tracing back decades. Ek shows a pilot program where the AI can search through old maintenance notes, find common failures and how to fix them. The data is extracted from a database that the company has built up over the years. It is already used by maintenance personnel when encountering issues they have no previous experience with, effectively searching through the notes of their previous coworkers. The AI model can answer questions, make summaries and even give suggestions on what can be done to solve the issue, shortening the problem-solving process drastically. By enabling interface and data translation through AI, SKF aims to significantly reduce maintenance time and accelerate the learning process for new employees. As a fun fact, he mentions that the first beta version took half a day’s work to create.

Jan Ek argues that extracting all data from PLCs is inefficient due to the presence of unnecessary data. However, he sees potential for extracting more relevant data by implementing data modeling techniques. He exemplifies this by saying that 95% of data is slow data, updated every second, and the rest is fast data, updated in nanoseconds. Even though it would be possible to extract data in nanoseconds, you rarely benefit from it, and when you do it is highly specific and is not relevant for the process. Jan Ek believes that one solution is to let the AI determine what data is relevant on its own, instead of saving all data from PLC.

Jan Ek explains that they have implemented a UNS like structure in their production already but do not call it a Unified Namespace. For example, there are some architectures that are implemented alongside the data flow to Intelligence Hub, meaning not precisely everything is centralized to an MQTT broker. Ek continues to describe a UNS structure built not according to the Schultz method but instead built around functionality. He says that it is necessary for his team and the organization to build a UNS that relates to all different functions of the organization. This can be seen as the maintenance team having their own UNS and the supply chain having theirs and so on. The digitalization team looks at what is the smallest common denominator for the different systems, the machines, and standardizes the nomenclature for these. Instead of saving data in terms of topic structure, it is obtained based on use case. Therefore, their structure, according to Jan Ek, differentiates somewhat from the widely accepted MQTT based UNS structure following ISA 95. This is due to the data transferred and consumed being limited to Gothenburg, thus only using different factory buildings and machines. There is however an ISA-95 structure of the same data in databases, and therefore it can easily be converted to it if necessary. However, most data is still converted into the MQTT communication protocol for future applications, as a consequence of the topic structure and how easily a new software or application can be subscribed to it. The main difference since implementing their UNS structure according to Jan Ek, is that data becomes easier to manage, mainly due to it being standardized.

SKF currently lacks a concrete parameter for evaluating the digital transformation, and instead looks at internal satisfaction. This is mainly measured through complaints, numbers of active users and informal feedback. However, there is notable value in the transformation for users. For example, somebody working within the supply chain saves time by finding information more easily, and therefore also the company saves money. Another more abstract value is that an operator feels fulfillment, knowing that they completed all their tasks, as they can now easily track their own progress. As the digital transformation continues and the predictive tools are implemented, it is likely that parameters like savings, scrap and indicators of shortenings in downtime will be used. Jan Ek elaborates by saying that SKF in Gothenburg is focusing on automation, as it is a high-cost environment, and therefore it is essential to automate the data flow as well. To be able to do so, data needs context and must be somewhat standardized. The goal of SKF is to build an end-to-end data solution where the product can be traced from the source mine all the way to it being recycled. Unfortunately, most suppliers are not mature enough today to be able to share data with SKF, Jan Ek informs. However, SKF and Ek are positive to opening their UNS/data for suppliers and customers that they have close relationships with. When asked if there are any security concerns, he answers that he does not have any, as it is always possible to set up security.

Another reason for Jan Ek having to restart their project, except for scalability and learnings, is because they had to Figure out what they wanted their UNS to be. He says that this is a huge barrier, as there is no one-fits-all-solution and it is, in most cases, necessary to find one by trial and error. In the case of SKF they had to restart after merely 2 months and then again after half a year. Jan Ek says that it is not impossible that SKF has to restart the project again, as it is still evolving. However, the next implementation of the system will be faster as most things are standardized, and the team has codified/automated much of the process. Although the sheer quantity will likely make the process longer. He estimates that it would only take three months to switch the entire system. Jan Ek sees no downsides with UNS as long as it has some flexibility. If companies are allowed flexibility with ISA-95, the standardization and the implementation parts of it, it is a useful and perhaps even essential tool. He continues by saying that the necessity for ISA-95 becomes larger if the UNS is intended for multiple factory sites. Another example of the necessity for flexibility is when purchasing factories with existing digital infrastructure, where the integration process can become difficult. The world is also dynamic, which adds another layer of complexity if there is no flexibility.

Jan Ek's focus on continuous improvement throughout the interview suggests that SKF Gothenburg's digital transformation is far from over. SKF Gothenburg does, however, present a compelling example of a data-driven manufacturing transformation journey. Digital transformation within SKF emphasizes the importance of a user-centric approach, the power of data standardization, and the potential of AI to unlock hidden value within industrial processes. Despite the progress outlined by Jan Ek, challenges persist as integrating a global network with varying digital maturity is a significant hurdle.

4.3.4 Interview with Stena Recycling International

Date: 5th of April 2024
Person: Åsa Fast-Berglund
Title: Development Portfolio Manager
Location: Lund / digital interview

The following interview was conducted with Åsa Fast-Berglund, Development Portfolio Manager at Stena Recycling International. Stena Recycling is one of Europe’s leading companies within recycling and circular economy solutions, recycling residual waste from industrial and civil applications. Some of these industries include manufacturing, transportation, energy and logistics. The company is part of a concern called Stena Metall, that has several other branches focusing on areas such as aluminum, steel or oil. With several sites ranging greatly in size scattered around Sweden and Europe, they are truly one of the key players within creating circular economies at scale.

Åsa Fast-Berglund has a background in both industry and academia, including previously working as professor in smart automation production systems, as COO for Stena Innovation Lab and as Program Director for

Mechanical Engineering at Chalmers Technical University. Furthermore, she has been active as a member of the Swedish Standards Institute (SIS), served as Vice Chairman of the Internation Federation of Automatic Control and is currently Board Member of the Nordic MTM Association. As of now, she is leading a team focusing on technical development within both production and the organization, consisting of people based in seven different countries. Her team is mostly working with Stena Recycling, but their work also branches out to other subsidiaries within the Stena Metall group.

Stena Recycling’s leadership is positive when it comes to implementing digital transformations. Around the time she transitioned into her current position at the company is when they started implementing ISA-95, and since then, they have been advancing in the level of digitalization and data management. The team always tries to quantify key Figures, such as return of investment (ROI), to measure business benefits of digitalizing. To further strengthen their argument towards management, they would often present a business case to represent the problem a new solution solves. However, she admits that work focusing on backend, within structuring and data management, is something that is hard to show in ROIs, but still necessary. If in-house users can access and make sense of data fast and efficiently, it allows them to focus on using the data instead of analyzing it. Data management, therefore, consists of mostly time savings, reduction of human errors and of mismanagement.

Stena Recycling stores industrial data in databases, data lakes, and Microsoft Azure. Manufacturing data is retained for two years for analysis before automatic deletion. When analyzing production data, the company focuses on monitoring downtimes and prevention of future downtimes, Overall Equipment Effectiveness (OEE), energy use and CO2 emissions among other parameters. Most of the analysis is currently done in Microsoft Excel, although PTC ThingWorx is also implemented as their IoT software. One of the biggest challenges is that the analysis is heavily knowledge and experience dependent, meaning to gain valuable insights stemming from the analysis, one needs a detailed knowledge of the processes. For instance, adjusting the processing speed for incoming recycled materials relies on operator experience, not historical data analysis.

As mentioned, Stena Recycling is using PTC ThingWorx as their preferred IoT platform, where they also have virtual 3D models of the machines. Other software the company use include Microsoft Azure for cloud applications and Microsoft's ERP program, namely Dynamics 365. For the SCADA system, they use SIMATIC SCADA developed by Siemens, and for communication between PLCs, Stena Recycling implemented PTC Kepware as middleware, which integrates well with ThingWorx. Apart from the industrial machinery, the company also owns laboratory equipment and machinery, which tend to use specific communication protocols interpreted by Kepware. Most of the communication between devices is through OPC-UA, as Siemens manufactured equipment comes with the protocol already integrated. Some machines also use MQTT, although these are an absolute minority. Fast-Berglund however points out that as most devices lack the integration for MQTT, it is significantly more troublesome to implement. The protocol needs third-party software responsible for parsing and translating, whereas the OPC-UA implementation can be integrated directly and therefore substantially faster. For IoT communication, mostly between sensors added to older machinery, Rest API is used, which Berglund finds simple to use. Currently, communication between the connected devices is one way, meaning data is only collected and devices do not receive information for control purposes. One reason is that this is simply harder to implement, therefore the company prioritizes collection. The second reason is the safety aspect, as controlling these machines is complex, and hence is done by experienced operators through the SCADA interface.

Most machines, in the production, have a PLC that is connected to the IoT network. The majority of the PLC data is extracted and later filtered by a template specific to that machine within the IoT system, before finally getting saved for later use. To older, mechanical machines, the team has added vibration sensors that can signal if a machine is running or not. Some mobile machines such as wheel loaders are sparsely connected to the network, however, in the future might also receive data from a GPS tracker to know their exact movement. Apart from this, the next step for the company is to implement ML models, to make use of their collected data mostly for different predictive purposes. According to Berglund, many control processes still rely on manual inputs, of which she would like to decrease the reliance on experienced professionals and instead use their available data to automate operations.

