LAPPEENRANTA-LAHTI UNIVERSITY OF TECHNOLOGY LUT
School of Engineering Science
Degree Program in Industrial Engineering and Management

Milena Myllyniemi

DATA MODELING IN IMPLEMENTING PROCESS INFORMATION MANAGEMENT SYSTEM

Master's Thesis

Examiners: Professor Timo Kärri
Post-doctoral researcher Lasse Metso

ABSTRACT

Lappeenranta-Lahti University of Technology LUT
School of Engineering Science
Degree Program in Industrial Engineering and Management

Milena Myllyniemi

Data Modeling in Implementing Process Information Management System

Master's thesis
2021
82 pages, 27 figures, 3 tables, 2 lists and 3 appendices
Examiners: Professor Timo Kärri and Post-doctoral researcher Lasse Metso

Keywords: data model, data modeling, Industry 4.0, cloud computing, OPC UA

The amount of data collected is already high, but to analyze and use it to meet the business strategic need is still evolving. The data models organizes the data to be used easier and they are build based on requirements set by use cases. Industry 4.0, cloud, OPC UA and ISA-95 are providing standards and ways to better collect, process and use the data. The purpose of this master's thesis is to create an evolved data modeling process which supports a flexible and transparent data model.

This study is done using qualitative research. A literature review and semi-structured interviews are used as data collection methods. The purpose of this thesis is to find out best practice process for implementing information management system from the perspective of data modeling. The effects of industry trends are also considered. The data gathered from the literature review and empirical section were analyzed and compared to reach the aim. A total of 10 interviews were conducted.

The literature review revealed that the data model process should support flexible data models and that the Industry 4.0 encourages with implementation requirements to development in manufacturing industries. The results from the interviews show that the process of implementing data modeling is still under development. The main challenges of data modeling in implementing process information management systems in process industries are the slow development, high volume of manufacturing data and data model life cycle management.

TIIVISTELMÄ

Lappeenrannan-Lahden teknillinen yliopisto LUT
School of Engineering Science
Tuotantotalouden koulutusohjelma

Milena Myllyniemi

Tietomallinnus prosessitietohallintajärjestelmän käyttöönotossa

Diplomityö
2021
82 sivua, 27 kuvaa, 3 taulukkoa, 2 luetteloa ja 3 liitettä
Tarkastajat: Professori Timo Kärri ja tutkijatohtori Lasse Metso

Hakusanat: tietomalli, tietomallinnus prosessi, Industry 4.0, pilvilaskenta, OPC UA

Kerätyn tiedon määrä on jo suuri, mutta tiedon analysoinnissa ja käytössä liiketoiminnan strategisen tarpeen täyttämisessä on vielä kehittävää. Tietomalli järjestää tiedon helpommin hyödynnettäväksi, ja se rakennetaan tarvittavien tietojen perusteella. Industry 4.0, pilvi, OPC UA ja ISA-95, tarjoavat standardeja ja tapoja kerätä, käsitellä ja käyttää dataa paremmin. Tämän diplomityön tarkoituksena on luoda kehittynyt tietomallinnusprosessi, joka tukee joustavaa ja läpinäkyvää tietomallia.

Tämä diplomityö on toteutettu kvalitatiivisena tutkimuksena. Tiedonkeruumenetelmänä käytetään kirjallisuuskatsausta ja puolistrukturoituja haastatteluja. Tavoitteena on selvittää parhaat käytännöt prosessitietojen hallintajärjestelmän käyttöönottoon datamallinnusmenetelmän näkökulmasta. Myös teollisuuden trendien vaikutukset otetaan huomioon. Kirjallisuuskatsauksesta ja empiirisestä osasta kerättyjä tietoja analysoitiin ja verrattiin toisiinsa tutkimuksen tavoitteen saavuttamiseksi. Haastatteluja tehtiin yhteensä 10.

Kirjallisuuskatsaus toi ilmi, että tietomalliprosessin pitäisi tukea joustavaa tietomallia ja että Industry 4.0 kannustaa kehitykseen käyttöönotto vaatimuksien kanssa valmistavan teollisuuden alalla. Haastattelujen tulokset osoittavat, että tietomallinnuksen käyttöönottoprosessi on vielä kehitteillä. Tietomallinnuksen suurimmat haasteet prosessitietohallintajärjestelmän käyttöönotossa prosessiteollisuudessa ovat hidas kehittyminen, suuri määrä valmistustietoa ja tietomallin elinkaari hallinta.

ACKNOWLEDGEMENTS

This thesis was done for ABB. Thanks to Jukka Kostiainen, Ville Sällinen and Jyrki Peltoniemi for all support and ideas. Also, I want to thank all interviewees for their time and insights which were valuable for the research. Especially big thanks to Outokumpu for development ideas. I want to thank you my supervisor Timo Kärri for suggestions and guidance during this project.

Special thanks to my family and friends. They have supported and motivated throughout my studies.

Helsinki, 31.8.2021

Milena Myllyniemi

TABLE OF CONTENTS

1 Introduction 8

1.1 Background 8

1.2 Objectives and scope 9

1.3 Execution of the study 10

1.4 Structure of the study 11

2 Process data modeling 13

2.1 Data model 13

2.2 Data model life cycle 15

2.3 Time series data 22

2.4 Process information management systems 23

3 Industry trends 25

3.1 Cloud computing 28

3.2 Open Platform Communications Unified Architecture (OPC UA) 30

3.3 International Society of Automation (ISA) 32

3.4 Effects of industry trends on data modeling 34

4 Implementing ABB Ability™ History 38

4.1 Introduction to ABB 38

4.2 Introduction to ABB Ability™ History 38

4.3 Current process workflow 41

5 Results 47

5.1 Process data 48

5.2 Data modeling 50

5.3 Data modeling lifecycle 52

5.4 Industrial digitalization and industry trends 55

5.5 Implementing equipment model 58
5.6 ABB and other suppliers' services 60
6 Evolved data modeling process 62
7 Conclusions 69
7.1 Answers to research questions 69
7.2 Limitations and further research 72
References
Appendices

LIST OF FIGURES

Figure 1 Data modeling as a design activity (adapted from Simsion & Witt 2005) 16
Figure 2 Data Stewards (Henderson et al. 2017)..17
Figure 3 Data modeling activities (adapted from Henderson et al. 2017)...............................18
Figure 4 Database design tasks and deliverables (adapted from Simsion & Witt 2005) 21
Figure 5 Examples of PIMS functionalities (Du et al. 2018) 24
Figure 6 Technological trends forming Industry 4.0 (Gilchrist 2016) 26
Figure 7 Industry 4.0 Design Principles (Gilchrist 2016) 27
Figure 8 Equipment hierarchy (adapted from Chen 2005)...34
Figure 9 Effects of industry trends on data modeling, PIMS and time series data 37
Figure 10 ABB Ability™ History database management system (ABB 2021) 39
Figure 11 The conceptual model (ABB 2021) 40
Figure 12 Current data modeling process of implementing equipment model 42
Figure 13 Current process, defining requirements and building the data model......................43
Figure 14 Current process, maintaining the data model...44
Figure 15 Current process, modifying the equipment class 45
Figure 16 Current process, modifying the hierarchy..45
Figure 17 Current process, creating a new equipment instance 46
Figure 18 Process data characteristics...48
Figure 19 Conclusions of main differences between variable models and equipment model .52
Figure 20 Familiarity of Industry 4.0 and considering it in data modeling..............................57
Figure 21 Main challenges in implementing equipment model 59
Figure 22 Evolved process roles..63
Figure 23 Evolved process, planning and building the data model...65
Figure 24 Evolved process, reviewing the data model..66
Figure 25 Evolved process, modifying the equipment class 67
Figure 26 Evolved process, modifying the hierarchy...67
Figure 27 Evolved process, creating a new equipment instance 68

LIST OF LISTS

List 1 Base principals of UA modeling (Pauker et al. 2016)...32
List 2 What should be considered in data modeling lifecycle management 55

LIST OF TABLES

Table 1 Structure of the master's thesis 12
Table 2 Methodological approaches (Lee 1999) 15
Table 3 Interviewed companies, roles and dates 47

ABBREVIATION LIST

API	Application programming interface
CPM	Collaborative Production Management
CPS	Cyber-physical system
CRM	Customer relationship management
E-R	Entity-relationship
ERP	Enterprise resource planning
I4.0	Industry 4.0
IaaS	Infrastructure as a Service
IIoT	Industrial internet of things
IoT	Internet of things
ISA	International Society of Automation
MES	Manufacturing execution system
MIS	Management information systems
M2M	Machine to machine
O-O	Object-oriented
OOP	Object-oriented programming
OPC UA	Open Platform Communications Unified Architecture
PaaS	Platform as a Service
PIMS	Process information management system
RDBMS	Relational database
SaaS	Software as a Service
SME	Small and medium-sized enterprises
SOA	Service-oriented architecture
TSMS	Time series management system
UML	Unified Modeling Language

1 INTRODUCTION

In this chapter the background of the data modeling, Industry 4.0 (I4.0) and process information management system (PIMS) are introduced. The execution and the structure of the study are also presented.

1.1 Background

There are many different data management activities. These activities are all from the ability to make coherent decision on how to receive strategic value from data and the technical deployment to behavior of databases. Therefore, data management requires both, technical and non-technical skills. To ensure, that company has high quality data fulfilling strategic needs, the business and information technology roles should together share the responsibility of managing data. (Henderson et al. 2017)

The main goal of data modeling is to more effectively align applications with current and future business requirements by validating, documenting and understanding different perspectives. Well-performed data modeling can provide benefits in terms of lower support costs and increasing reusability opportunities for future initiatives, and it also reduces the cost of building new applications. Planning more advanced data demands architecture, modeling and other design functions to have a strategic approach. Data models helps organizations to understand their data assets (Henderson et al. 2017). Data modeling has also been understood as data analysis which may cause confusion. Data modeling is a design activity. Data analysis can be characterized as description and design as prescription. (Simsion & Witt 2005)

The I4.0 is changing the industrial landscape requiring, new capabilities for information management. Manufacturing companies are required to take measured data, analyze it, obtain information from it and support with the knowledge of their employees. This is causing difficulties in the companies. One success factor is the flexibility, which enables producing companies to produce and deliver products of high quality and to adapt fast to customer requirements. When the data and the information are visible and available, companies can recognize things similar to each other and perform faster decisions. Manufacturing companies want to become agile companies with reacting in real-time to occurring events and make data-based decisions. (Stich et al. 2017) Advanced information and communication technologies are growing in the industrial automation field and the I4.0 is based on them (OPC Foundation 2021).

PIMS can provide faster and more accurate data and reports to management systems. Therefore, companies are more interested in PIMS to improve the knowledge and intelligence of the production process. Even though PIMS is used by numerous process industries, many small and medium-sized enterprise (SME) are lacking funds and understanding of information management and are neither developing their information management. (Du et al. 2018)

The difference between an information model and a data model is that an information model models managed objects at a conceptual level and a data model describes a lower level of abstraction and contains a lot of details. An information model also defines relationships between managed objects. Even though the data model and information model are used for different purposes, it is not always clear what details should be included in an information model and which ones belong to data model. (Pras & Schönwälder 2003) The benefits of information models are that it can offer a shared, stable and organized structure of information requirements for a domain context (Lee 1999). The study is focusing more on the data modeling but also some information modeling perspective is considered since the relationships are affected by the changing environment.

1.2 Objectives and scope

The aim of this study is to present a picture of what are the required practices and workflows of implementing data model in PIMS. The study examines the existing processes and workflows of creating PIMS in ABB, using data modeling-based approach. The study documents and finds the challenges of a case example of implementing the information model, ABB Ability™ History's equipment model at one metal industry customer. The requirements are examined also from the perspective of industry trends that may affect the data models in the future. The business requirements are constantly changing, and this study presents how the changes affect the data modeling and how they can be considered in the future. The study also examines how the data model is built and developed supporting the flexibility. There are three research questions this master's thesis addresses to get a clear picture on the topics mentioned. Next, each research question is presented.

1. *What are the best-practice processes and workflows to implement a PIMS data model?*

2. *What are the requirements for the software product supporting the proposed best-practice process to create a PIMS data model?*

3. *What are the impacts of Industrial internet of things on data modeling process?*

1.3 Execution of the study

This master's thesis consists of a theoretical and an empirical part. Both parts help to identify the challenges in process data modeling, best-practice process and workflows to implement PIMS data model. The outcomes are evaluated, compared and analyzed in this study. The results of the empirical part are compared with results of the theoretical part. The study is done as case research. The data collection is done using a literature review and qualitative research. The company's internal ABB Ability™ History documents are also used. The theoretical part is conducted as a literature review. The aim of the literature review is to provide an outlook on the topic. In the literature review, the main sources are scientific publications and books. The data discussed in the literature review is not defined but the review is done by concentrating on aspects that may affect process data, and the process data is usually time series-based data. The data discussed in the empirical part focuses more on the process data, meaning the data that is collected from the factories.

The aim of the literature review is to get to know the related literature and to ensure the knowledge about the topic. Qualitative research is a term for various approaches to and methods used for study natural and social life. The data collected and analyzed is usually non-quantitative in character; these are, for instance interview transcripts and documents. (Saldana 2011)

The empirical part is conducted as a qualitative research. The qualitative semi-structured interview was conducted with eight ABB employees and two ABB's customers. The aim was to get a view and development ideas regarding data modeling and industry trends, as well as to answer the research questions. The questions are both closed-ended and open-ended to have a wider perspective of the answers and to be able to follow up with specifying questions. The interview structure is constructed based on the literature review topics. The interview requests were sent to all interviewees one to three weeks before the interview with the description of the study and topics. The interviews were conducted via Microsoft Teams and recorded with the permission of the interviewees. After the interview the summary of the interviews were sent to the interviewee for approval.

1.4 Structure of the study

The content and the structure are presented in Table 1. The study starts with a literature review which consists of two parts described in chapter 2 and chapter 3, process data modeling and industry trends. In chapter two, data modeling and its challenges and benefits, lifecycle are identified as well as the process information management system and time series data. Chapter three addresses Industry 4.0, Cloud computing, Open Platform Communications Unified Architecture (OPC UA) and ISA-88 and ISA-95 standards and their effects on data modeling, PIMS and time-series data.