When it comes to selling data, Åsa Fast-Berglund expects Stena Recycling's manufacturing data to become even more valuable than it is today. As circular economy solutions become even more relevant, there is lot to gain from data sharing between companies involved in production, product development and recycling. End-of-lifecycle data is not only important for other recycling businesses but is something that can be incorporated during the whole lifecycle of a product. Stena Recycling can provide upstream developers with exact information of how much of the product is recycled, how old they are and in what condition did they get thrown away, which can be incorporated in the design process for future products. However, data sharing is currently quite limited, even within the company, but is something that they are actively working with. For instance, Stena Recycling has several sites in Sweden specializing in battery recycling, where one site could make great use of another site's data, especially for optimization and ML purposes. They are already sharing data with business partners such as Volvo and Electrolux, as these companies require follow-up and traceability for their product parts. A possible concern mentioned by Berglund is the importance of cyber security, especially regarding the connection between IT and OT. Much of the transferred data is sensitive, therefore the company has allocated great focus on building an infrastructure that minimizes the risk of data breaches. As a consequence, the team has segmented the machine park into virtual units inside the network architecture, reducing the repercussions caused by a possible breach.

Discussing Unified Namespace, she and her team analyzed whether it should be something worth implementing officially. However, in the end they ended up deciding against it, for several reasons. One reason being the sufficient level of standardization in nomenclature already actively implementing ISA-95. Another reason is the ample amount of context given in data flows, at least for the use cases currently in use. She continues by saying that there is certainly value in UNS, but mostly in cases with great amounts of data and with clear flows all the way between PLCs to ERP systems, which is just somewhat the case for Stena Recycling. Kepware and ThingWorx are sold by the same developer, meaning compatibility between programs has not been an issue. Apart from this, they are, at the moment, mostly using the data internally, within the site, which has been functioning well. Contextualization and translation between protocols is, moreover, currently achieved through ThingWorx. Overall, the idea of a Unified Namespace is something she looks at positively, especially the structuring and the use of ISA-95. The MQTT protocol however is suboptimal, as the implementation is too much of a hassle as opposed to simply sticking with OPC-UA, according to the professional.

Although not outspokenly having implemented a Unified Namespace, Åsa Fast-Berglund and her team are without question actively using several important parts of the concept. The rigorous use of ISA-95, contextualization of data, translation between different communication protocols and a central platform to which most systems connect are all important aspects of the approach. For Stena Recycling, the missing link seems to be a publish/subscribe protocol, however that is not something they are looking to implement in the near future. The company plans to continue improving their data management practices, realizing the business value in actively working with improving their existing processes. Decreasing the reliance on experienced operators will most likely demand yet better contextualization, deeper data analysis and higher levels of automation.

4.4 Interview with a Software Supplier

4.4.1 Interview with HighByte

Date: 18th of March 2024
Person: John Harrington
Title: Co-founder and Chief Product Officer
Location: USA - Sweden / digital interview

The following interview was with John Harrington, co-founder and Chief Product Officer of HighByte, focusing on product management at the company. HighByte is an Industrial DataOps business, among other things responsible for the development of the Intelligence Hub software that was used in the demonstration of a UNS implementation for this report.

John Harrington has a background in mechanical engineering, software engineering and business, with time spent working for firms active in manufacturing for several years before co-founding HighByte. Together with the company's other two co-founders, current CEO Tony Paine and current CCO Torey Penrod-Cambra, they identified the industry's need for a new way of handling industry data that is in line with what companies implementing Industry 4.0 solutions need. This need, he explains, evolved from several things coming together at once. General sensors becoming extremely cheap, networks becoming ubiquitous, the maturing of cloud platforms and data analytics, and more recently enormous leaps in AI. Despite these advances in technology, the trio noticed the inoperability of industrial data for people that want to use factory data outside of the factory itself. Data for teams in supply chain, maintenance, R&D and management that do not necessarily understand the factory, and certainly not the implemented control systems that are in use, which are outputting semi structured, uncontextualized data for a different target use case. In his experience, getting access to the data was never usually the problem, but rather contextualizing it and therefore making it usable for other purposes than control engineering. Traditional industrial software companies leveraging entire ecosystems lack the incentive to create solutions that can communicate between vendors, as they rather want to lock users onto their platform, he explains. Instead of consolidating data usage, the key is to deliver data in forms that IT systems can make use of. In the future, there will be more smart data networks with nodes spread out all over the ecosystem, consisting of nodes producing or receiving data, increasing the value in delivering the data rather than being one of those nodes.

The industrial sector is inherently slow to adapt to change, with machinery that has been in use for 10, 20 or 50 years. Despite excitement about new data implementations, many existing industrial companies lack the know-how to be data driven. According to Harrington, they simply try to keep up with production, get the products out of the door, and lack the expertise to properly analyze their data in this new age. If companies are not automated with controls in MES, or SCADA, they are surely not ready for jumping to Industry 4.0 either.

Harrington says that UNS is an ever-evolving concept. The definition and message that Walker Reynold gave just a few years ago has changed in the past years and will probably continue to do so. He exemplifies with SparkplugB, which used to be one of the main communication protocols for UNS, but today he would not recommend use of it any higher than SCADA level. Companies often struggle with SparkplugB at higher levels because it limits data context compared to pure MQTT, which offers more flexibility. He continues, in some cases, especially below the SCADA level, OPC-UA is a powerful protocol and has many advantages compared to others. For his customers he still recommends OPC-UA for a high-speed and control-oriented supervisory level, i.e. the SCADA level. According to Harrington, the UNS should be above the SCADA level and pull data from the MES, ERP and so on, but also from sensors and PLCs. However, the purpose of data collected at SCADA level or below should also be for analysis -, supervisory - and persona use cases, namely for management purposes and not only to control processes. Harrington also mentions that he would like to see the conversation start to move away from building a UNS, to the applications that are consuming the data. He illustrates the problem by describing how MQTT functions. The pub/sub protocols that MQTT offer are great, but less than ideal for applications that need historical data. Essentially, there are not a lot of applications that can consume pure MQTT as of now, and therefore it is necessary to save data in a data lake, database, or something else to be able to access historical data. Widely used analytics software, such as Power BI, Grafana or Tableau need to query a database first, which complicates the architecture somewhat. Therefore, as MQTT is not the ideal protocol for all existing use cases, it should be used when multiple consumers are subscribed to data and need different parameters of data. It should not be used to simply pipe all the data flows together, as in control - or analytics applications this could even be counterproductive.

Harrington does not believe that there will be a new market segment where industrial data is sold, similar to oil. He elaborates by saying that data certainly is valuable and there will be some exchange of data if companies are not able to create their own AI models or analysis tools. This enables companies to focus on their core capabilities, which is positive. Harrington also believes that there will be companies that make AI models as their core business model, using training data from outside industrial sources. These companies will use data from manufacturing companies to train models, to later sell those models back to the same companies or to others. Rather than them buying the data through one platform or marketplace, he believes that it will be a contractual obligation, where a one-time solution is created, or a discount is offered for companies supplying data. The analytics company will be able to implement and train their software, while the industrial company supplies data and reaps the rewards from increased analytics, making it a win-win. He problematizes by wondering how effective those models will be as industries and factories vary extensively, which is another argument for why there would not be one single marketplace that companies use to sell their data. Another aspect is secrecy; therefore, the data will likely be from processes that are not critical for that sector or business.

All data should not be extracted from the PLC, as much of it cannot be used and it is expensive to store, according to Harrington. There are important datasets not currently extracted in most cases that could be valuable, and a company should aim to have these in their UNS. However, there are many data points that essentially say the same thing in different ways, and these should not be included. It is imperative that data extraction is modeled, as extracting all of it is overwhelming. This is one reason why Harrington says that UNS does not elevate the need for all specialists currently working in a production. It does, however, reduce the amount needed and makes training of new specialists easier.

Specifically talking about HighByte's Intelligence Hub, but also UNS, it provides four different concepts.

1. *Data orchestration* - is about managing flows of data, i.e., transforming data and extracting it from one location and distributing it to others.
2. *Data quality* - Ensuring data quality is that the data is accurate and complete. This can be achieved by having data in real-time, establishing standards or ensuring data transfers are consistent and without data errors or missing segments.
3. *Data observability* - Like data quality, observability is achieved by having the correct data broadcasted, at the correct time, and to the right person.
4. *Data governance* - Governance of data is about creating a set of rules and standards for data. Who has access to data, how can it be used and who is responsible. It can also be about ensuring that an organization follows the regulation of data, for example how long it has to be stored.

When Harrington is asked what value HighByte and UNS offer for an organization, he answers that there are two types of values. There is an enterprise value that is achieved through digital transformation, that increases agile business decisions, improves factory uptime, reduces scrap, enhances quality and boosts the yield of production cells. The second type is the IT/OT value, as it allows businesses to deploy projects faster with reduced maintenance costs, cutting the need for resources while at the same time providing higher quality data. When evaluating enterprise value, one main parameter he focuses on is the time saved, both by implementing a UNS system instead of another system, and time spent managing it after implementation. Projects can be rolled out across a wider range of equipment, systems and sites with less resources. While the overall added value is high, it is difficult to pinpoint one single component, rather than looking at the big picture. Therefore, enterprise value can be described as shared value.

There is also value on the OT/IT side as it streamlines processes and acts as a bridge between the two. UNS can simplify projects, allow for faster reaction times, and thereby decrease costs. One of HighByte’s customers could earlier perform one predictive maintenance project in a year, he mentions. After having implemented Intelligence Hub, they could get the same number of machines up and running in 30 days, and now have a platform to manage future changes. A common issue many customers of the company face is the lack of technical resources combined with complicated legacy solutions. Another customer could reduce the number of integration software and approaches by a factor of five, which means they could install and manage fewer technologies, boosting the learning curve and reducing infrastructure costs.

He continues by saying that UNS also adds values to AI, as without contextualized data AI is useless. If a company wants to use AI for business optimization, they should implement a UNS like structure regardless of what they call it, as AI is only as good as the data it receives. UNS will also add value for the supply chain as it is possible to integrate suppliers and customers in your processes. Another benefit he points out is that for larger companies with many alarms or defects it is easier to analyze the cause of an incident and trace it back in real-time. Something that Harrington also mentions is that if a company hasn’t reached at least Industry 3, it is going to take a lot of work to implement a UNS system. While the added values of UNS still remain the same, the costs are going to be very high, and thus the overall economic performance worse. Therefore, if a company is not already actively working with digitalization, it would probably be too expensive to implement both Intelligence Hub and UNS.

In conclusion, John Harrington emphasizes the need for a data-driven approach in manufacturing. He highlights the challenges of data context, interoperability, and talent gaps. His vision of a Unified Namespace structure offers a potential solution for these challenges, promoting collaboration and unlocking the true value of industrial data. While Harrington acknowledges cost reduction as a potential benefit of UNS, he emphasizes its broader value. UNS facilitates collaboration within the supply chain, improves traceability for larger companies, and empowers data-driven decision making across various departments. John Harrington is one of the key Figures within industrial digitalization space and has been forming both public opinion and the direction that Unified Namespace has been taking these past years. Together with his colleagues, they have been effectively using their combined industrial experiences to create a product that is growing rapidly. The Intelligence Hub is a software that was central to the Unified Namespace demonstration for this report, but also a program that Novotek Sverige AB are exclusive resellers of in Sweden.

4.5 Summary of Interviews

Based on the conducted interviews, it is possible to draw several parallels regarding the digitalization, data management and DataOps practices of the participating companies. All the businesses asked actively work with these questions and is something they view as vital to improve upon in order to stay competitive in the long term. Each company demonstrates tailored approaches to digitalization based on their unique industry needs. The mission of becoming more data-driven is almost generally agreed upon, however there are still considerable differences in how these companies picture themselves completing these transformations.

Main differences can be spotted between the work culture, primarily with regards to how they face failures. Some of the companies interviewed could be seen as using a more traditional, industry-typical approach. They are generally afraid to change processes that are proven to work, as taking a leap of digitalization-faith is viewed risky and is therefore done carefully. Meanwhile, other companies have more of a learning approach, recognizing that implementing new and innovative technologies will certainly give room for failures. This is where the support from management is crucial, as without the true commitment of company leadership, any greater infrastructural development is doomed to fail. One needs to realize that this is a continuous process, with a lot of hard-to-measure benefits that are hardly broken down into KPIs. Most companies therefore do not have a single universal unit of measure but focus on total time saved when searching and analyzing data, reduction of human errors and software license costs.

Another main difference between the researched companies is the level of current automation, the amount of implemented data management solutions and the use of distinct specific software. Communication protocols they use include OPC, OPC-UA, MQTT, Modbus and Rest API in different ratios. There is a clear tendency towards using more standardized and interoperable protocols, such as OPC-UA. Software choices, as for ERP systems, IoT or cloud platforms tend to be from the same developer, if possible, most likely to ease integration. For OT connectivity, all responders use KepServerEX developed by PTC, which they describe as user friendly and useful. For ERP systems and MES applications, there seems to be a larger variety including open-source software.

All of the industrial companies interviewed operate machinery of different ages and complexity, where a large part of the machine park consists of mechanical instruments, many times lacking even a PLC or basic sensors. No company representative considers exporting all PLC data to be reasonable, as it gets overwhelming without adding much value. For sorting out useful information, companies commonly use *templates* used as filters put on top of a data stream, designed by a system specialist. In the future, the interviewees see AI models as a possible alternative solution, where it would be in charge of the analysis and simply present its findings to a human. Decreasing the need for specialists is another important aspect that these ML models may aid with, which companies value highly. Furthermore, respondents agree that good quality data is key to train useful models, ease the work of human colleagues and shorten the overall time spent on analysis and therefore ultimately reduce costs.

There are great differences between the companies asked regarding their openness to share industry data with suppliers, customers or selling it to competitors. Opinions range from only sharing data for production scheduling purposes, exclusively with trusted suppliers, to opening up their whole system and selling most data freely. Usually however, industrial companies prefer gaining an edge from sharing their own data, either by financial compensation, or by collaboration benefits. For instance, being granted access to the AI model that was trained on their data. They would only consider buying data from outside sources if there was a specific use case for it in their operations. Similarly, sharing data within sites in the same enterprise is only deemed useful if they work with the same processes, using the same type of machinery. Sites within effectively researched companies seem to have minimal industrial data sharing and tend to instead function as separate units with their own infrastructure, practices and processes. Despite the tendency, for example Alfa Laval and Stena Recycling lately do intend to standardize and share practices between sites, trying to use similar data architecture regardless of geographical location. It can be stated, however, that among these exists a huge variety in the level of digitalization between sites within the same company, where some use IoT platforms and AI while others use pen and paper.

Despite operating in different industries, every company is either looking for, or already has implemented real-time production tracking software. This is usually made possible by an IoT platform and/or a visualization software that stakeholders can use. The point is to ease supervision and tracking, as the same software usually can be utilized by several users, such as operators, maintenance and management.

Nomenclature of production variables are commonly a source of frustration for the companies asked. The most frequent solution is based on ISA-95, that presents a very intuitive terminology to name datapoints. Usually, however, the ISA-95 standard seems to be something companies rather strive towards and is not something they follow rigorously. Frequently, companies would aim to unify nomenclature systems across sites, commonly using some kind of contextualization. Firms that have implemented some sort of standard earlier in their digitalization development seem to think of it naturally, while facing less difficulties backtracking information and implementing new systems.

Each of the responding companies works with the foundational goals of UNS. Alfa Laval claims to be actively working towards the concept, while the international packaging business mostly just acknowledges the value of it. They are, however, striving towards a centralized repositor of data, where all data points can be easily found and interpreted. SKF has come the furthest in their outspoken UNS implementation, with a fully functioning version on their Gothenburg site. Stena Recycling is working thoroughly with key concepts that make up Unified Namespace, however without calling it officially that, primarily as they are skeptical of the MQTT protocol. The exact definition of UNS is something still evolving, although the basic concept of having a single source of reliable truth is something that companies have been striving towards for decades.

When possibly implementing a UNS, asked company representatives seem to be overall positive, while still having doubts about some aspects. Forsberg at Alfa Laval is concerned about the newness and implementation costs, realizing the high resource allocation required. The international packaging business is careful about external access to operational data, as having everything consolidated and contextualized exposes the company to possible data breaches. For this reason, they would strictly use a UNS internally. SKF already implemented the concept, with focus on scalability, user-centric design and architectural flexibility as crucial for future profitability. Stena Recycling has decided against implementing an official Unified Namespace, citing sufficient current levels of standardization and data contextualization. Instead, they remain open to revisiting this stance as their business and the concept matures.

HighByte aims to resolve the industry's data management challenges, particularly issues of interoperability and contextualization. Traditional industrial setups struggle with interoperability and are locked in a developer-specific system that hinders broader use of shop floor data. COO John Harrington describes UNS as an ever-evolving concept, moving towards more focus on flexibility using communication protocols such as MQTT above, and mostly OPC-UA on and below the SCADA-level. He highlights the dual value that UNS offers to organizations. Firstly, there is enterprise value to be gained as more agile business decisions, improved factory uptimes, reduced waste and better quality. UNS implementation saves time in system management and deployment, enhancing the overall business value without pinpointing one single benefit. Secondly, there is IT/OT value to be gained in faster project deployment and reduced maintenance costs. It helps with reducing complexity among software integrations, lowering operational costs and easing the adoption of new machines and software.

5. Discussion

5.1 Direction of Future Industrial Development

Implementing a UNS in an organization seems like a natural step as a company evolves digitally, implements more automated processes and wishes to develop future Industrial AI capabilities. As industries advance from Industry 3 to Industry 4.0, the connections between physical and digital infrastructure become increasingly complex, and if no centralization is made the multiple connecting architecture can become as disordered as a plate of spaghetti. The introduction of UNS does not only decrease the need for specialists that understand the implemented ad hoc solutions, but mainly affects the overall time spent searching for, understanding and analyzing data. Upgrading infrastructure, finding specific data and adding new capabilities are a few examples of operations that take longer time without appropriate data management practices. All companies interviewed are currently working with data management, implementing standards based on ISA-95, while seeking and implementing new software that can simplify dataflows within the company. Seemingly all companies have realized the need for uniform data, with clear traceability and contextualization. They have committed resources for development of their data management, even though there is no one Unit of Measurement for the prognosis on Return of Investment for commitments in the back end, but rather mostly soft benefits. This can be an indication of how vital the implementation and development of digital capabilities are for companies to operate profitably.