The empirical part of the thesis starts from chapter four which introduces the case company, the ABB Ability™ History platform, including two data models - the variable model and the equipment model. The study concentrates on developing the equipment model and its lifecycle. The chapter also includes the introduction to the current process of implementing the equipment model. The workflow is built based on the interviews and discussion with ABB's employees. The interview includes questions regarding the implementation of the equipment model and it refers mostly to the process at Outokumpu.

Chapter five includes analyzing the interview answers and results. In this chapter, the results of each interview topic are discussed. Chapter six contains suggested improvements. It includes the evolved data modeling process with equipment model based on the results of theoretical findings and interviews. The last chapter, Conclusions summarizes shortly all the findings, answers to the research questions and it also covers limitations of the research and future research ideas.

Table 1 Structure of the master's thesis

INPUT	CHAPTER	OUTPUT
Objectives of the thesis	**Introduction**	Introduction of the topic and research questions, description of the research methods.
Theory of process data modeling	**Process data modeling**	Definition of data model, data modeling, time series and process information system management.
Theory of industry trends	**Industry trends**	Definition of Industry 4.0 and understanding of its effects on data modeling.
Description of ABB Ability™ History and discussions with ABB's employees	**Implementing ABB Ability™ History Equipment model**	Description of ABB Ability™ History and equipment model and introduction to current data modeling process with equipment model.
Interviews	**Results**	Description on interviewees views and ideas regarding data modeling process.
Theoretical and empirical results combined	**Suggested improvements**	Introduction of evolved process for equipment model.
Assessment of results	**Conclusions**	Comprehensive look at the findings. Answers to research questions.

2 PROCESS DATA MODELING

Data models provide a general vocabulary around data. They also collect and record specific information concerning an organization's data and systems and act as primary means of communications during projects. Another advantage of data model is that they offer a starting point for customization, integrations, or even replacing an application. The purpose of data modeling is to validate and document the understanding of various aspects. This leads to applications that better meet current and future business requirements and lays the foundation for the successful implementation of large-scale initiatives such as information management programs. Data model is making data easier to consume by demonstrating the structure and relationships in the data. (Henderson et al. 2017)

2.1 Data model

The data model is specifying the database, defining what kind of data it includes and how it will be organized. Data models can be done in different ways since there is not one correct answer how to design the data model. (Simsion & Witt 2005) Data model can be called a map to help understand data structure within the environment for professionals, project managers, analysts, modelers and developers. Data models are the result of the modeling process and mainly used to convey data requirements from business to IT and in the IT field from analysts, modelers and architects to database designers and developers. Data model consists of symbols with text labels that represent data requirements visually. Data models include metadata necessary for data users. Data models are a form of metadata. Other information management functions can benefit from metadata revealed during the modeling process. Metadata defines what data an organization has, what it represents, how it is categorized, where it came from, how it works in the organization, how it evolves during use, who can and cannot use it and whether it is high quality. (Henderson et al. 2017)

An information model includes concepts, relationships, constraints, rules and operations to define data semantics for the selected discourse domain. (Lee 1999) Comparing to data model which according to Henderson et al. (2017) contains a set of components which can be for example entities, relationships, facts, keys and attributes. An attribute is a feature that describes, defines or measures an entity. An attribute in an entity can be a column, a field, a tag or a node in a table, a view, a document, a graph or a file. An organization collects information about an object which is the entity. An entity can answer to who, what, when, why, or how. Entity-type refers to a type of something that is being represented. Occurrences or values of a specific entity are entity instances. An association between entities is a relationship. (Henderson et al. 2017).

According to Lee (1999) there are three modeling methodological approaches which are the entity-relationship (E-R) approach, the functional modeling approach and the object-oriented (O-O) approach (Table 2). Though, Henderson et al. (2017) is listing six most used methodologies to represent data which are relational, dimensional, object-oriented, fact-based, time based and NoSQL. E-R is the most used data modeling approach for database applications compared to O-O approach (Halpin 2001).

The E-R approach concentrates on how the concepts of entities and relationships can be used to specify information requirements. E-R approach uses the graphical notation technique. (Lee 1999) E-R model does not organize data into tables but stores data as relationships. (Pfrommer et al. 2016). The approach views an application as entities that have attributes and are involved in relationships. There can be a lot of variations of E-R approach and that may cause issues, since there is no one single standard. (Halpin 2001)

O-O modeling is an approach that covers data and behavior in objects. It is typically used for designing codes for object-oriented programs but it is also used to design databases. (Halpin 2001) O-O approach concentrates on first recognizing objects from the application domain and then on operations and functions. The base of O-O approach is the object, including data structures and functions. O-O model consists of object classes, attributes, operations and relationships. (Lee 1999) O-O and E-R relationships can be combined with extending O-O with relations as first-class concepts. On the other hand, O-O can be built starting from triple-relations as the underlying abstraction. (Pfrommer et al. 2016)

The functional approach focuses on defining and decomposing system functionality. The data-flow diagram is usually used in this approach to show the transformation of data when it flows through system. The diagram includes processes, data flows, actors and data stores. The functional approach uses object and functions when addressing system's processes and the flow of the information from one process to another. (Lee 1999)

Table 2 Methodological approaches (Lee 1999)

Methodological approaches	Consists of
E-R approach	Stores data as relationships
O-O model/approach	Object classes, attributes, operations and relationships
Data-flow diagram (functional approach)	Processes, data flows, actors and data stores

The appropriate modeling methodology should be chosen in the beginning of the design work. The information model always includes entities, attributes and relationships. Each information model, though, has a different viewpoint. The viewpoint defines the information modeling methodology type that should be used. The E-R approach should be selected when data requirements are really detailed. However, the E-R model can lack preciseness in supporting the detailed levels. The functional approach should be used when functions are the most important and more complex than data. Though, the O-O approach could be more easily extended and be more compatible with the planned implementation environment. The data requirements of the application are commonly changed and the changes usually concerns the functions. In functional approach the amount of changes could be high. In the O-O approach, data and functions should be considered carefully and not just from the data perspective. (Lee 1999)

2.2 Data model life cycle

Data modeling consists of finding, analyzing, and reviewing data requirements, as well as building a data model representing and communicating the data requirements. Before initiating the modeling process, organizations should identify and document how their data fits together. Designing how the data fits together is more the modeling process. (Henderson et al. 2017) According to Simsion and Witt (2005), data modeling can be considered as a design activity (Figure 1), not only a process of documenting requirements, as it is seen sometimes. Even though data modeling is a design activity, a set of business requirements is set for the data model. Generally, the data modeling task consists of analyzing the business requirements and then designing a response to those requirements. The design starts before full understanding of the requirements in real life. (Simsion & Witt 2005)

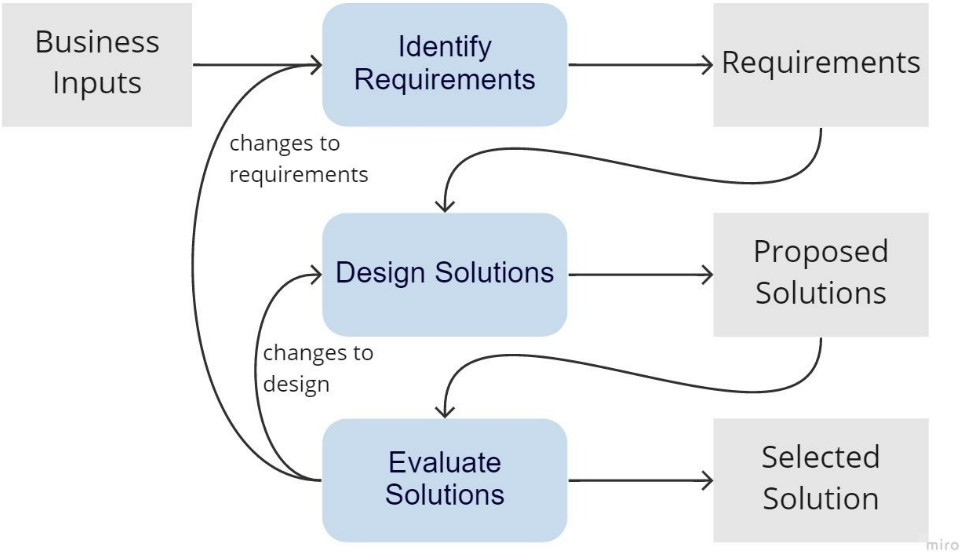


Figure 1 Data modeling as a design activity (adapted from Simsion & Witt 2005)

Henderson et al. (2017) lists seven data stewards. Data stewards are responsible for data and processes that ensure effective control and use of data assets. The data stewards are constructing and managing metadata, documenting rules and standards, managing data quality issues and performing operational data governance activities. The types of data stewards are Chief Data Stewards, Executive Data Stewards, Enterprise Data Stewards, Business Data Stewards, Data Owner, Technical Data Stewards and Coordinating Data Stewards (Figure 2). Chief Data Stewards can lead data governance. Executive Data Stewards are senior managers in Data Governance Council. Enterprise Data Stewards controls data domains of every business functions. Business Data Stewards are business professionals responsible for a subset of data. They specify and manage data with stakeholders. A Data Owner is a business Data Stewards and approving decisions about data within their domain. Technical Data Stewards are IT professionals, for example, Data Integration Specialists, Data Administrators, Business Intelligence Specialists, Data Quality Analysts or Metadata Administrators. Coordinating Data Stewards lead and represent both team of business and technical Data Stewards. (Henderson et al. 2017)

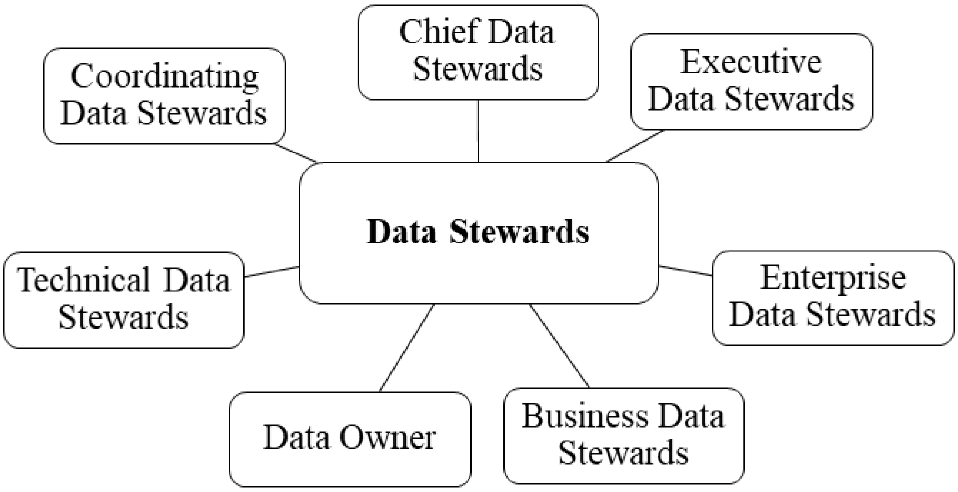


Figure 2 Data Stewards (Henderson et al. 2017)

According to Lee (1999) the development process of information model includes defining the scope, information requirements and a specification and then building the model. Data models need to be planned, built, reviewed and maintained (Figure 3). The planning includes tasks such as evaluating organizational requirements, constructing standards and deciding on data model storage. (Henderson et al. 2017) The data use cases should also define which are the types of data that the system is expected to manage. They are for example ways in which the information is presented. (Halpin 2001)

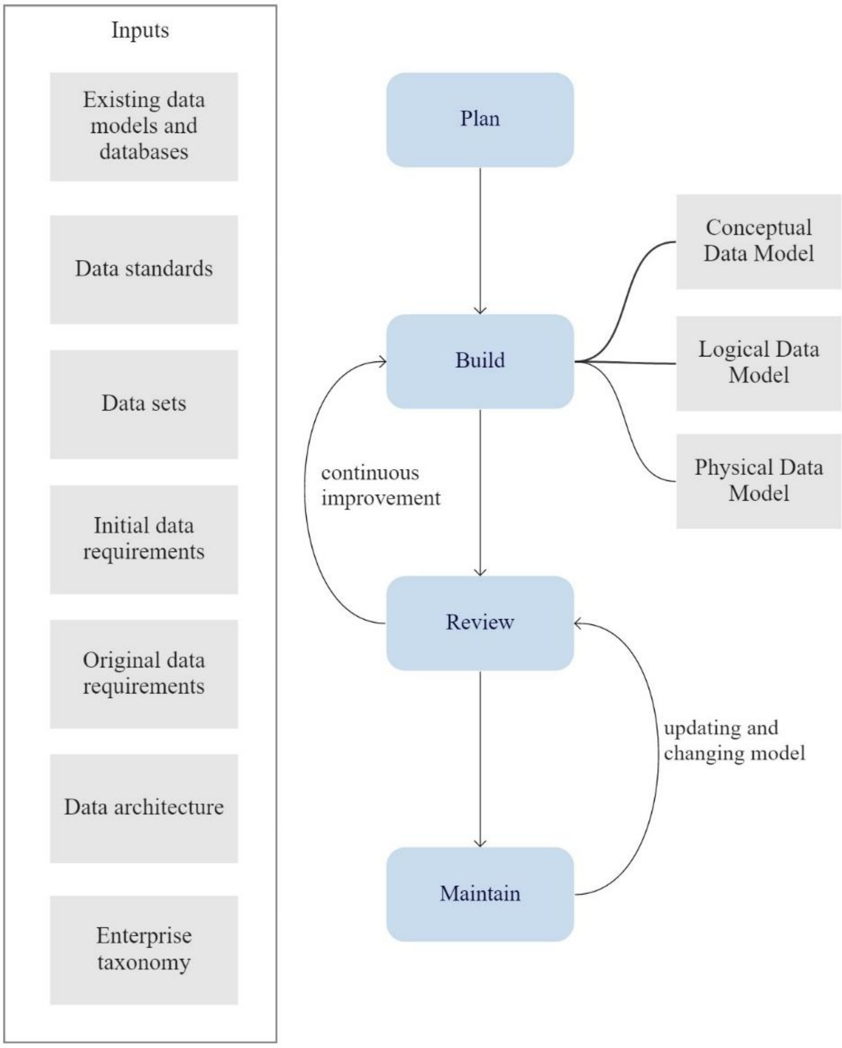


Figure 3 Data modeling activities (adapted from Henderson et al. 2017)

The developing starts with defining the scope of the information model's applicability (Figure 4). The scope specifies processes, information and constraints that fulfil the industry need. The scope statement consists of purpose and viewpoints of the model, the type of the product, the type of data requirements, the supporting manufacturing scenario, the supporting manufacturing activities and the supporting stage in the product life cycle. (Lee 1999)

The step 2 is requirement analysis. The data requirements should also be collected for the application scope. (Lee 1999) The interviews and workshops are mostly used techniques in collecting the requirements. People who understand the requirements of the system and people who might have something to say should be invited for interviews and workshops when collecting requirements. (Simsion & Witt 2005)

Simsion and Witt (2005) are describing the requirements phase and its deliverables from two perspectives. The first perspective is that there are not separate requirements phase and associated statement of requirements. The requirements are described during data modeling process and defined in the data model. This approach is mostly used in practice and might cause confusion whether the purpose of data modeling is to document all data structures. This approach is typically used, for example, when most of the requirements are well-known to the designer and customer and there is no need to document them all or after the customer has seen the design, there may come new requirements. The second view is that the requirements should be developed completely according to the business needs so that there is no need to refer to the customer. This approach might not be practical but, for instance, when the business has already high-level directions and rules that affect the design of the data model but cannot be described directly using data modeling constructs. Another situation could take place when the requirements should be documented in another form than in data model to be able to trace the changes easily. (Simsion & Witt 2005)

When the scope and information requirements are defined, developing the model follows. The conceptual model is developed based on the information requirements. The model should fulfil the data needs of the application. (Lee 1999) The idea of presenting the model first as conceptual level is that people can easily work with it (Halpin 2001). Requirements planning and analysis activities include the conceptual data modeling and logical data. The conceptual data model collects high-level data requirements as a collection of relevant concepts. It includes only basic and critical business units within a specific area and functions and a description of each unit and the relationship between the units. (Henderson et al. 2017) The conceptual data model helps the communication between data modeler and business stakeholders as it is frequently described as diagram. (Simsion & Witt 2005)

When the conceptual design is done, it can be mapped to logical data model. Logical data models are for example network, hierarchic, object-oriented and relational approaches. (Halpin 2001) A logical data model is a detailed description of data requirements, frequently in support of a specific usage context, like application requirements. Logical data models are independent of technology or implementation constraints. When the conceptual data model is expanded by adding attributes, it becomes a relation data model. Attributes are determined for entities by applying normalization techniques. A dimensional logical data model is in multiple situations a fully- attributed perspective of the dimensional conceptual data model. The logical relational data model collects the business rules of the business process, while the logical dimensional captures the business questions to determine the condition and performance of the business process. (Henderson et al. 2017)

After the logical data model is developed, it is possible to build the physical data model. The physical data model is a detailed technical solution and built for specific technology. Physical data modeling is a design activity. Building data model is an iterative process since the modelers return the draft of the model to the business analysts to clarify terms and business rules. Then the model is updated and more questions are asked. (Henderson et al. 2017) The physical data model includes all the necessary changes to accomplish sufficient performance and is also presented in the form of tables and columns. Also, it includes a specification of physical storage and access mechanisms. (Simsion & Witt 2005)

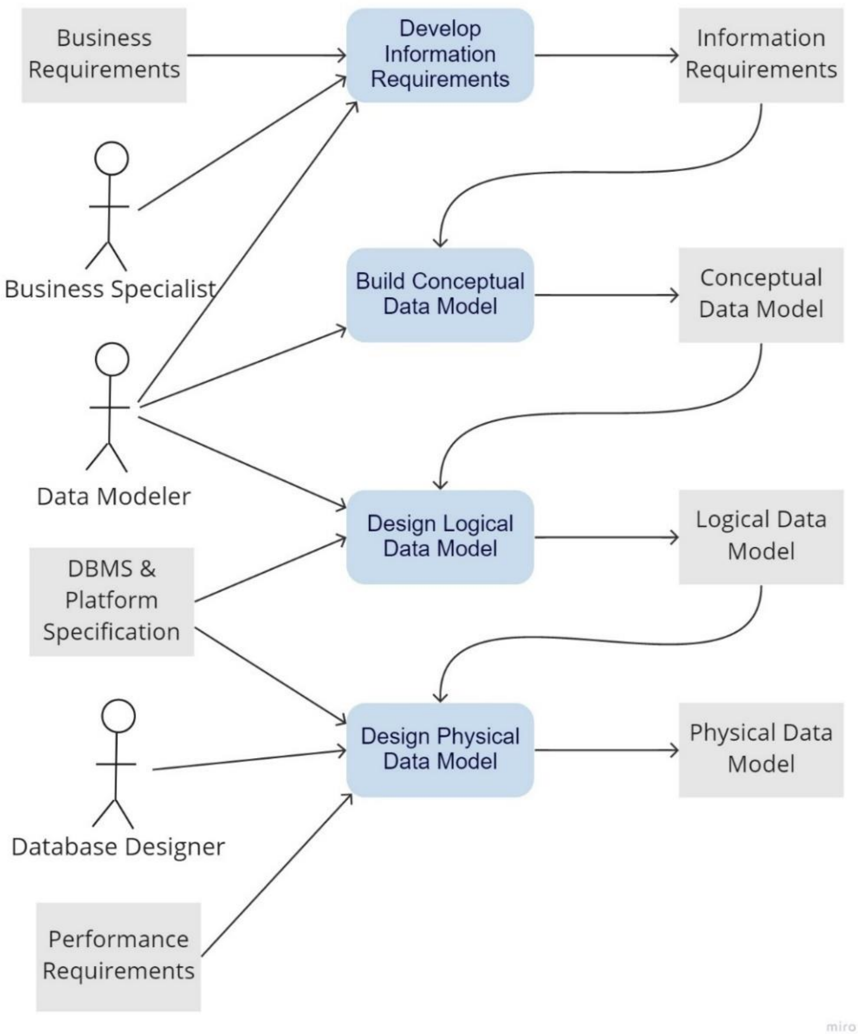


Figure 4 Database design tasks and deliverables (adapted from Simsion & Witt 2005)

After the model has been built, it should be reviewed and, once approved, maintained. The maintaining should be done when the business requirements change or when the process changes. One model level can have many changes. When, for instance, adding attributes, they should be atomic - containing one piece of data that cannot be separated into smaller pieces. (Henderson et al. 2017)

Sometimes the actual data model, which is the logical database structure, should be changed because of new requirements or changes in business. This is a big challenge when implementing the changes to the database and the consequent changes to the applications. The bigger problem is to assure the ongoing usefulness of archived data, which remain in the old format. Frequently, the copies of the original applications and all data conversion programs are archived. (Simsion & Witt 2005) The changes should be recorded, and a change control kept, like with requirements specifications. Each change should be explained with why the project required the changes, what and how objects changed, when the change was done, who made the change and where the change was made. (Henderson et al. 2017)

The changes are simpler and cheaper to implement to the data when it is well-designed. Therefore, data organization is important because small changes made to a data model affect the applications in consequence. Many applications are using the data from database for example for updating, deleting and displaying it. With modern database management software, the database can be organized to match the new model without major difficulty. The modifications impact the rest of the systems, for example the report formats should be redesigned according to the modifications. Though, the changing of the database may be straightforward, they affect all the applications which uses the data. A data model is stable when it does not need to be modified when requirements change. A data model is flexible if it can be easily extended to meet new requirements with only a minor impact on the existing structure. (Simsion & Witt 2005)

2.3 Time series data

Data collection processes are increasing fast since the use of embedded systems and sensor networks is growing. These collection methods are giving the opportunity for collecting large amounts of data. After the data is collected, the information systems are analyzing and processing the data. The collected data instances have a timestamp. The data with specific timestamps are formalized as time series. (Llusà Serra et al. 2016) Time series is a sequence of data points, for example a series of numbers. The numbers are collected with certain intervals within a period of time. Generally, time series consists of successive measurements made over a time interval. (Namiot 2015) A data model and a group of operations are the components of time series management system (TSMS). With the operations the time series can be manipulated. For example, the relational model operations, operations over time series avoid the actual semantics of the data. In a real application, it must be determined whether the function is semantically consistent or not, so if it should be applied. For example, adding values from two different phenomena may be semantically incorrect. Set operations that consider time series as sets, sequence operations that consider time series as sequences and temporal operations that manipulate time series assuming they are representations of functions. (Llusà Serra et al. 2016)

Time series is a sequence of data points, for example a series of numbers. The numbers are collected with certain intervals within a period of time. Generally, time series consists of successive measurements made over a time interval. (Namiot 2015) The time series consists of collected observations at specific timestamp. A data model and a group of operations are the components of time series management system (TSMS). With the operations the time series can be manipulated. (Llusà Serra et al. 2016) While the usage of time series is growing, the observation amount is growing and this causes time series database development. Time series databases are integrated with other systems and usually offer web service-based interfaces. (Leighton et al. 2015)

2.4 Process information management systems

Monitoring and control systems are affected by the changing supplies and energy of industrial production process and PIMS can address these problems. PIMS is a platform focused on a process, for information integration and management, designed to cooperate with common control systems. PIMS has many functionalities (Figure 5) like collecting process information, real-time process view, trends display, alarm record, historical trend report generation, data storage, production statistics report generation under the network environment. PIMS is a bridge between production process data and data users. PIMS collects and processes field data and then transmits it to customers, the data users. (Du et al. 2018)

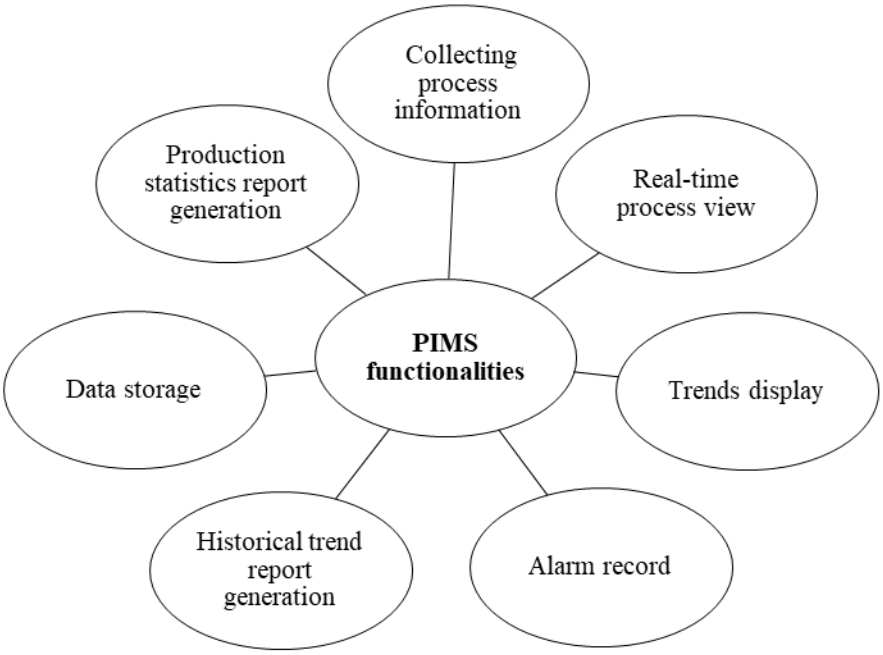


Figure 5 Examples of PIMS functionalities (Du et al. 2018)

PIMS can integrate various existing control systems into an information platform providing field data for more advanced management networks, like customer relationship management (CRM) and management information systems (MIS). This means that PIMS helps management system data and reports to be more up-to-date and accurate, which improves data and intelligence of production process. (Du et al. 2018) PIMS should be well implemented and designed to ensure that all data is collected and stored. PIMS provides a platform for advanced tools that can be successfully deployed in the future. One example of the PIMS benefit is that is provides a tool to help with long-term optimization of the process. With PIMS the processes can be optimized and continuously improved. (Muza 2005)

3 INDUSTRY TRENDS

Internet of things (IoT) describes technologies that have not been connected and are now connected to an IP-based network (OPC Foundation 2021). Industrial internet of things (IIoT) means using these networked technologies in industrial applications (Pfrommer et al. 2016). With IIoT technologies, the inefficient processes can be found to be developed and turned into working capital. IIoT offers new challenges and opportunities for manufacturers regarding machine and plant design. The new opportunities include benefits like connectivity, efficiency and reliability and these usually lead to financial benefits. The machine and plant data can be communicated better and more efficiently with IIoT technologies. This combined to analyzing historical data helps finding the inefficiencies in production and improve them. (Neubert 2016) I4.0 is related to IoT and the Internet of services becoming integrated with the manufacturing environment. The future goal of this fourth industrial revolution is for industrial companies to create global networks to connect machines, factories and warehouses to cyber-physical systems (CPS). CPS intelligently connect and manage each other by sharing information that triggers actions. This will improve the industrial processes in manufacturing. I4.0 will demand integration of CPS in manufacturing and logistics. CPS are integrations such as computation, networking and physical processes. This differentiate microprocessor-based embedded systems from more complicated computing systems that integrate with their environment. (Gilchrist 2016) I4.0 platform has embedded and collaborative intelligence like smart factories, smart services and smart devices and it is a part of IoT world (Neubert 2016). Scientifically speaking, I4.0 can be described as real-time, multilateral communication and data transmission between cyber-physical devices with high data volume rates. The main benefit of I4.0 for companies is the transformation to an agile and learning company to be competitive in a growing dynamic business market. (Stich et al. 2017)

While implementing I4.0 successfully, the obtainable data should be prepared and processed in a way that it supports making decisions. The data may be useful if the technical requirements for real-time access are met and if there is an infrastructure with the necessary data processing and seamless data transmission. Another principle for successful I4.0 implementation is that manufacturing companies need an IT integration to improve data use and increase agility. (Stich et al. 2017)

There are nine technological trends that are forming the Industry 4.0 (Figure 6). They are big data and analytics, autonomous robots, simulation, horizontal and vertical system integration, the IoT, cyber-security, the cloud, additive manufacturing and augmented reality. The big data and analytics are a big part of the I4.0, since the data amount is increasing from different sources in manufacturing and there is a need to collect all the data, assemble and organize it in a coherent manner and then use it in analytics to support decision-making. (Gilchrist 2016) The big data analytics is also causing challenges regrading speed, space and automation when it comes to data collection, processing, transportation, integration, transformation, storage, computation and knowledge extraction from big data. (Khan et al. 2017)