With more OT data being treated in IT-heavy environments, there will be more value in the systems leveraging this transformation. OT/IT software companies are focusing to become the middlemen leveraging data, exploiting the fact that the industry is still very unconsolidated regarding technology and operational practices. Industrial data will become ever more valuable, and industrial businesses are realizing that their manufactured products are not the only valuable element that they create.

5.2 Viability of the UNS Example Simulation

As described in section 2.3.3., the simulation, or example implementation was performed in order to gather information on how a realization could be performed, with possible advantages and disadvantages. It was a rather simplistic showcased example, with only one complete production line containing three machines. Despite this, the demonstration was still highly functional and enough to be used to draw conclusions on what software, what communication protocols and connections a company looking to implement UNS might use. This will be discussed further in section 5.3. The intention was to choose software that often appear in the industry, mixing those that Novotek uses or promotes with those that they knew less of.

Furthermore, the simulation aimed to help understand whether UNS can benefit business operations and data management. It is deemed that the concept can help ease data management, as using the intermediary software to model, contextualize and standardize input data makes it easier to manage, as it can be organized easier and be used for decision making. To truly evaluate the effect UNS has on business operations, however, one would have to gather more data from using the concept day-to-day, apart from the example implementation. Hopefully, this is something that the conducted interviews can give insight to, but more information is subject to future research.

5.3 Implementing Unified Namespace for Industrial Companies

UNS is a reasonable and promising concept of achieving contextualization, transformation, consolidation and distribution of data. As depicted in Figure 23 below, instead of thinking of Unified Namespace as a single technology, it should be seen as a set of guidelines for structuring data flows. UNS is still a young concept which has evolved significantly just in the last couple of years. What the interviews and the practical implementation in this report showed is that companies require flexibility for the adoption of any concept. Perpetually working with limited resources, the ease of implementation, genuine business benefits, total cost and ROI are evaluated harshly by company leadership for large data infrastructure projects. Based on our findings, we have developed several recommendations to assist industrial companies in successfully launching and integrating Unified Namespace.

5.3.1 Not Strictly all Data Streams Should be Centralized

An observation made during the practical simulation of a virtual production site in 4.2.3 is that all distribution should not be facilitated by the UNS, meaning not every connection should be facilitated through the central broker. For example, the SCADA level should not retrieve its control data from the UNS, since it is more reliable and easier to establish a direct connection with shop floor, or edge devices. Data should nonetheless be collected from the shop floor and SCADA levels into the UNS, but communication vice versa should be limited. The reason behind this is the complexity issue. Traditionally, the shop floor (level 0) and control (level 1) levels are next neighboring levels in the automation pyramid. Control applications and communication protocols used for these are designed specifically for communication between them, therefore the extra connection and translation into (most likely) MQTT is an unnecessary step.

Another example of the complexity issue is applications that need a database to function, as the data is both contextualized and standardized in the database as it is collected, which can be seen in Figure 23. Moreover, the database also establishes stable transmission of data in SQL which is what the software needs. Historical data will be queried in SQL, translating it to another protocol, such as MQTT, will simply add complexity and extra connections, without the benefits. Therefore, the database should facilitate the transmission of the data directly to the querying application, mainly due to a simpler structure.

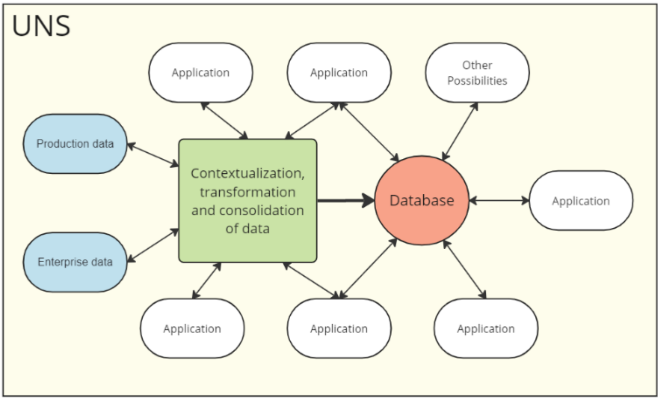


Figure 23. Representation of how a UNS should be constructed with a Database, highlighting that the UNS is not only the green box, but the entire concept.

5.3.2 Use of Communication Protocols for UNS Implementation

There are many companies that are trying to capitalize on UNS since it is a fairly undefined concept, and therefore claiming that their solution is a true UNS, leading to some confusion about the term. UNS is a concept, and not a single software, meaning there are various ways of implementing it and no one-size-fits-all solution. Through collected literature and practical examples (from mainly Walker Reynolds and software companies like HiveMQ), it is easy to imagine that MQTT is the de facto standard for backend industrial messaging protocols, and that existing UNS implementations strictly use it. However, through the simulation and the conducted interviews with manufacturing companies operating in different sectors, it does not seem to be the whole truth. While there certainly is some MQTT communication between current industrial infrastructures today, it seems highly limited, especially for older production sites with legacy equipment. Another unexpected discovery was that few software use MQTT as a native communication protocol, which strengthens the claim that it is not as widely adopted as literature might suggest. However, it should be mentioned that data transformation extensions or middleware software can be added to ensure that a MQTT broker can act as a central repository. For example, Grafana, a commonly utilized visualization tool, does not communicate in MQTT and instead uses a database or historian as source to collect datapoints, like many other visualization platforms. By adding an intermediary database, communication is commonly ensured through using historical data, whereas the MQTT broker only communicates in real-time. Forsberg at Alfa Laval mentioned that he had built a UNS-like structure using a database acting as a single source of truth. SKF is also building a UNS-like structure with their systems, gathering and dispersing data with HighByte's Intelligence Hub acting as an intermediary, while heavily relying on historians and databases. Neither of these architectures use MQTT as their primary communication protocol, and neither currently use it for the unique topic structure nor added level of context. Instead, SKF uses the Intelligence Hub to structure, add context to data, and transform between mainly OPC-UA and SQL.

The limitations of OPC-UA are, according to the results and some interviewed companies, exaggerated as it is possible to build large systems with it as the main protocol, spanning multiple factories. A study from the University of São Paulo comparing MQTT and OPC-UA pub/sub with regard to size of the total payload divided by total bytes transmitted, meaning the amount of data needed to transmit the same load of data. The study also researched transfer speed with a set payload of 16 bytes and varying amount of listener nodes, or devices connected to the same network getting that data (Rocha, Sestito, Dias, Turcato, & Brandão, 2018). It found a significant difference mostly when raising the amount of listener nodes, suggesting that the upside of MQTT is mostly seen when connecting 100+ nodes with regards to speed. It is important to note, however, that the study was limited to data exchange, even though OPC-UA offers other services such as data modeling, alarm and event management, variable history, access control among others. This is to say, that for companies using less than around 100 connected machines, the easy implementation of OPC-UA is not to be understated. To get started on their digitalization journey it is not incorrect to continue using OPC-UA, although at larger scale there is a clear upside of using MQTT instead.

Stena Recycling is a company that successfully uses a substantial amount of OPC-UA (and Rest API), at sites with up to 300 devices. However, the company predominantly has one-way communication at their sites, where the additional latency is of less importance. This is a system that works for now, but perhaps in the future will need to be restructured if more control is implemented. For controlling heavy machinery, latency is critical, whereas for pure tracking it may be acceptable with some latency. The topic is frequently discussed in the UNS community, referred to as “technical debt” by persons such as Walker Reynolds, pointing out the lesser scalability attributes.

Therefore, if a production site already uses predominantly OPC-UA and has less than around 100 edge devices, consisting of legacy equipment and aims to build a UNS for the site only, based on our findings, it is acceptable to continue using existing protocols. Although, there is a need for intermediary software processing those protocols and transforming them for other applications, while adding context and being able to transmit data in real-time. This will limit complexity and the number of programs needed, and thus likely keep operating and implementation costs down. Lowering the barrier for starting the digital transformation might allow companies to start heading in the right direction instead of being stuck in the same routines. However, if a company aims to connect multiple medium-large sites with 100+ devices, it is likely that bandwidth will start becoming a problem, thus MQTT is a better alternative to use long term. Acknowledging the truth in “technical debt”, especially with regards to scalability, will yield a more sustainable solution. Likewise, if a company is building a new factory, it is favorable to implement MQTT as the primary communication protocol, because of modern equipment likely having it preinstalled. By utilizing MQTT when it is applicable, the factory, and in extension the company, will be better prepared for the future, when multiple sites are connected, and network bandwidth may play a larger limiting role.

Another solution considered is one that takes inspiration from the Schultz method, suggesting that a company has multiple UNS's, one for each site, and one central for decision making, i.e. a UNS for headquarters. The multiple UNS's at a site level would use the dominant communication protocol for that specific site, while simultaneously transitioning towards the use of MQTT, hence minimizing costs and complexity for the factory. This in theory allows the factory to focus more on the production, retaining their systems and communication protocols that have historically performed well, while minimizing the learning curve for employees. For the local UNS, one could also implement Sparkplug B, as presented in Figure 24, to make use of its properties for control purposes. As a consequence of its topic depth limit, it is not advised to use it further upstream than the area - and lower levels.