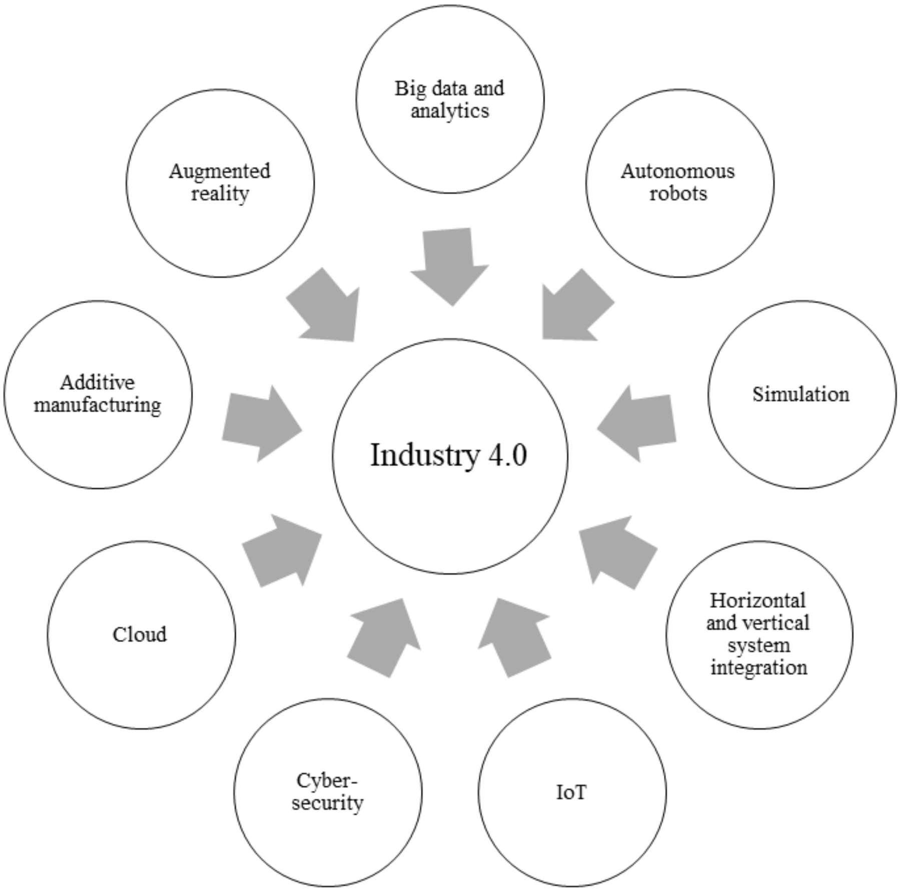


Figure 6 Technological trends forming Industry 4.0 (Gilchrist 2016)

The six design implementation principles for I4.0 systems (Figure 7) are interoperability, virtualization, decentralization, real-time capability, service orientation and modularity which are used in automation and digitization for production processes. The interoperability requires the whole environment with flexible collaboration between all parts. The virtualization is about linking physical processes and machinery and returning sensor data to virtual models. This way process engineers and designers can, for example, test changes through the virtualized processes without affecting the physical processes. Decentralization allows intelligent factories' versatile systems to make decisions independently without deviating towards a single, ultimate organizational goal. Production process, collecting data and the feedback and the monitoring of processes should be achieved in real-time, as it supports the idea of making everything real time. Internet of Things creates services that others can use, which leads to the fact that internal and external services are required by smart factories. That is why Internet of Services is an significant part of I4.0. Modularity means that a smart factory should be flexible, so that is can easily adapt to changing circumstances and requirements. With modularity changing production and replacing individual product lines are flexible. (Gilchrist 2016)

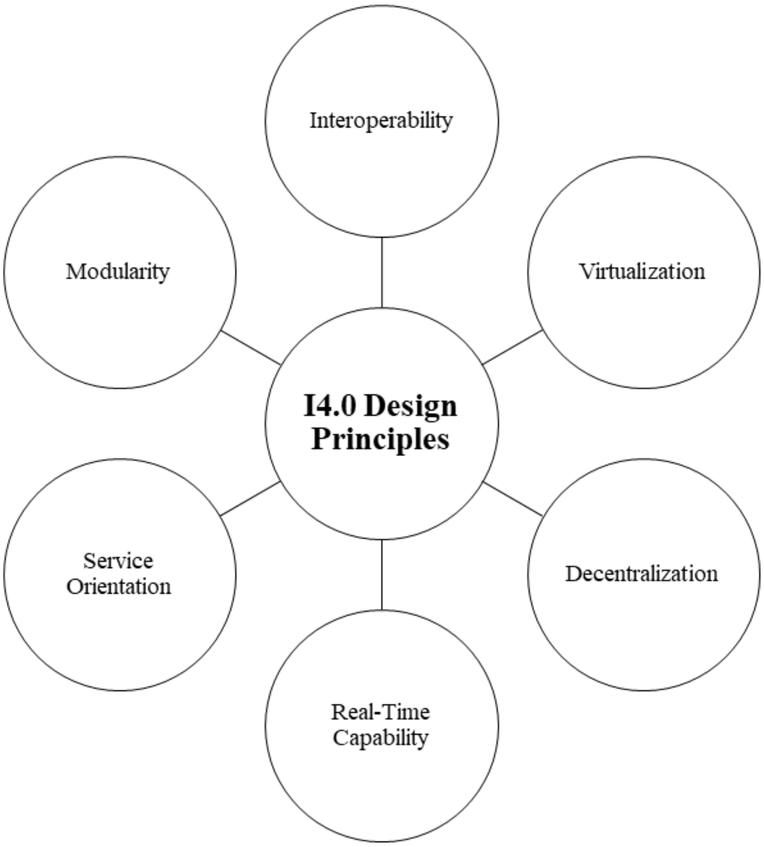


Figure 7 Industry 4.0 Design Principles (Gilchrist 2016)

3.1 Cloud computing

IoT cloud computing architecture plays a big role in IoT data management. IoT data and applications are stored in the cloud for easy access in any client software web browser. The cloud computing architecture suits I4.0 because of its centralized control accessibility for various users like managers, customers, operators and programmers. (Khan et al. 2017) Those who want to analyze and access data from machines, systems or other products over the internet are not able to know the amount of customers who will benefit from such services when implementing an internet-based service in public cloud. It is similar to the suppliers of consumer goods or services for end users. The public cloud is open to every end user on the internet, while the private cloud is only available for a defined group of people. The cloud computing is a key feature in IoT. (Sendler 2018)

Cloud computing technology is about improving the provisioning of computing resources. The main improvement is that the location of resources is moved to the network to reduce costs regarding the management of hardware and software resources. Cloud computing simplifies hardware provisioning, hardware purchasing and software deployment. Therefore, there are many benefits to deploying data-intensive applications, such as resource flexibility and detection of unlimited resources and endless scalability. (Zhap et al. 2014)

Cloud computing’s characteristics are on-demand self-service, broad network access, resource pooling, rapid elasticity and measured service. On-demand self-service means that the customer can unilaterally provide billing functions like server time and online storage automatically, when needed, without the need for human interaction with each service provider. The second characteristic describes the availability of capabilities over the network and accessed through conventional mechanisms that promote the use of heterogeneous thin or thick customer platforms. Resource pooling is about combining the provider’s computing resources. This enables numerous customers to use a multi-tenant model and various physical and virtual resources are dynamically designated and reassigned according to consumer demand. Location independence is that the customer frequently does not have control or knowledge of the specific location of the resources provided. Rapid elasticity means that capabilities can be quickly and elastically conducted, even automatically. The features available to the consumer to provide services often seem to be limitless and can be purchased at any time. The last characteristic is the measured service. Cloud systems automatically control and optimize the use of resources by utilizing a measurement function at a certain level of abstraction, according to the type of service such as storage or processing. Resource use can be monitored, managed and reported by providing transparency to service provider and consumer. (Zhap et al. 2014)

Satyanarayana (2012) defines cloud computing as a model for enabling ubiquitous, appropriate, on-demand network access to a shared pool of configurable computing resources. Which can be conducted and released quickly with minimal management effort or service provider interaction. Cloud services includes three service models, which are Infrastructure as a Service (IaaS), Platform as a Service (PaaS) and Software as a Service (SaaS) (Satyanarayana 2012). The three models are built on each other. The internet-based software requires a platform that allows it to run on the internet and the platform requires an infrastructure of servers that meet its requirements. (Sendler 2018)

In IaaS, clients rent a virtual server for their own needs. They can increase or decrease capacity as it meets their requirements. In PaaS, the clients include, for instance, software application vendors that use ready-made platform as development, testing and runtime environment. The platform is more virtual than the infrastructure model, in which the client is a least allocated to a virtual server of differing capacity. In the platform model, the client has no idea which server structure will runs part of its application. (Sendler 2018) SaaS is a software distribution model where vendor or service provider host the applications. SaaS is made available to wide variety of customers hosted in its cloud infrastructure, usually available in Internet (Satyanarayana 2012, Zhap et al. 2014). The applications are available for different customer devices through a thin customer interface, like web browser. Managing the cloud resources or individual application capabilities does not belong to the customer (Zhap et al. 2014). SaaS is becoming an increasingly common delivery model as underlying technologies supporting web services and service-oriented architecture (SOA) evolve and new development methods become popular. SaaS is a development platform but also a resource platform where all data and software can be used as services. (Satyanarayana 2012)

3.2 Open Platform Communications Unified Architecture (OPC UA)

OPC UA is the data exchange IEC standard for safe, reliable, manufacturer- and platform-independent industrial communication. OPC UA provides one standardized way to do information modelling, especially in the context of automation industry. OPC Foundation is a non-profit organization and develops the standard with users, manufacturers and researchers. (OPC Foundation 2021) OPC UA applications are trying to cover many levels of automation pyramid, from the field level to management level like enterprise resource planning level (ERP) (Graube et al. 2017). OPC UA standard is used for exchanging information and controlling industrial domains, like manufacturing system domain and power system domain (Lee et al. 2016). OPC UA is not only a transport protocol for industrial applications, it also specifies how information is encoded and specifies the semantics that allow that data to be interpreted. Unlike the classic OPC which only offers the ability to represent basic data, the OPC UA provides mechanisms for revealing certain semantics to the data. For instance, information about the sensor type of the device that implements the sensor functionality can be modeled, in addition to the measurement value of the sensor. (Graube et al. 2017)

OPC UA is considered a reference standard for the communications inside I4.0. OPC UA has an important role in industry environments and it is an approved protocol that harmonized the interaction of machine to machine (M2M). OPC UA is one of the key candidates to lead the standardization of current and future frameworks and systems integration. (Cavalieri & Salafia 2020) OPC UA is important for CPS communications since it covers both middleware communications technology and broad data modeling framework for digitalization and I4.0. An integral part of OPC UA are concepts to provide flexible and secure communication but its most important enabling feature are the large data modeling capabilities and the capability to communicate information rich in semantic content. (Graube et al. 2017) Expertise from various domain experts is required when developing and implementing OPC UA interfaces for virtual presentation as a “virtual twin” of the systems. Especially, the design and implementation of the data model requires a lot of work and increases the initial investment required to know OPC UA technology. (Pauker et al. 2016) Interoperability of equipment from various suppliers required a harmonized presentation of data. Thus, the OPC UA has specific information models for varied application domains. These can be used directly or by extending them with suppliers’ own domain specific knowledge. The end customers can trust in all OPC UA servers to have the same base model that reveals their data. (Graube et al. 2017)

OPC UA is used in industrial communication and it describes a meta-model for information modeling that uses triple-relations to represent object-orientation (Pfrommer et al. 2016). By default, OPC UA defines the information model, the message model, the communication model and the conformance model. With these models, it is possible to exchange messages between clients and server over different network environment for various types of systems and devices. The structure, behavior and semantics of UPC UA server are represented in OPC UA information models. (Lee et al. 2016) The data models are based on references between nodes with attributes. With OPC UA, existing information models can be extended or reused. This happens by creating new individuals' nodes, modifying or extending nodes in an existing namespace by creating a new namespace with reference to the existing ones. The later versions allow maintainable models. (Graube et al. 2017) Both customer/server and publish/subscribe communication models form the standard OPC UA and it is a semantically improved information model for presenting data. OPC UA Information Model offers a standard way for servers to disclose information to customers. The OPC UA Information model is based on object-oriented programming (OOP). Which means that some nodes describing instances inherit from other nodes determining types; numerous inheritances are not suggested in OPC UA, although the definition does not limit type hierarchies to a single inheritance. (Cavalieri & Salafia 2020)

The OPC UA specification consists of 13 parts. The first seven parts concern the core specifications like the concept, security model, address space model, services, information model, service mappings and profiles. The following six parts concern the access type specifications such as data access, alarms and conditions, programs, historical access, discovery, and aggregates. The meta model of OPC UA is the address space model described in Part 3 of the specification. The nodes are the base component of the meta model. Node classes that specialize in the base node class are defined as objects and variables. Every node has a set of attributes, depending on the node class. Some attributes are required and others are optional. In addition to the communication part, data modeling is the second basis of the OPC UA. The base principals allow to form a simple but also complex OPC UA information models (List 1). (Pauker et al. 2016)

Base principals of UA modeling

- Object-oriented techniques with type hierarchies and inheritance
- Type information and an instance are provided and accessed the same way
- Full meshed network of nodes allows to connect the information in different ways
- Extensibility of the type hierarchies as well as the reference types between nodes
- No limitations of modeling to allow an appropriate information model design

List 1 Base principals of UA modeling (Pauker et al. 2016)

3.3 International Society of Automation (ISA)

The ISA-88 standard of International Society of Automation initially addressed batch processes control issues and was later extended to address separate manufacturing and continuous processes. It organizes information from three perspectives: physical model, process model and procedural control and they are all hierarchical representations. The physical model organizes the company hierarchically into locations, areas, process cells, units and devices and control modules. The process model is a multi-level hierarchical model for presenting a high level of batch process and is the basis for defining hardware-independent recipe procedures. The batch process is divided hierarchically in ISA-88 standard to process stages, process operations and process actions. The third hierarchical representation is the procedural control model and it describes the orchestrations of procedural elements for performing process-oriented tasks. (Vegetti & Henning 2014)

The ISA-95 standard consists of five parts. The first part includes standard terminology and state models that can be utilized to decide what information is exchanged. The second part contains attributes for each object defined in part 1. The object and attributes in Part 2 are used for data exchange between various systems and also as a basis for relational databases (RDBMS). The third part focuses on functions and activities at level 3. It is an excellent guide to describe and compare production levels in different places in a standardized way. Part 4 is under development entitled “Object Models and Attributes of Production Management”. The part 5 development has also started and it is about “Business to manufacturing transactions”. (ISA-95 2021)

The ISA 95 standard is titled as “Enterprise-Control System Integration” and as the title describes, the standard is about how Enterprise/Business systems should be integrated with Manufacturing and Control systems. The standard is widely used by vendor and end-user community. (Johnsson 2004) ISA-95 standard is published by the ISA Committee for developing automated communication between enterprise planning and shop floor control systems. (Unver 2012) ISA-95 is an international standard for enterprise and control system integration. The standard can be used by several user groups, such as vendors, end-users and integrators. The ISA-95 common benefit is that a set of common terms and terminology is defined. The standard is also helping the software development team to structure the user requirements very carefully. (Johnsson 2004) ISA-95 includes models and terminology that can be used to determine what information needs to be exchanged between sales, financial and logistics systems, production, maintenance and quality systems. This data is built on Unified Modeling Language (UML) models, which are the basis for the development of standard interfaces between ERP system and manufacturing execution system (MES). (ISA-95 2021) The biggest contribution of the ISA-95 standard is to formalize the interaction of the manufacturing system with the company’s other business processes. The purpose of the standard is to define data flows and interfaces between a company’s business systems and manufacturing control systems using company modeling techniques. (Unver 2012)

ISA-95 is based on hierarchical structure which is defined by 4 different levels. The level 4 is the Business Planning and Logistics, and it includes, for example the plant productions scheduling and operation management activities. The level 3 is the Manufacturing Operations and Control level which consists of activities like production dispatching, detailed production scheduling and reliability assurance. Level 4 and 3 are similar regardless of the type of industry they are used. Level 2 conforms to the process control systems, level 1 to the sensors and actuators and level 0 is the production process itself. Levels 2,1 and 0 differ in the type of industry– batch, continuous and/or separate – which they are used in. (Johnsson 2004) These five levels of functions form the functional hierarchy (Chen 2005).

The equipment hierarchy model of ISA-95 (Figure 8) defines how production resources are frequently constructed and involved in manufacturing. It is an extension of the model defined in IEC 61512 and ISA S88.01 by including a definition of separate and continuous manufacturing. The function hierarchy model defines the different function levels and the area of responsibility for various function levels are specified in the equipment hierarchy model. The model maps physical assets to the function levels. (Chen 2005)

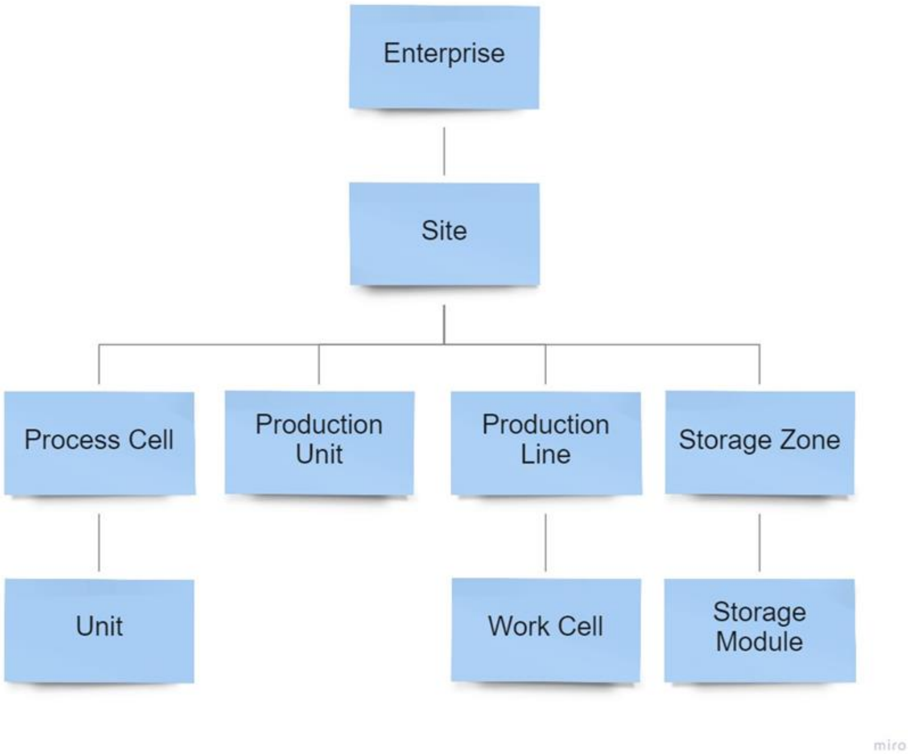


Figure 8 Equipment hierarchy (adapted from Chen 2005)

3.4 Effects of industry trends on data modeling

I4.0. has caused the need of actions on the information management side in companies. A key to ensuring that company data and information is available for decision making is efficient information management. Now the relevance of information management and its influence on production processes are not evident for manufacturing companies. One key capability required by companies in I4.0 is a faster reaction to events achieving agility. (Stich et al. 2017) PIMS is designed to provide a single online, real-time information system that is online and accessible to all parties (Muza 2005). Which means that reacting to the events can be done fast. All effects are listed in Figure 9. Other effects are described below.

In data integration and modeling the interoperability is the key principle of automation in I4.0 because numerous types of devices communicate with each other. For remotely controlled and operated machines where real-time action is needed, the integration of data is highly important. Designing I4.0 and industrial internet applications requires consistent, reliable, scalable and secure data models. All attributes must be defined in the data collection up to the end users. (Khan et al. 2017) When using the time series data in the data model, it is already capable of real time access since the time series describes the time the data is collected. For real time action, the data should be the data that is directly collected at the same time it is generated. This is helping to make decision based on the real-time data. As mentioned before, according to Henderson et al. (2017), data model is for using the data easier by illustrating the structure and relationships in the data.

The role of time-series database in I4.0 is to provide precision monitoring of events even in nanoseconds, using various data sources for monitoring and a context on the data. Another important aspect is the processing of the manufacturing data as well as the need for scalability and open exchange. The manufacturing data can vary a lot in its volume so the core time-series database is required to ingest the high throughput of data and sustain the real-time querying. If the data architecture is not well designed and implemented, it can lead to data silos where the critical data needed to optimize the real-time process is not available. (Hall 2020) Which means that the data model should be also able to handle such an amount of data to receive all the possible benefits of it.

The implementation of I4.0 principles in industrial automation challenges the architecture of the entire system. Classically, automation systems follow the organizational model of an ISA-95 layered architecture, distinguishing between the system and their communications. This strong layering is the result of different requirements in the application areas. Concepts consistent with I4.0, like big data analytics, require greater interconnectivity and harmonization of communications. Therefore, the monolithic ISA-95 system architecture interferes with these ideas and should be changed. (Trunzer et al. 2019) Though, the ISA-95 is used nowadays more as a tool to build systems.

For successfully implementing I4.0, seamless data transmission is needed. Thus, system topology should be well designed and planned before building a data model. Changing requirements and various applications are affecting the data models so they need to be flexible to changes, for instance with modularization. In the future, the information needed for integration is provided by modules in their own way through OPC UA. The changes in a single module or restructuring of the plant need to be traced and that is why the traceability mechanisms must be created. This means that the data model is evolving within changing environmental influences. Changing types and instances are then important to be able to trace. OPC UA supports traceability in numerous ways. One example could be semantic versioning schema which includes the version number in a specific format. Though, they do not offer all information about the changes itself. So, the old version of the model needs to be kept and browse or query after a change event happened and the dissimilarity between the old and new version need to be created to conclude required actions. One solution to this drawback would be the integration of the change itself within the event. Other solution would be to map the concepts already available for review control functions to, for example OPC UA. This is enables the services (Read, Browse, Query) to be used directly in the old version of the data model. This would be really useful in a distributed application. Also, small OPC UA server can rely on using the nodes with the version number and they do not have to worry about the referenced meta model changing and having their own instance information inconsistent. (Graube et al. 2017)

According to Graube et al. (2017) OPC UA is suitable for new applications in digitalization in process industries, because applications can benefit from the power of OPC UA data model. It can be used to build flexible and smart applications. These applications are advantaged by creating new scenario-specific namespaces that are subclassed by existing ones. Still, there are few limitations of OPC UA in software support related to object targeting, server aggregation and data model checking.

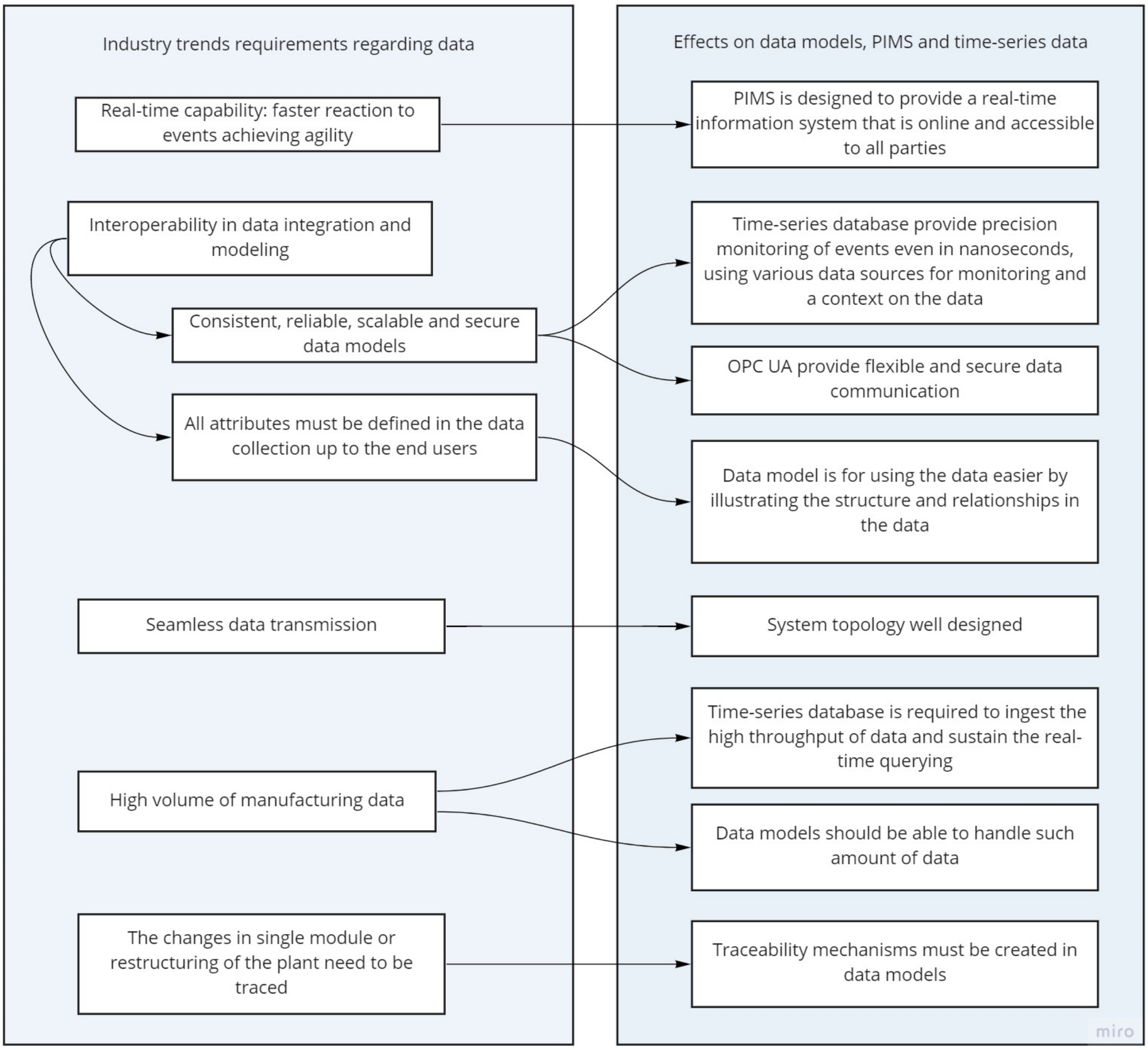


Figure 9 Effects of industry trends on data modeling, PIMS and time series data

4 IMPLEMENTING ABB ABILITY™ HISTORY

In this chapter, the process of implementing ABB Ability™ History equipment model is introduced. The ABB and the ABB Ability™ History time-series database are presented, as well as the two data models in ABB Ability™ History.