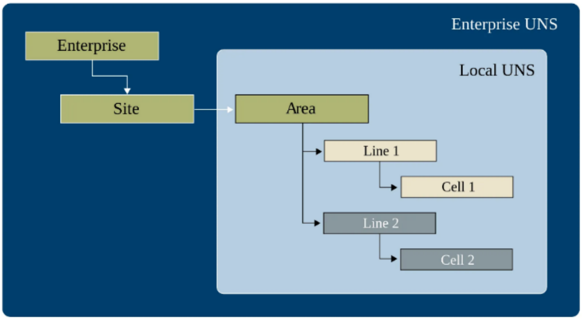


Figure 24. Building a UNS using Sparkplug B. (Manditereza, 2022)

The local UNS would communicate with the headquarters' UNS using pure MQTT, as the quantity of information from multiple sites is likely to overload network bandwidth if a poll/response protocol is used. Therefore, MQTT allows for a more scalable solution and as the company expands, with more sites and machinery, bandwidth will not act as a barrier. Furthermore, production – and enterprise data amount is expected to increase as productions continue to develop, by extracting more data, increasing autonomy and integrating more AI solutions, meaning the maturity of a company integrating more MQTT is higher. A drawback of this solution is that a company will not decrease costs relating to multiple software, as the production sites will continue to operate independently. It does, however, minimize risks that are tied to solely using one software provider while allowing a faster UNS implementation that bypasses the inherently slow-moving industrial complex.

5.3.3 IIoT Platform Instead of the Traditional SCADA System

Another realization is that the traditional SCADA system can be removed, like SKF has accomplished, since modern machinery can be controlled via the IIoT platform. By removing the traditional SCADA layer, the machines can be controlled through a tablet, computer or anything else capable of transmitting data, while simultaneously being able to control the machine from anywhere. This allows an operator to become responsible for multiple machines simultaneously, that are located in different locations within the factory. Something that usually is not possible with current systems. Removing the SCADA level offers a clearer data path and removes one data transfer node from the edge device but can compromise safety. As Berglund-Fast of Stena Recycling mentioned in her interview, it would be a security concern if a machine could be started from another location, since it increases the risk of injury among other things. The person starting the machine, or process, cannot see if somebody is performing maintenance or another activity that requires them to be close to or inside the machine. By adding a locking mechanism for people performing maintenance, or similar activities, this risk can be minimized, however, it is not eradicated.

5.3.4 The Necessity of a Data Modeling Intermediary Software

Technically speaking, it is possible for companies to write software for translation between protocols, contextualization and standardization of data in house. This, however, will require enormous resources with a questionable outcome and will inherently multiply complexity and cost. Instead, it is recommended to use existing software with these capabilities. One should look for ease of implementation, software cost, developing team credibility and product roadmap, and software functionality of course. SKF and Stena Recycling are two companies that use different software for this purpose, and both seem to be equally content with their choices. The representative of the international packaging business expressed openness for the UNS implementation in case there was software to do it with in a simple manner, further highlighting its importance. For scalability, an embedded (or at least easily integrable) MQTT broker is desired, making use of its light bandwidth messaging and native pub/sub architecture. There are several possible software that can be used as intermediary, some examples of these are HighByte’s Intelligence Hub, ThingWorx by PTC, OPC Router developed by Inray or HiveMQ broker. Recommending one falls outside the scope of this report and is subject to future research.

5.3.5 Growth Mindset

While not a technical perquisite, the importance and presence of growth mindset in the company culture is not to be understated. Most industrial companies can get trapped in not being forward thinking enough, settling for proven processes and ways of working. The opposite can be said for agile IT companies, that are faster to adapt into new environments and make use of new technologies, else they will never be successful. Today, the line between OT and IT is getting ever blurrier, with strategies such as DataOps gaining ground in industrial fields. Acknowledging and understanding the value of data, especially high-quality, structured and contextualized data is key to becoming data-driven. Industrial companies resistant to this change will get left behind, lagging after competitors that actively work with not only their product portfolio, but also their operations in detail. In the background for all of this stands the company culture, allowing for ownership, failure and personal growth. DataOps solutions are traditionally foreign for players in the industry, many times playing it safe and focusing on keeping up existing operations. At times of change, however, fortune does favor the brave. In order to be brave, trying new methods and ways of working, the company culture (stemming from the example of leadership) must embrace failure. Implementing a new and somewhat unproven data architecture will most likely bear missteps along way, and having this clear from the beginning will save a lot of headaches, and hopefully result in an even better outcome. Growth mindset, embracing failures and working iteratively might sound like buzzwords, but in the case of Unified Namespace, they are necessities.

5.3.6 Knowledge and Data Sharing Between Sites in the Same Enterprise

During the interviews, it was noted that industrial company sites have a tendency to operate as separate entities, without much regard taken to them being in the same enterprise. Sites have their own teams for much of the same sort of tasks that other sites are implementing, in a somewhat distinct manner, solving the same arisen issues that somebody else within the enterprise has likely already solved. Implementing Unified Namespace enterprise-wise can be a challenging task, especially if some sites are greatly lagging behind others regarding their level of digitalization. As discussed previously, the concept cannot be implemented in places that barely have reached Industry 3, as the jump is simply too much to be completed simultaneously. Industrial companies need to have teams working on an enterprise-level, to share both knowledge and data with sites that are lagging behind. There needs to be an effort for standardization, else the ad-hoc solutions implemented will quickly take over, rendering data management efforts ineffective.

5.3.7 ISA-95 Defined Nomenclature

The role of nomenclature for data standardization is central. If a company aims to reduce dependency on specialists, as well as ease the flow of information and reduce time spent on analysis, it is essential that they follow some kind of standard for nomenclature. Among the participants for the interviews in this report, nomenclature defined in ISA-95 is predominantly what they aim to follow. At times, questioned companies use it as overall guidelines as opposed to following it to the letter, but nevertheless, the general consensus is that it is a useful and intuitive nomenclature to adopt. Moreover, the topic structure that pub/sub communication protocols allow for work perfectly together with the proposed nomenclature, creating an intuitive architecture that stakeholders can easily make sense of. For these reasons, we recommend that teams looking to implement a UNS structure in their own companies can use this nomenclature as a basis. This is not a requirement for a successful UNS implementation, it can however be considered good praxis for its many upsides.

5.4 Limitations of Research

During the thesis work period, several limitations were encountered, complicating the project. Using the aid of advisors from the Faculty of Engineering at Lund University and Novotek Sverige AB, it was however possible to overcome most issues to the best of our abilities.

Keeping impartiality throughout the report could be difficult at times, as Novotek is active in this very field of industrial digitalization. For this reason, their coworkers that could be queried for specific investigations could be biased towards the company and its affiliate companies, business partners. Realizing this, it was deemed best to use a mix of software for the implementation, as well as companies for the interviews, to include both those that are affiliated with Novotek, and those that are not. As introduced in section 1.3, the number of interviewed companies was limited to five total, out of which four were manufacturing companies. The amount was deemed adequate and in line with a qualitative report, prioritizing understanding these companies in-depth, rather than finding superficial information from several businesses. One purpose, as stated in the same chapter, was to research companies both with and without background in digitalization. After completing the interviews, it can be concluded that all companies are working with digitalization, although the actual advances may vary.

Reliability of sources is yet another thing to be addressed. There is a genuine lack of scientific publications regarding Unified Namespace, most likely as a consequence of its novelty. The majority of web-based literature can be connected to companies active in the sphere and should be viewed critically regarding impartialness. Scientific sources could mostly only be found with regards to underlying technologies that make up the concept.

6. Conclusion

To summarize this report, a definition of a UNS has been created:

A repositor of structured and contextualized data, acting as a single source of truth using scalable technologies, in one, or multiple centralized locations within the enterprise or factory site.

6.1 Concluding Results

RQ1: Why is Unified Namespace relevant for manufacturing companies and what problems does it solve?

Manufacturing companies are with increasing interest looking at UNS, since it provides a platform for solving common data management and data transformation problems. UNS facilitates faster and more efficient operational processes, while streamlining accessibility and common data management. Below some of the essential issues that UNS solve are highlighted:

- *Ease of accessing data and data distribution.* UNS facilitates easier access to data and retrieval of the information that the individual needs. This leads to the company being able to respond quicker, perform analysis with greater accuracy and overall efficiency is increased.

- *Enhanced analysis capabilities*. In line with previous argument, as data within a UNS is standardized in easy to access repositories and widely available, it makes it easier for an enterprise to perform up-to-date analysis.
- *Future proofs the company.* With UNS, due to contextualized data and standardization, it for instance becomes possible to truly leverage the power of AI. It becomes easier to add new machines in the virtual and digital space as the data sources are standardized, which in turn will decrease the need for specialists.
- *External utilization of data.* With a UNS, factory data can easily be consumed by the enterprise, but organizational data (enterprise and factory) can also be consumed outside the boundaries of the organization. This could mainly impact the supply chain of the organization, but also enables a new market segment where data is traded. The company can sell data or negotiate mutually beneficial deals that leverage data from the organization, creating models, optimizing production and more, leading to increased focus on core business areas.
- *Consistent Data.* Every individual at the organization has access to the same data, preventing discrepancies caused by data interpretation.
- *Reduce human errors.* By automating data collection, contextualization etc. human errors are reduced, while analysis based on uniform data becomes more accurate.

RQ2: Are there any barriers, challenges or drawbacks to implementing and using Unified Namespace?

While UNS offers many benefits, it has multiple barriers and challenges tied to implementation which should be considered and avoided to successfully implement the solution.