4.1 Introduction to ABB

ABB's Process Automation Business Area enables efficient operations that are safer, smarter and more sustainable throughout the lifecycle of customers' investments. The business area consists of five divisions Energy Industries, Process Automation, Marine & Ports, Turbocharging, and Measurement & Analytics. In 2020 21 500 employees were working in ABB's Process Automation. (ABB 2020)

4.2 Introduction to ABB Ability™ History

ABB Ability™ History is a time-series database management system (Figure 10) that is designed and optimized for industrial process information management and history recording. It collects and transfers data accurately with the least possible delay, supporting both on-premises and in cloud implementation. Various enterprises use the ABB Ability™ History from a stand-alone embedded data logger to enterprise-level Collaborative Production Management (CPM) solutions. The platform consists of tools and service and of independent, but integrated, software technology components. The components are functionalities like data acquisition, data processing, data storage, analytics, notification and visualization. The base for all the functionalities are the information models and one of the main models is the equipment model. ABB Ability™ History also provides built-in support for data acquisition from 3rd party data sources like control systems and devices. The database consists of built-in columnar features, optimized for time series signal processing and storage. Customer data abstraction interface allows customers to access to any data source for which a driver is available. (ABB 2021)

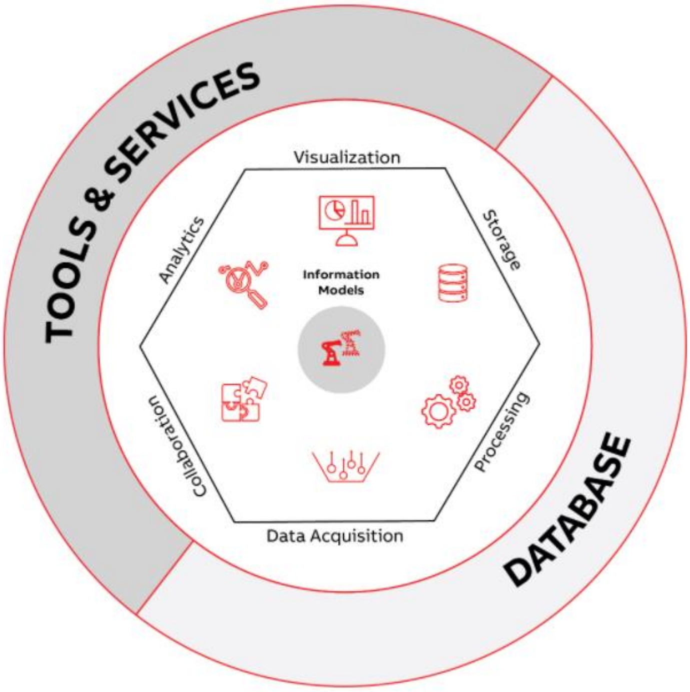


Figure 10 ABB Ability™ History database management system (ABB 2021)

Equipment modelling enhances the data, since it includes a physical description of the physical equipment, the subsystem, a comprehensive physical and functional facility and all available operational data. The real-world assets, defining process and implementing application can be modeled with the equipment model. Combining this with powerful information modeling tools, it enables to query the database to provide on-the-fly status reports, along with time series data and/or speed up the development of model-based applications. The equipment model is a predefined metadata model for modelling industrial assets and processes and implementing applications against them. It contains properties, functions, interfaces and data tunnels. The efforts while communicating between systems and subsystems can be reduced with the unified equipment model presentation. The equipment model allows to collect time series history for an equipment property which is a big advantage, while it is also simple to add new equipment instances to an existing system. (ABB 2021)

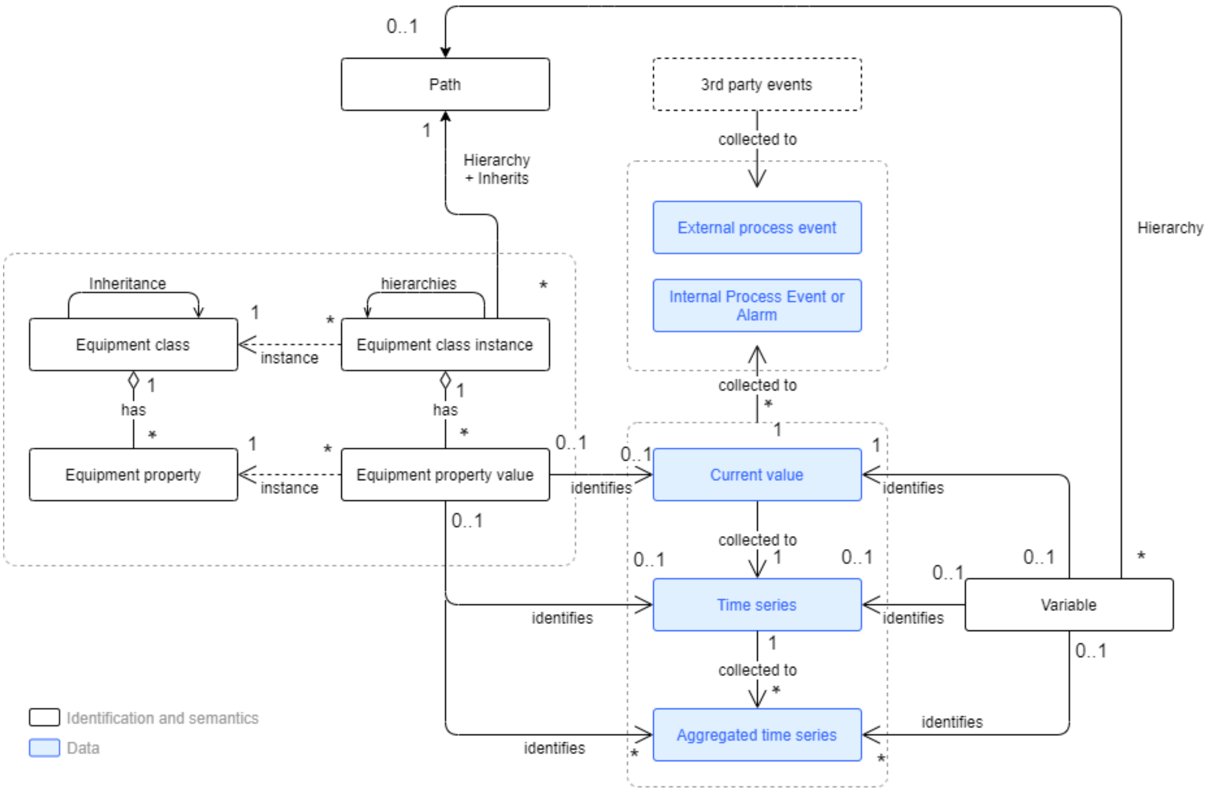


Figure 11 The conceptual model (ABB 2021)

The conceptual model (Figure 11) defines how time series concepts are related in ABB Ability™ History. The ones highlighted in blue color are referring to time series data and the rest are metadata that described the time series data. Current value is the value of a measured or calculated property at the current time and it also contains the quality status and time stamp, describing when is has been measured or produced. Time series represent the history of the current value or the future value in case of predicted value. Statistical aggregate of the time series is the aggregated time series, like one minute time average, one hour maximum or 15 minutes standard deviation. For every variable or equipment instance, there can be any number of aggregated time series. Aggregated time can be collected automatically by ABB Ability™ History or be calculated or produced by an external application. Aggregated time series can be used with all application programming interfaces (API) and used in calculations. (ABB 2021)

Variable is a basic object type used to engineer time series data. Variable represents the measured or calculated property at the current time. Data type, engineering unit, description, value limits, other attributed of current value and how the data is preprocessed and compressed before storing are defined by the variable. The tags form the Variables list, which is not connected to the equipment model. (ABB 2021)

The equipment model supports inheritance to enable model type hierarchies between equipment of the same type. As an example, an “electric motor” type can be derived from a base equipment “electrical apparatus”. The electric motor can be inherited as a “low voltage motor” and so on. Inheritance allows common properties to be automatically available for all the inherited types. (ABB 2021)

Equipment instances can be organized into hierarchies for many purposes and they are created from equipment classes. Each equipment, an instance of an equipment class, is placed at least in one logical hierarchy. The purpose of the logical hierarchy can be a process hierarchy, a location hierarchy or any logical or natural hierarchy. Each equipment instance has a type or non-typed parent instance, except the root. Non-typed equipment instance is called path and it is intended to create structures with equipment instances. The data type specifications for equipment class are equipment properties. Equipment property can be a static value maintained by any basic data type, a reference to another object or a historical property that includes a reference to a current value. Every equipment property has equipment property value. (ABB 2021)

4.3 Current process workflow

The current process of implementing equipment model includes five main steps (Figure 12); define requirements, build data model, connect live data, develop and deploy visualization and analytics and maintain data model. The data model refers to equipment model. Below the flowcharts are describing the main steps and their sub process steps. The inputs of the process are signal list, customers defined hierarchy and use cases. These inputs affect the whole process and the data model is build based on them. The steps with grey colors mean that they are mostly done by the customer.

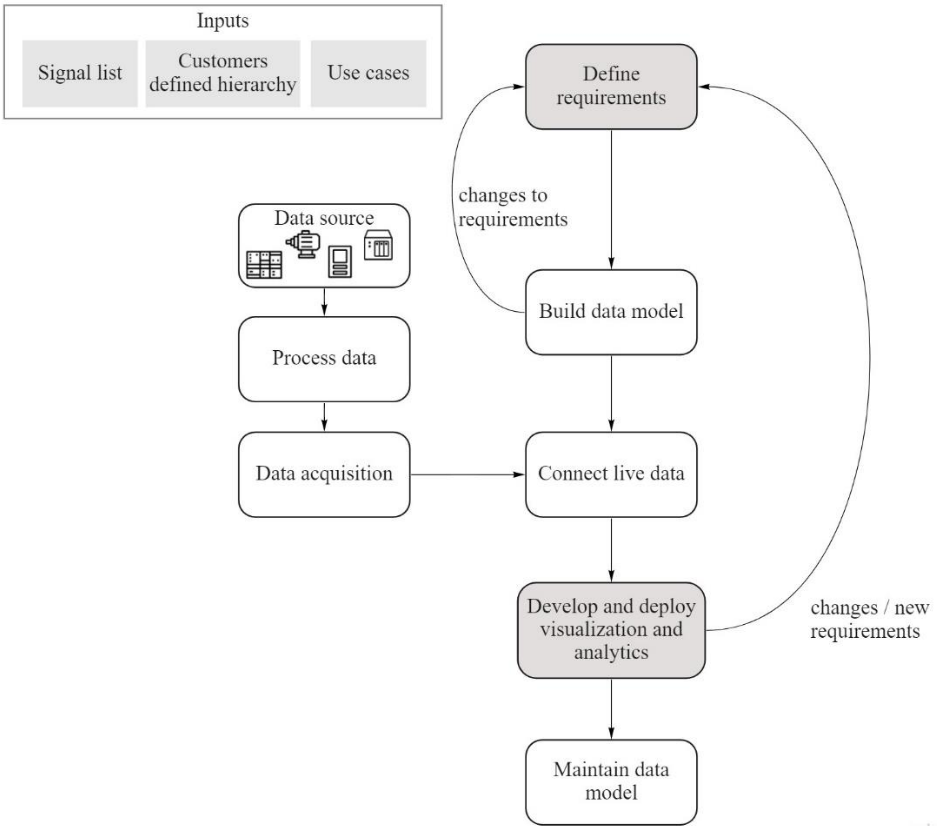


Figure 12 Current data modeling process of implementing equipment model

Defining requirements (Figure 13) includes defining the use cases and the needed signals. These are defined mostly by the customer. After the requirements are defined, the data model is built (Figure 13) by multiple ABB's employees. It starts with forming the equipment model hierarchy based on the customers already defined hierarchy. Then the data signals are grouped manually with the knowledge of the production process that the modelers have into the hierarchy. Based on the grouping the equipment classes and their equipment properties are defined and built. Next the equipment instances are defined and built from the equipment classes. During the building process there might come needs for new signals, especially when the equipment instances are built.

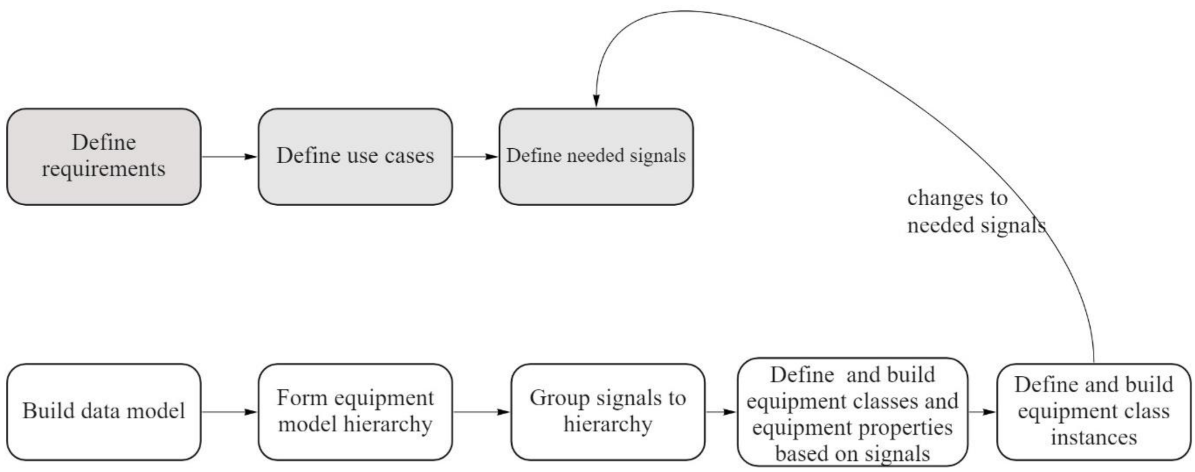


Figure 13 Current process, defining requirements and building the data model

After the equipment model is defined and built, the model is connected to live data. Which means that the signals are connected to each equipment instance property. At this point, the data collection starts and the data begins to accumulate in the ABB Ability™ History real-time database. That means also that the equipment model is ready to be used. Thus, next the equipment model is used by referring to it in for example applications and calculations. If, for example, some properties are missing that are needed in applications, new requirements are defined. This leads to the new requirements being defined, the properties are built to the model and then the live data is connected. After this the property can be used in the application.

Finally, there comes the maintaining part of the whole process (Figure 14). This is the most time-consuming, since after the equipment model is in use by the data users, the changes have more impacts on the applications so they need to be considered carefully and executed with customers approval. The request to change the model comes from customer and the effects are identified with the customer and the data users. The possible steps in maintaining the data model that is considered in this process chart concern modifying equipment instances, equipment classes and hierarchy.