- *Industry 4.0 ready and technology compatibility.* To benefit from implementing UNS in an organization it is essential that the organization is at an Industry 3 level beforehand or has an industry

4.0 mindset. Implementing it without any previously implemented necessary technologies makes it extremely expensive, while the added value of it becomes more ambiguous.

- *MQTT*. MQTT presents a significant barrier as devices should be MQTT ready to make the integration as smooth as possible, and the organization familiar with the protocol. However, some companies do not need the scalability dimension added by MQTT.
- *Initial investment.* The initial investment of developing and adopting a UNS is high (SKF's system that is UNS-like has cost around €1 million as of today, and they have a lot of previous knowledge of digitalization and automation) and calculating a ROI is challenging. It is also vital that a company implementing a UNS solution is not afraid to restart the project, scrapping all previous progress, as you learn along the implementation and development of UNS.
- *Existing systems and lack of expertise. C*ompanies generally have systems implemented that fulfill their needs, which act as a major barrier for implementing new systems. It becomes harder to justify the investment costs if systems already work adequately. Likewise, there is a knowledge gap, often with individuals with power, that further enhances the challenges of switching system.
- *Lack of standardization.* Every company can, and should, adapt the UNS to fit their organization's needs, since everybody has unique needs. Unfortunately, the byproduct of this is that there is rarely a one-fits-all solution, meaning every company have to build their unique UNS.
- *Security*. While robust security measures can be implemented to protect the UNS, should a hacker gain access to the UNS, all organizational data is available and contextualized like never before. This leads to detailed information being readily available, like location of machines, all parameter data and possibly even allowing unauthorized control. The system is designed to facilitate "easy" access for authorized users, meaning if an unauthorized user bypasses security all data is easily available.

- *Program dependencies.* Having one system for the UNS where all other applications extract data and send data poses a possible risk. If that system stops working, or for some reason gets cancelled, all other systems will shut down, likely leading to significant costs, downtime and operational losses.

RQ3: How can/should a Unified Namespace be implemented in a manufacturing company?

UNS is a versatile concept, which leads to companies being able to make their own interpretations and adaptations, aligning with their vision and organization. This is similar to ISA-95 which leaves the concept open for interpretation, with specific operational frameworks and requirements in mind. In the end the organization has molded their own definition of UNS that fits into their organization well. With this in mind, there are some specifications that affect how it should be implemented.

- *Scalability and protocol selection.* The choice of communication protocol can vary depending on the site, in terms of scale and machinery age, and also the requirements of the project from the board. For small-scale projects with less focus on scalability, bandwidth is not a limiting factor, allowing for virtually any communication protocol. Projects of larger scale are recommended to MQTT due to it being lightweight and no technical debt is acquired. However, such a project likely has a high initial investment cost and will not allow for as natural of a transition as a slow-paced project does.
- *Control systems.* Control systems that require real-time data should not rely on the UNS, ensuring that critical operations are not delayed or interrupted due to lag or similar failures. However, it should be mentioned that it is possible to build a new type of control system, separate from the SCADA, which allows control of machinery through the UNS. An appropriate implementation of such a system is to add a long transition period, ensuring functionality or use both systems simultaneously.

- *Data modeling and communication efficiency.* Source data should be modeled and stored in historians and databases, and communication with some programs requiring said data should be facilitated through the database or historian. Meaning, not all communication should go directly through the UNS, and it is essential to identify data interaction beforehand to optimize flow, reduce overhead and potentially avoid bottlenecks.
- *Transformation.* Crucial for UNS is transformation of data, as different programs use different communication protocols. Therefore, it is essential to choose programs that ensure data transformation between different platforms and that all applications can use the contextualized data. To exemplify, most applications do not speak MQTT today.
- *Restart projects and autonomize data.* During the implementation phase it is appropriate to allow the team to restart the project as they feel it is required. During this phase, they will make realizations and find improvement potentials, and even though the current version of the project works it can be essential to restart, to future proof. To simplify the process, it is necessary to codify as much as possible and autonomize data flows and modeling. This will also facilitate easier overall changes after the program is implemented, for example changing a machine name.
- *Use a nomenclature standard like defined in ISA-95.* Standardization of data starts with a common nomenclature, and the one defined in ISA-95 is highly recommended to use. It is easy to understand and works in perfect harmony with publish/subscribe communication protocols.

RQ4: How does Unified Namespace differ from existing infrastructure?

UNS represents a significant shift and evolution in data management and system integration compared to traditional concepts and methods.

- *Connects factories to the enterprise.* Historically, data management systems were confined to a production site or area. UNS allows for scaling of these capabilities up to enterprise level, creating a more comprehensive and uniform management. Compared to different systems the factory merges with the entire enterprise unlike many current systems that are limited to KPIs.
- *Inclusion of non-factory data.* Unlike many other concepts, UNS encompasses a broader spectrum of data. This supports a more holistic view of the organization, providing insights beyond the production lines, perhaps explaining why certain downtimes occur and so on. Areas that typically can be included in a UNS, apart from production data, are supply chain data, customer interaction, warehouse status and more.
- *Focus on scalability.* Scalability is one of the core focus areas of UNS and it is designed to allow multiple users to access data simultaneously, from anywhere. This is rarely possible in traditional systems.
- *Standardization of data management.* UNS offers a more defined concept of data management than other systems, something that companies have struggled with creating on their own. The standardization will likely give different systems and software suppliers a direction to continue developing, while simplifying IT architectures.
- *Reduction of system complexities* Traditional automation architectures connections can from a snapshot look like a messy plate of spaghetti, as there are multiple connections from one driver to another. UNS streamlines this and can provide a platform where inefficiencies and errors associated with system complexities are removed. By reducing the number of connections for an edge device, the risk of it overloading also decreases, leading to a more stable connection.
- *Enabling real-time applications.* Another transformative aspect of UNS is the ability of it being able to support multiple real-time applications. Traditional PLCs can become overloaded with multiple connections and databases are usually not used for real-time

applications, as data transfer generally is slower, due to the amount of data.

6.2 Future Research

This master's thesis has managed to establish a foundation for implementing a UNS concept in an organization, while defining it and highlighting strengths and barriers with it. We suggest that further research should be conducted in a real-life environment, particularly within a manufacturing setting that includes several machines or operates at an enterprise level. Applying learnings from this report and knowledge gaps in a real company would validate theoretical findings in the report, while hopefully providing insights into knowledge gaps and uncovering new insights.

A suitable environment for doing this is a factory with more than 100 connected devices, getting access to the data, and build a UNS implementation. This scale is chosen based on the identified scalability challenges of OPC-UA as noted by Reynolds, who argues that OPC-UA presents barriers to effective UNS deployment due to technical debt and limited scalability. Do larger servers limit the barrier and how can the sub/pub protocol within OPC-UA impact a UNS? This report concluded that at around 100 connected devices advantages of MQTT will become noticeable. To test this claim, connecting 100 or more machines to a UNS using both OPC-UA and MQTT for comparative analysis. Additionally, exploring other protocols could provide new insights and perhaps even better performance. This could potentially also be achieved by collecting interviews from multiple companies that have implemented a UNS. As Walker Reynolds has implemented multiple systems, he could perhaps provide some companies with it or further explain why OPC-UA does not work.

Another critical area for further research is the cost associated with UNS implementations and the potential savings. It is important to find the costs, including labor and software expenses, while examining the tangible savings from reduced software needs and time savings through easier data analysis and availability. Expanding on this, research could also explore how improved data availability can lead to better and more accurate organizational developments and savings. These investigations should aim to quantify a typical return on investment (ROI) for UNS implementations as it is a critical investment tool for organizations. Another aspect of UNS that this report only shortly has described is how it influences AI applications, which not only predicts maintenance, but can lead to meaningful savings. Further research into Industrial AI in the context of contextualized data should also be made.

Future research could also explore other concepts of data management in a company comparing them to UNS.

Lastly by changing research method to quantitative, future research becomes more objective and produces more widely applicable results. By doing this relatively simple change, the result of new research is less general and Figures like, for example, ROI can be found, and speed comparisons between OPC-UA and MQTT can be made.

7. References

Alguliyev, R., Imamverdiyev, Y., & Sukhostat, L. (2018). Cyber-physical systems and their security issues. *Computers in Industry, Volume 100*, 212-223.

Brandl, D., & Johnsson, C. (2021). *Beyond the Pyramid: Using ISA95 for Industry 4.0 and Smart Manufacturing.* Retrieved from International society of automation.

Chen, Hsing-bung, (2018). A Universal Namespace Approach to Support Metadata Management and Efficient Data Convergence of HPC and Cloud Scientific Workflows. *IEEE International Conference on Big Data (Big Data)*, 516-521.

Chen, W., Yan-Yi, L., Tie-Zheng, G., Da-Peng, L., Tao, H., Zhi, L., Qing-Wen, Y., Hui-Han, W., Ying-You, W. (2024). Systems engineering issues for industry applications of large language model. *Applied Soft Computing*, 111165.

David, J., Martikkala, A., Lobov, A., & Lanz, M. (2019). A Unified Ontology Namespace for Enterprise Integration - a Digital Twin Case Study. *Conference: Instrumentation Engineering, Electronics and Telecommunications – 2019*, 13-22. November 20–22, 2019, Izhevsk, Russian Federation

Elsevier, (2024). *Choosing the Right Research Methodology: A Guide for Researchers.* Retrieved from Elsevier: https://scientific-publishing.webshop.elsevier.com/research-process/choosing-the-right-research-methodology-a-guide-for-researchers/ (Accessed 2024-03-25)

Erumban, A. A. (2018). *Lifetimes of Machinery and Equipment. Evidence from Dutch Manufacturing*. University of Groningen.