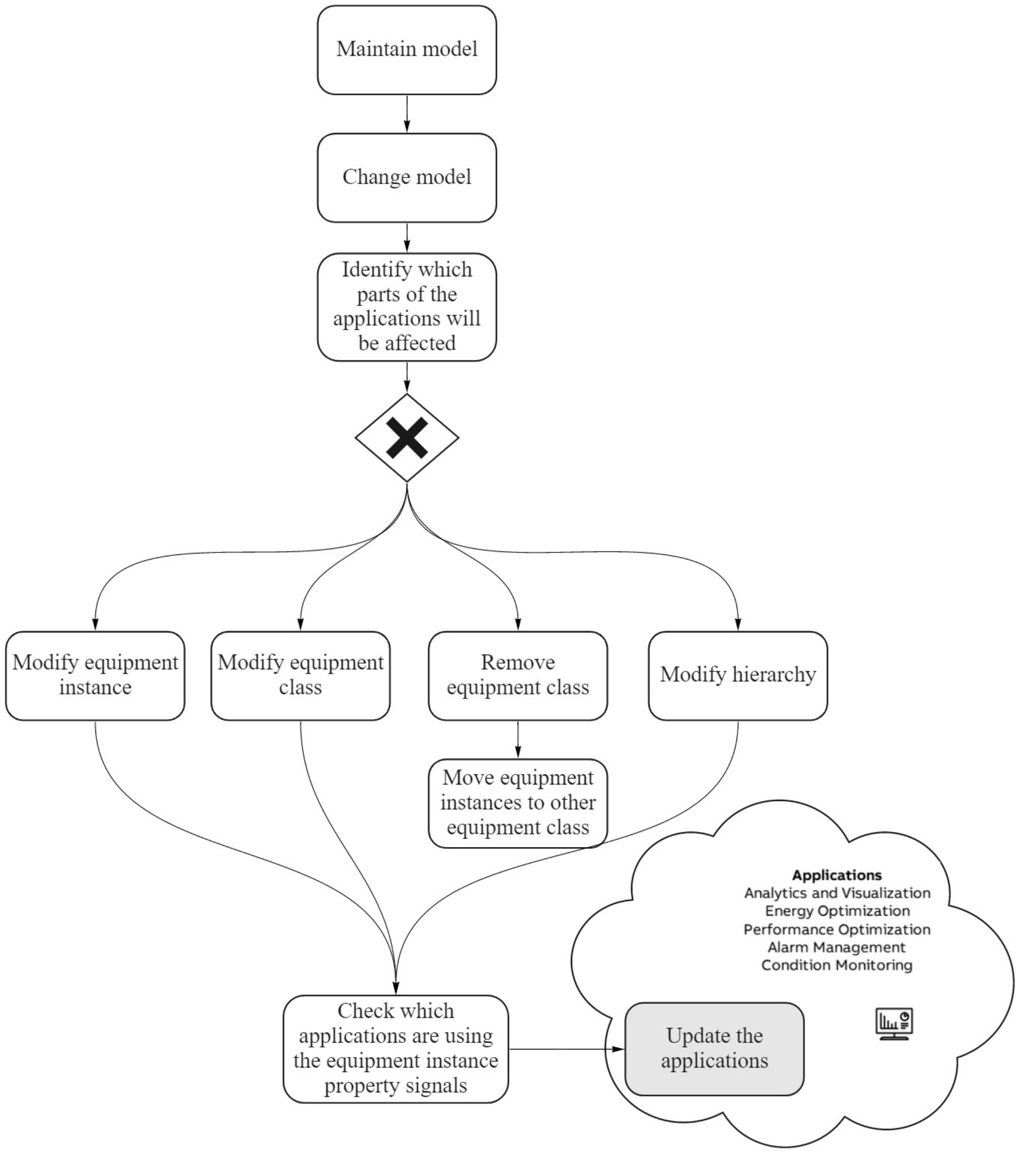


Figure 14 Current process, maintaining the data model

Equipment class can be added, modified or removed (Figure 15). Usually, in maintaining process when the equipment class should be added, it is in a situation that new equipment instance does not have any existing equipment class that it can be created from. When removing the equipment class, the equipment instances below the equipment class should be moved to another equipment class. When modifying equipment class, the possible activities are to add new property, to update property or to remove property. If needed to remove property, the if is affecting so much the equipment instances and, therefore, also the applications so it is be communicated with the data users. If the acceptance to remove the property is received the property can be removed totally from the data model. If it is the removing that is not accepted, the property is left in the data model and notified to the data users who are not using the data model tool.

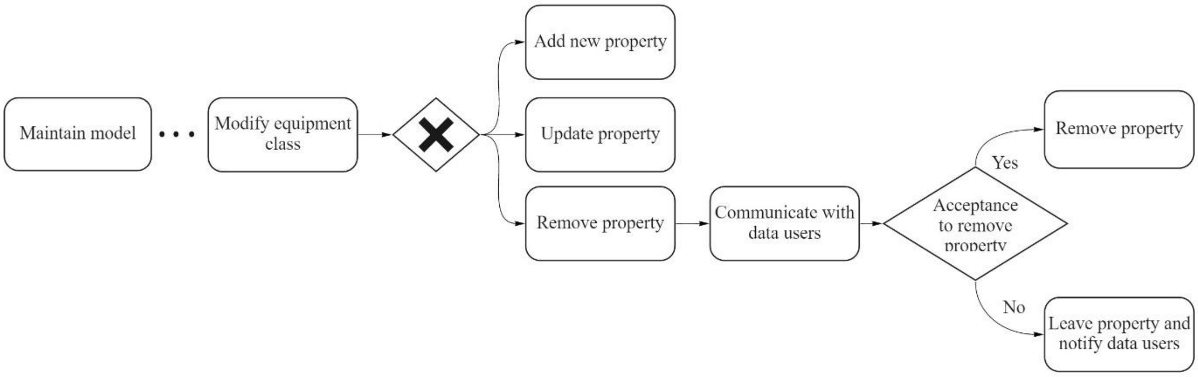


Figure 15 Current process, modifying the equipment class

When modifying the hierarchy (Figure 16) of the equipment model, the possible activities are to create equipment instance, move already existing instance to other place in hierarchy or to remove equipment instance. If removing equipment instance the history data can be moved/copied to new equipment instance or the history data can be left.

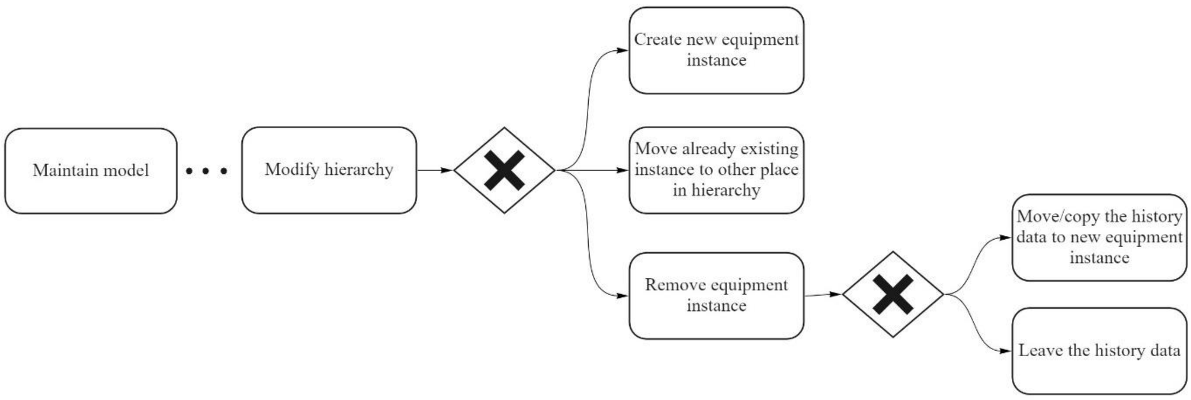


Figure 16 Current process, modifying the hierarchy

When creating new equipment instance (Figure 17), it should be considered if the equipment class has been instantiated, from which the equipment instance is created. If the equipment class exists, the equipment instance is created from the equipment class. If there are not any equipment classes from which the equipment instance can be created, the equipment class should be created first with its equipment properties. After that, the equipment instance can be created. When the equipment instance is in both situations created, the path is added and then the signals are connected to properties.

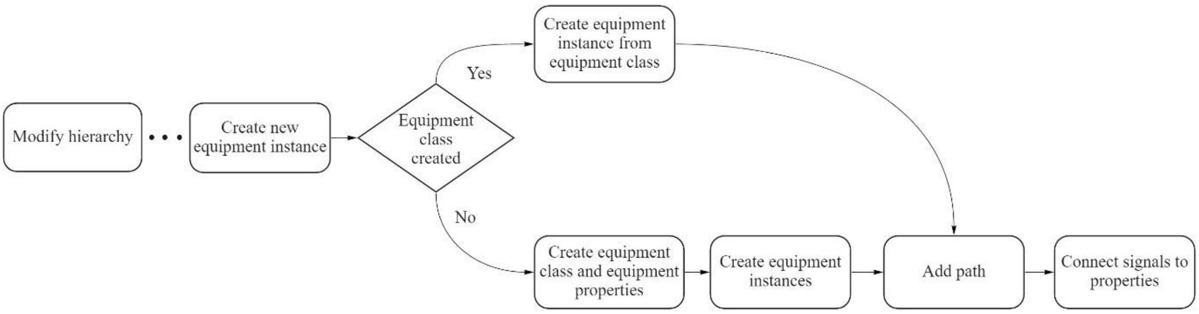


Figure 17 Current process, creating a new equipment instance

After all of maintaining steps, it is checked again which of application uses the modified equipment property signals. Next, the applications that are using the data model are updated according to the changes.

5 RESULTS

The interview results are discussed in this chapter. The purpose of the interviews was to find out development ideas and current situation regarding the data modeling in ABB Ability™ History and view of the effects of industry trends. The results are analyzed with comparing the answers to the theory and to other interviewees' answers.

Position and job title of interviewees varied which led to the fact that the answers given as knowledge, expertise and perspective was different with each interviewee (Table 3). Some of the interviewees worked more on the technical implementation and some from a larger perspective, with experience of many projects. It also gave more comprehensive look on the topics with diversity of answers. The interviews included of seven different topics: industrial digitalization, process data, data modeling, data modeling lifecycle, industry trends, implementing equipment model and ABB and other vendors services regarding the topics (Appendix 1). Next, the topics are discussed.

Table 3 Interviewed companies, roles and dates

Company	Interviewee role	Date
ABB	Project Manager	11.05.2021
ABB	Software Engineer	12.05.2021
ABB	ABB Lead Engineer	18.05.2021
ABB	Consultant	19.05.2021
ABB	Software Project Engineer	20.05.2021
ABB	ABB Ability™ History Product Owner	20.05.2021
ABB	Global Solution Architect - Metals Digital	25.05.2021
ABB	Project Engineer	26.05.2021
Company X	System Owner	15.06.2021
Outokumpu	Technology Architect – Data Management	15.06.2021

5.1 Process data

Characteristics of process data

The characteristics of process data that interviewees answered are listed in Figure 18. Six interviewees mentioned that characteristics of the process data are that it is time-based data and reflects the events of a specific time. Interviewees mentioned also that amount of the process data is also large, since it can be collected with millisecond resolution, or even more frequently. Often the time series can be compressed without sacrificing its usability. The process data has no correlation itself, since it is individual data of some observer meter. Process data is also structured data and, therefore, suitable for relational databases. The process data can be also batch-based when it is related to the product batch.

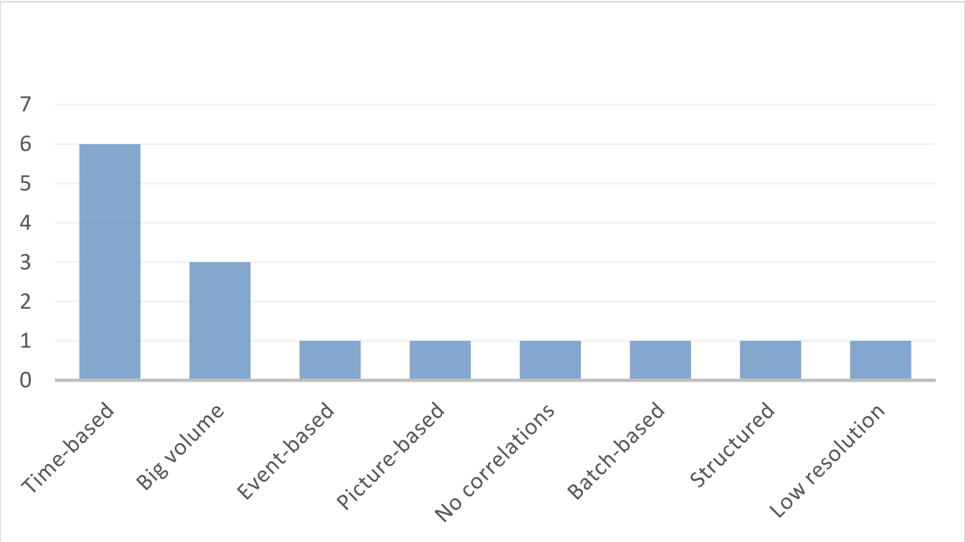


Figure 18 Process data characteristics

Usage of process data

The purpose of the process data is to bring knowledge of the production process to people who are interested of in. The process data describes the current production process data and the behavior of the process data describes the situation of the production process. The energy used to the production from another supplier can be also counted as process data, because it affects the production and the data collected from the production.

To achieve more efficient production, it is important to consider all the data that affects the production, not only the data that is collected form the plant. The process can be used to monitor the production process and in ensuring the quality. In the future it could be used in predicting the production process with the history data, comparing what is expected to happen to what is happening. This is already used in Outokumpu. The quality of the process data is important and in Outokumpu the aim is to provide data users information, information related to data quality, such as the latency of each signal. The latency means how long it has been since the signal has entered the system from the moment it was originally generated in automation.

Challenges regarding wider use

Six interviewees were able to recognize what could be the challenges regarding wider use of process data. Challenges could be, for example, human mind and lack of knowledge of how to utilize the data in developing the business. There is so much data to manage and new analysis methods are constantly being developed. Technically, there are not any challenges for a wider use. One interviewee mentions that one big challenge in wider use of process data is not understanding the difference between importing the data available and processing the data further. For data to be processed, it must first be available. “Available” means that signal data is obtained from automation systems imported into computers, where they can be processed by information technology. Processing projects are easily started on the assumption that the data is already available. The first challenge is usually forgotten, how to make the best use of the data so that it can be most easily processed in the future.

The use of the process data is transforming from reactive analysis to predictive analysis. Nowadays the process data is used very locally, for example only reacting to failures when it is happening and then correcting it, and maybe even analyzing what was the reason. The digitalization makes the factory use the data better and predict if and when something could go wrong. Four interviewees mentioned that in the future the machine learning could be also used when analyzing process data.

Organizing process data

To make data easier to structure or combine, devices should be able to structure and produce structural data on its own. The process data should be arranged so that the data in the database can be linked with the measured quantity of the process. The analysis is facilitated by the fact that the information is structured so that, for example, behind the control circuits, you can find all the necessary data for the analysis. Since the process data is time-based data, the data must be structured in time order. Due to the compression used to store the data, the stored discontinuous process data must be interpolated as accurately as possible between measurements to ensure that the data is continuous. The process data is also linked to product batches so that the problem detected in the process batch can be compared to potential problem areas in the production processes.

5.2 Data modeling

Variable model

The variable model collects data and includes the history of all variables. Each process must have a variable model independently. The variable model has been in used for a long time in one company and they are moving to use the equipment model to be able to move more data to cloud. One of the benefits of the variable model is that the data is collected and the history of the data can be found easily. If a calculation or report is made in the variable model, then you have to select the wanted variables individually related to a specific device. When you want to use the same report with other variables, you must select them again. If there is a situation that the calculation must be changed, the change must be made separately for the reports. Thus, manual work is multiplied when not all reports can be changed at the same time as they are separate. With an equipment model, applications are easier to maintain.