Foegeding, E. A. (2020). Basic, applied, and developmental R&D should be under one roof. *Journal of Food Science*, 2264.

Ghobakhloo, M. (2020). Industry 4.0, digitization, and opportunities for sustainability. *Journal of Cleaner Production, Volume 252, 119869*.

Harrington, J. (2021). *DataOps: Fundamental for Industrial Transformation.* Retrieved from International Society of Automation.

HighByte, (2024). *MQTT Broker*. Retrieved from HighByte: https://www.highbyte.com/intelligence-hub-mqtt-broker.html (Accessed 2024-02-22)

HighByte, (2024). *The Industrial DataOps Solution for Industry 4.0*. Retrieved from Highbyte: https://www.highbyte.com/intelligence-hub.html (Accessed 2024-02-21)

Höst, M., Regnell, B., & Runeson, P. (2006). Att genomföra examensarbete. Lund, Sweden: Studentlitteratur AB.

IBM, (2024). *IBM*. Retrieved from What is Industry 4.0?: https://www.ibm.com/topics/industry-4-0 (Accessed 2024-01-26)

Inductive automation, (2024). *Ditch Data Silos: Create a Unified Namespace with Ignition UDTs & MQTT*. Retrieved from Inductive automation: https://inductiveautomation.com/resources/webinar/ditch-data-silos-create-a-unified-namespace-with-ignition-udts-mqtt (Accessed 2024-04-15)

Inductive automation, (2024). *SCADA Software*. Retrieved from Inductive automation: https://inductiveautomation.com/scada-software/ (Accessed 2024-03-02)

Internet Society, (2024). *TLS Basics*. Retrieved from Internet Society: https://www.internetsociety.org/deploy360/tls/basics/ (Accessed 2024-02-28)

Jaylin. (2024). *Unified Namespace (UNS): Introduction and Its Applications in IIoT*. Retrieved from emqx: https://www.emqx.com/en/blog/unified-namespace-next-generation-data-fabric-for-iiot#what-is-unified-namespace (Accessed 2024-04-13)

Koprov, P., Ramachandran, A., Lee, Y.-S., Cohen, P., & Starly, B. (2022). Streaming Machine Generated Data via the MQTT Sparkplug B Protocol for Smart Factory Operations. *Manufacturing Letters*, 66-73.

Larochelle, P. (2022). *MQTT Sparkplug B, or Not to B? Unified Namespace Architecture Considerations*. Retrieved from neomatrixinc: https://neomatrixinc.com/blog/sparkplug-b-or-not-to-b-unified-namespace-architecture-considerations/ (Accessed 2024-02-25)

Lasi, H., Fettke, P., Kemper, H.-G., Feld, T., & Hoffmann, M. (2014). Industry 4.0. *Business & Information Systems Engineering, Volume 6*, 239–242.

Law, A. M. (2015). Simulation Modeling. Tucson, Arizona, USA: McGraw-Hill Education.

Lindqvist, N. (2024). (Á. Péter, & S. Werner, Interviewers)

Manditereza, K. (2022). *Implementing Unified Namespace (UNS) With MQTT Sparkplug*. Retrieved from HiveMQ: https://www.hivemq.com/blog/implementing-unified-namespace-uns-mqtt-sparkplug/ (Accessed 2024-03-05)

Manditereza, K. (2022). *What is Unified Namespace (UNS) and Why Does it Matter?* Retrieved from HiveMQ: https://www.hivemq.com/blog/what-is-unified-namespace-uns-iiot-industry-40/ (Accessed 2024-01-30)

Manditereza, K. (2023). *Data and Functional Modeling for Unified Namespace*. Retrieved from HiveMQ: https://www.hivemq.com/blog/data-functional-modeling-dataops-unified-namespace-iiot/ (Accessed 2024-03-05)

Manditereza, K. (2023). *HiveMQ*. Retrieved from Foundations of the Unified Namespace Architecture for IIoT: https://www.hivemq.com/blog/foundations-of-unified-namespace-architecture-iiot/ (Accessed 2024-02-30)

MQTT, (2024). *MQTT: The Standard for IoT Messaging*. Retrieved from MQTT: https://mqtt.org/ (Accessed 2024-02-15)

Navathe, S. B. (1992). Evolution of Data Modeling for Databases. *OMMUNICATIONS OF THE ACM Vol.35*, 112-123.

OPC Foundation. (2024). *What is OPC?* Retrieved from OPC Foundation: https://opcfoundation.org/about/what-is-opc/ (Accessed 2024-02-25)

Peres, R. S., Jia, X., Lee, J., Sun, K., Colombo, A. W., & Barata, J. (2020). Industrial Artificial Intelligence in Industry 4.0 - Systematic Review, Challenges and Outlook. *IEEE Access 8*, 220121 - 220139.

PTC. (2024). *What is Kepware?* Retrieved from PTC: https://www.ptc.com/en/products/kepware (Accessed 2024-03-02)

Ragatz, G. L., Handfield, R. B., & Scannell, T. V. (1997). Success Factors for Integrating Suppliers into New Product Development. *Journal of Product Innovation Managment*, 190-202.

Reynolds, W. (2023, September 10). Unified Namespace for Industrial IoT: The Masterclass. (Wrighter, USA, Interviewer)

Rocha, M. S., Sestito, G. S., Dias, A. L., Turcato, A. C., & Brandão, D. (2018). Performance Comparison Between OPC UA and MQTT for Data Exchange. *2018 Workshop on Metrology for Industry 4.0 and IoT*, 175-179, 16-18 April 2018, Brescia, Italy.

Rose, K., Eldridge, S., & Chapin, L. (2015). *The Internet of Things: An overview*. The Internet Society. Retrieved from Internet Sociaty.

Rout, S. K. (2024). *IOT Architecture with MQTT Sparkplug(B)*. Retrieved from International Society of Automation: https://blog.isa.org/iot-architecture-with-mqtt-sparkplugb (Accessed 2024-02-10)

Scada International, (2024, 01 29). *What is SCADA?* Retrieved from Scada International : https://scada-international.com/what-is-scada/

Schneider Electric, (2024). *What is Modbus and How does it work?* Retrieved from Schneider Electric: https://www.se.com/us/en/faqs/FA168406/ (Accessed 2024-03-12)

Schuitemaker, R., & Xu, X. (2020). Product traceability in manufacturing: A technical review. *Procedia CIRP Volume 93*, 700-705.

Shi, Z., Xie, Y., Xue, W., Chen, Y., Fu, L., & Xu, X. (2020). Smart factory in Industry 4.0. *Systems Research and Behavioral Science Volume 37, Issue 4*, 607-617.

Sisinni, E., Saifullah, A., Han, S., Jennehag, U., & Gidlund, M. (2018). Industrial Internet of Things: Challenges, Opportunities, and Directions. *IEEE Transactions on Industrial Informatics*, 4724-4734.

Statista. (2024), *Average costs of industrial Internet of Things (IoT) sensors from 2004 to 2020*. Retrieved from Statista: https://www.statista.com/statistics/682846/vr-tethered-hmd-average-selling-price/ (Accessed 2024-03-15)

Tomczak, D. L., Lanzo, L. A., & Aguinis, H. (2018). Evidence-based recommendations for employee performance monitoring. *Business Horizons Volume 61*, 251-259.

Wally, B., Huemer, C., & Mazak, A. (2017). Entwining plant engineering data and ERP information: Vertical integration with AutomationML and ISA-95. *2017 3rd International Conference on Control, Automation and Robotics (ICCAR)*. 22-24 April 2017, Nagoya, Japan

Wang, S., Wan, J., Zhang, D., Li, D., & Zhang, C. (2016). Towards smart factory for industry 4.0: a self-organized multi-agent system with big data based feedback and coordination. *Computer Networks Volume 101*, 158-168.

Yin, R. (2009). Case Study Research and Applications: Design and Methods. Los Angeles, USA: SAGE Publications, Inc.

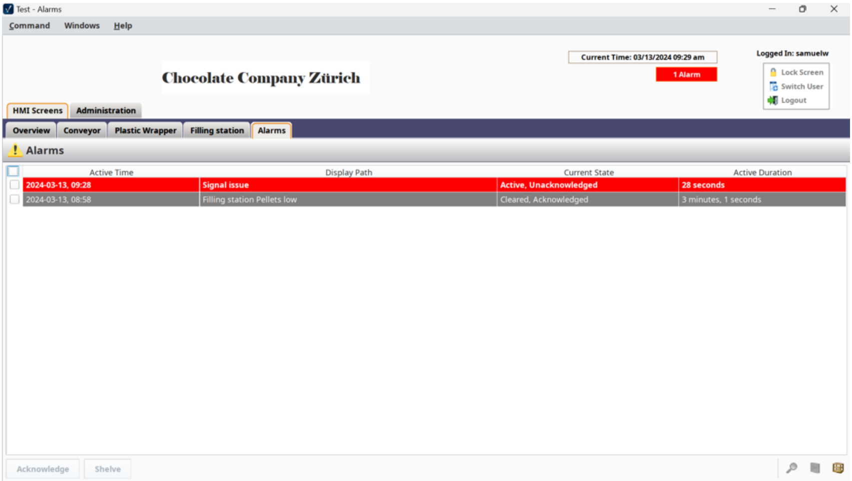
Yngvesson, U. (2024). (Á. Péter, & S. Werner, Interviewers)

Zohaib, J., Farhad, A., Wolfgang, M., Niki, P., Georg, G., Markus, S., & Ana, K. (2023). Artificial intelligence for industry 4.0: Systematic review of applications, challenges, and opportunities. *Expert Systems with Applications 216*, 119456.

Zou, P. X., & Xu, X. (2023). Research Methodology and Strategy. Hoboken, New Jersey, USA: John Wiley and Sons.

Appendix

Appendix A



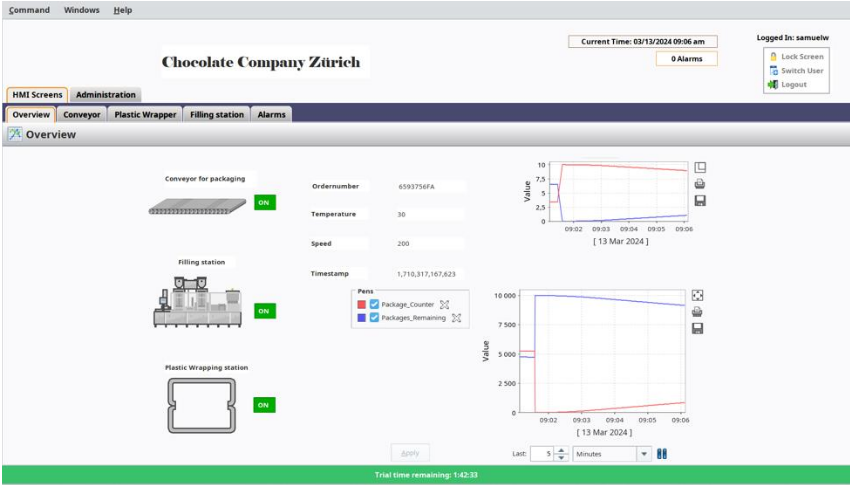
Current Time: 03/13/2024 09:29 am

Logged In: samuelw

Alarms

☐	Active Time	Display Path	Current State	Active Duration
☐	2024-03-13, 09:28	Signal issue	Active, Unacknowledged	28 seconds
☐	2024-03-13, 08:58	Filling station Pellets low	Cleared, Acknowledged	3 minutes, 1 seconds

The interface of the alarm screen where operators can acknowledge and clear alarms. An active error is indicated to the operator in the top right corner.



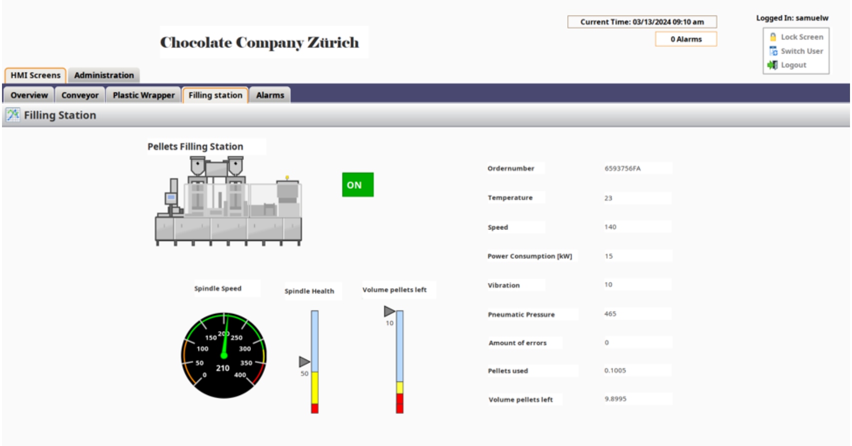
Current Time: 03/13/2024 09:06 am

Logged In: samuelw

Ordernumber	6593756FA
Temperature	30
Speed	200
Timestamp	1,710,317,167,623

Last: 5 Minutes

An overview of all of the machines in the line is shown in this tab of the system.

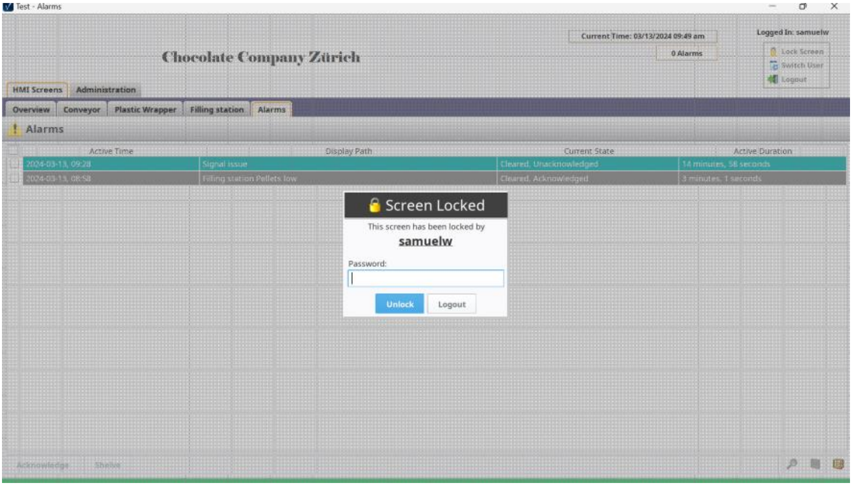


Current Time: 03/13/2024 09:10 am

Logged In: samuelw

Ordernumber	6593756FA
Temperature	23
Speed	140
Power Consumption [kW]	15
Vibration	10
Pneumatic Pressure	465
Amount of errors	0
Pellets used	0.1005
Volume pellets left	9.8995

More detailed information about one machine in the line can be seen in the tabs to the right. All parameters are shown to the right while some are highlighted in the bottom left.



Active Time	Display Path	Current State	Active Duration
2024-03-13, 09:28	Signal issue	Cleared, Unacknowledged	14 minutes, 58 seconds
2024-03-13, 08:58	Filling station Pellets low	Cleared, Acknowledged	3 minutes, 1 seconds

This screen has been locked by

samuelw

If no moment is detected for five minutes the screen automatically locks, which also can be done manually on the buttons on the top right of the screen.

Lund University **Department of Automatic Control** **Box 118** **SE-221 00 Lund Sweden**	*Document name* MASTER’S THESIS
	Date of issue June 2024
	Document Number TFRT-6254
Author(s) Ádám Péter Samuel Werner	*Supervisor* Nils Lindqvist, Novotek Sverige AB Jens Molin, Novotek Sverige AB Charlotta Johnsson, Dept. of Automatic Control, Lund University, Sweden Ulf Jeppsson, Dept. of Industrial Electrical Engineering and Automation, Lund University, Sweden (examiner)
Title and subtitle The Impact of Unified Namespace in Industry 4.0	

Abstract

Digitalization is part of the ongoing transformation for industries that are going through the fourth industrial revolution, Industry 4.0. The Fourth Industrial Revolution introduces significant advancements in digital technology, fundamentally transforming manufacturing environments through interconnected and intelligent systems. Organizations' willingness to adopt and integrate various AI solutions and other digital technologies presents a challenge for integrating, using, and storing data across various operational and information technology systems. Traditional data management concepts and solutions are not built for these new technologies and therefore, rarely able to fully capitalize on the potential of data.

This master's thesis investigates how Unified Namespace (UNS) can serve as a central hub for data communication, thereby addressing the challenges organizations face due to fragmented data systems, while supporting the implementation of advanced analytics and artificial intelligence. This report focuses on implementation of Unified Namespace systems, while identifying potential benefits, drawbacks, and barriers with the technology. It also considers alternative methods and compares them to Unified Namespace architectures.

The study adopts a comprehensive methodology that includes case studies of manufacturing companies, interviews with industry leaders, and simulations. These elements are complemented by an extensive review of existing literature, providing a deep dive into the theoretical foundations of digitalization in industry, UNS and its practical implications.

Key findings in this report include Unified Namespace significantly streamlining data management by acting as a centralized repository, enabling real-time data sharing at all levels of the organization. Integrating Unified Namespace enables significant time savings, improving operational efficiency, while facilitating higher accuracy decision-making, and at the same time reduces complexity of data systems. The main barriers presented are the initial complexity of transitioning, the upfront investment and potential resistance to change within the organization.

In conclusion, Unified Namespace, despite its challenges, prevails as a concept that can become critical in the transition to Industry 4.0. Not only does it enable seamless communication but is a necessary step for enabling multiple AI solutions and more. This report concludes that gradually transforming communication protocols to Message Queueing Telemetry Transport (MQTT), using the Schultz method, can be a way to decrease initial investment costs, while transitioning to the more effective protocol. Another insight is the
importance of using a platform that is able to communicate with all of the business-critical systems.

This thesis provides a foundation for future research to explore Unified Namespace and its long-term impacts on the manufacturing industry. It also serves as a practical guide for industrial entities going through a digital transformation and wanting to implement data management systems.

Keywords

Classification system and/or index terms (if any)

Supplementary bibliographical information

ISSN and key title 0280-5316	*ISBN*

Language English	*Number of pages* 1-125	*Recipient's notes*
Security classification		

http://www.control.lth.se/publications/