Equipment model

The advantage of the equipment model is that it connects the variables as properties under one equipment. When there will be more of this same equipment, the model can be repeated and just connect the process data to each of the properties. Another benefit of the equipment model is its flexibility. Many interviewees agreed that the possibilities of the equipment model are unlimited. The equipment model is a simple hierarchical model and it is easy to identify the required information. In the equipment model, all the variables and signals are in the structure, so they are easier to use. The equipment model can be used also, for instance, for process modeling, in supply chain, in cost management or in inventory. The equipment model could also be used to draw patterns of the behavior of equipment with the hierarchy, as it appears in the hierarchy if something goes wrong and associations can be detected.

Differences between variable and equipment model

Those interviewees who are familiar with variable model agreed that it is more simple and easier to implement than the equipment model. The variable model has a longer usage history and has been used for much longer. The equipment model is a newer data model and is still being developed a lot. Though one interviewee mentioned that equipment model is more structural, the whole structure is visible and it is possible to categorize the equipment, more modern way than the variable model. When the variable model is in use, the equipment model cannot be used at the same time for the same application but requires data replication and additional configuration. One interviewee adds that it would be good to be able to utilize both data models and not have to choose only one.

Now, the equipment model should be fully completed before it can be deployed. When an equipment model is used for existing facilities, the equipment model cannot be fully designed at the beginning. Data collection should start, for example, with a variable model on which the equipment model would be built. This also requires that the equipment model be adaptable. Customers should be able to use both of them flexibly, according to needs.

Hierarchy is essential in the equipment model, as are all related features and qualitative monitoring within the platform. The hierarchy of an equipment model could also be used only by referring to the hierarchy itself, utilizing the made structure. The goal is to use the equipment model as a reference data for applications. Conclusions of main differences between variable model and equipment model are listed in Figure 19.

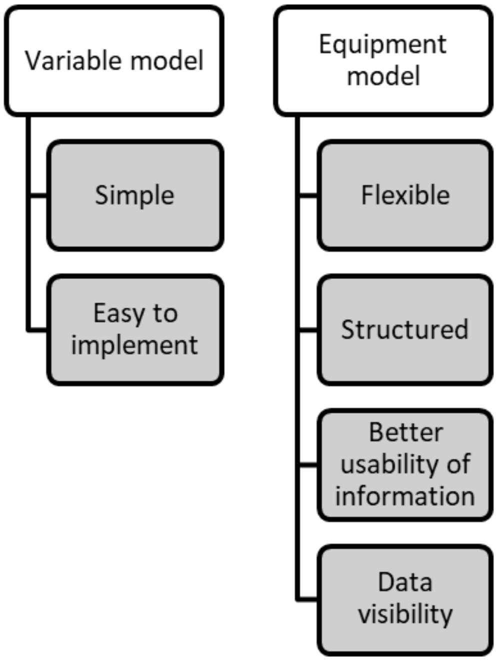


Figure 19 Conclusions of main differences between variable models and equipment model

Future usage possibilities

Machine learning will affect data models in the future. Now the amount of the data collected is high and the machine learning will help people in future to understand the data and process it. Only few interviewees were able to answer where the data models could be used in the future.

5.3 Data modeling lifecycle

Defining requirements

The requirements of the data model are based on what is wanted to be collected from the process. The data model is constructed to support the applications and the applications are setting some requirements for the data model. Many of the interviewees agreed that the requirements are changing before the data model is in production use and after it has been taken to production use. More challenging is to modify the data model after the data model has been taken into the production use.

The right competence used in structuring the data and creating the right equipment model is very critical to ensure that the full value is utilized of an equipment model. The importance of the competence was highlighted by eight interviewees. The production process should be understood by the persons building the model and the data users should understand the data modeling approaches. Before data modeling, the modelers should be aware of for example, the stages of the production process per equipment, what is measured with the properties that require signals and whether there are identical equipment on the production line. When defining a data model, one should think about and rationalize what is relevant data, as there is usually a little too much data available.

When one company was implementing the variable model, the requirements were considered very carefully and the necessary amount of time was spent on planning. Production workers were involved in defining the requirements. What was collected was initially planned and tested with one equipment. Since the model was put into production use, editing has been more challenging. It would be good if the changed name in the database would also appear in applications that use the data model, as naming changes may occur in the automation system or naming conventions will change with the new automation system. This, of course, is challenging to implement. The variable model adapts poorly to process changes because the names of the variables in the database cannot be changed.

Changing hierarchy

The hierarchy should describe where is what in the process, so naming equipment instances is very important. The naming conventional should be agreed with all stakeholders before building the data model. This helps the building and maintaining phase when for example finding a signal without knowing the path. The hierarchy can be changed. But according to some of the interviewees, the hierarchy is based on real-world objects, the equipment in the plant and the plant is not changing. The changes are also affecting the existing application, such as the dashboard and process models. The reference method that refers to the data in the model should remain constant. If it needs to be changed, the applications will also need to be changed. The current tool for data modeling is not having the information where the data is used and for what purposes.

According to one interviewee, the first thing is that data collection should be running to enable agile application development and the equipment model development comes along. Though, another interviewee thinks that first the equipment hierarchy and the equipment classes in the equipment model should be defined. The hierarchy should be the same as the plant is in the real world. The second thing is to define the equipment properties carefully. For example, if some information is missing, it needs to be added because it is more difficult to add when it is noticed in applications.

Process change or business change

A change in process data collection will cause changes in approximately three to five systems. Let us say a measuring circuit that has been in some calculations is removed from the automation. In this case, data collection from it must be stopped, calculations and history accumulations must be removed from the process data management system and in addition, it must be examined whether the data is transferred out of the process data collection system. After this, it is still necessary to check in which systems this information has been utilized, what has been used for it and what changes need to be made. For this, one interviewee is suggesting an option of a program on top of an RTDB database that would look for a variable in computational programs programmed in C # has been considered. Another way would also be for the end user to be able to retrieve a variable from the process database and get information on where it has been used. Obtaining data for variables passing through OPC UA is more difficult.

There are very few process changes in the process industry. When process changes occur, the hierarchy and equipment classes must be modified. The hierarchy is easy to modify but it directly affects applications that use the model so the connection must also be modified. In this case, it would be a good idea to keep a list of the applications that use the template and their links. There is no information in the equipment model about what data is used and for what purpose. If an equipment model is designed according to the equipment and data rather than a specific business need and all important and valuable data is collected, the equipment model should be able to adapt to changed business needs.

The data model should be ready for any changes and adapting to the changes easier. Like the configuration from the source systems should have an easier tool or mechanism. One company own model has a hierarchy and divides the variables by department. In the situation of new equipment and departments, each gets its own variables. When the equipment is turned off, its data collection is stopped, and old history data is preserved. The current hierarchy is so simple, so there have been very few changes.

According to one interviewee, the current tool for data modeling in ABB is not that well supporting the changes when the data model has been taken into production use. The changes have too big an impact on the applications, and it causes risks. Though, another interviewee mentions that the tool for the modelling is well-designed for changing the model before the production use. From the perspectives of reusability and lifecycle management, the equipment model still needs to be developed. Especially when different people model develop an existing model. Another challenge in general is that data collection automation systems are often customized solutions that always have something special to consider. Below is listed all that was mentioned what should be considered in data modeling lifecycle management (List 2).

What should be considered in data modeling lifecycle management:

- Requirements changes during data modeling process
- Naming conventional should be agreed in the beginning
- Live data connected before building data model
- Right competence in structuring data
- Production process should be understood in the beginning by the data modelers
- Flexible data modeling tool

List 2 What should be considered in data modeling lifecycle management

5.4 Industrial digitalization and industry trends

Almost all the interviewees agreed that the digitalization is slow in the industry. The industries have not considered all the benefits of digitalization. There are of course differences between various industries and there is still much to be done in the process industry for example. For many years the industries have been focused more on producing and revenue of the production. Nowadays the interest is growing towards developing the production to be more efficient.

Industries have growth potential; it is just a way to exploit that potential. To be able to find the possible ways to grow the data is needed. The collected data should be used to understand how the existing plants are running, to identify the inefficiencies in production capacity. In the long run the industrial digitalization will have an impact on sustainability because the industries are using the raw material efficiently and in an optimized way and the amount of industrial waste will be lower.

Industry 4.0

For eight interviewees the I4.0 is familiar as a concept but not in practice (Figure 20). The industrial internet of things is more familiar as a concept and more in practice. Generally, the value of the I4.0 is not that well understood. I4.0 has a greater impact on integrations. All the interviewees have not specifically considered the I4.0 in their operations or data modeling. Though, it should be considered, especially in data modeling. The reality is that there are few solutions, meters and automation systems that are directly integrated into the cloud.

Cloud computing

The Cloud is familiar for all the interviewees and has been in use in industries for the past few years. Although, it will take some time until most of the industries store the data in the cloud, since the industrial data processing is a slow-moving business. The customers of ABB are demanding nowadays that the digital solutions would be in a cloud. They all agreed that it affects implementing a new application or software, meaning the implementation is faster with a cloud. The capacity for data storage and computing capacity is unlimited in the cloud. Choosing the optimal cloud can be difficult. The challenge, however, arises from weighing on-premise and premise services, which is held on both sides. Cloud also provides scalability, flexibility and availability of applications.

The cloud is considered in data modeling. For example, the modeling has been done in such a way that it is utilized at the cloud level, meaning the data model is hierarchically built into three levels. This means that, at the lowest level, the data is collected from some factory or part of the process, at the second level, the data is gathered into different parts of the process and at the third level, it is utilized in the cloud through applications. The way the process data is integrated into the data model depends on the age of the existing equipment at site, interface capability and the digital readiness of the plant.

OPC UA

Not all OPC UA's capabilities have been exploited, for example the OPC UA data model would be capable of importing structured data. Thus, there is still room for improvement in making the data more structured and richer in content. OPC UA connectivity is used in data modeling in the data collection phase. The OPC UA data model works with the equipment model. OPC UA has been able to take advantage of the possibilities of the equipment model and bring information in the form of an equipment model. Only one interviewee recognized that OPC UA is part of I4.0 and considered in data model (Figure 20). Many other interviewee mentioned that OPC UA, is part of the data model but could not link it to I4.0 and therefore did not acknowledge that I4.0 is considered in data modeling.

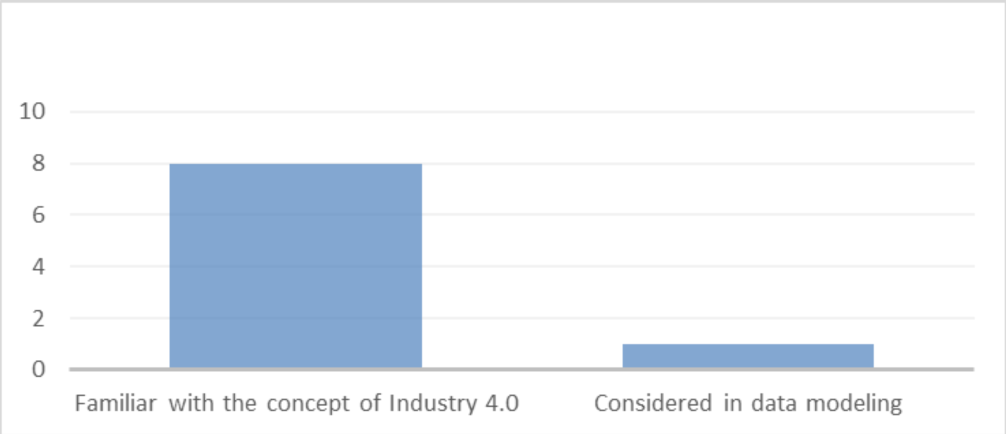


Figure 20 Familiarity of Industry 4.0 and considering it in data modeling

5.5 Implementing equipment model

Challenges in implementing equipment model

At Outokumpu the requirements were defined based on use cases, so the signals required by the equipment model were defined by the use case, as well as the characteristics and hierarchy. The top-level design was thus deficient, which has led to some inconsistency and duplication. New perspectives emerged in the construction process, which has created the need to rearrange the elements. This has caused confusion and led to maintenance challenges.

In the equipment model, it is possible to refer to a single signal by means of a hierarchy path or a unique ID. In the beginning of the implementation process at Outokumpu, the path of the hierarchy was used as a referencing method, which posed challenges because once the position of the signal has been moved, the path also changes, causing the reference to the signal to break down. The signal ID allows the end user of the data to reliably utilize the information produced by the signal.

When the equipment model was put into production, there has been a need for changes a few times. Changes must take into account the potential impact on the end user of the data. If it is known that the changes will affect the existing applications, the applications will be modified at the same time. Though, the biggest challenges have been related to data collection. Main challenges are listed in Figure 21.

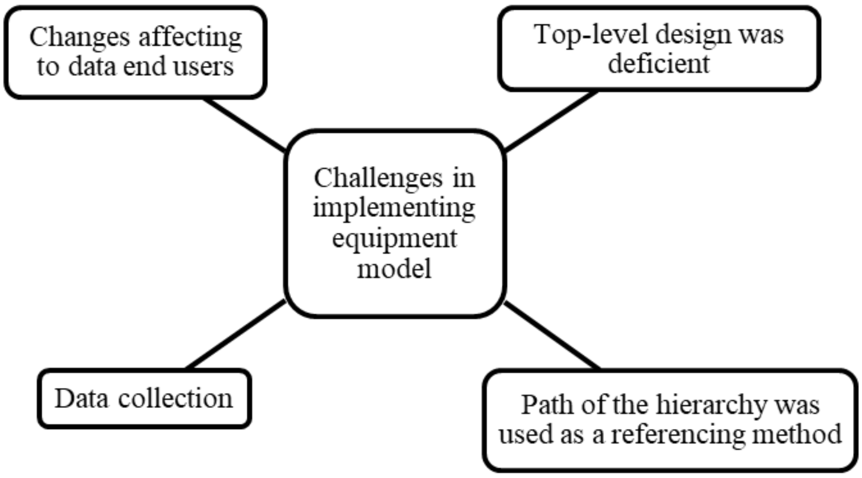


Figure 21 Main challenges in implementing equipment model

Overall picture and planning phase

It would have been better to take an approach by taking the factory to be modeled and starting to dismantle the equipment with an expert of the factory, meaning looking at what equipment is there and what potential features it has. Then classifying it into equipment classes and looked at how deep in the hierarchy you should go. So, initially make equipment classes and then instance hierarchy building, after which connected the automation data points to the model and make any fine adjustments if necessary. The overall picture of the equipment model should have been taken more into account when modeling the use case-specific part. For asset management and quality analysis applications it is important to have the overall material flow, the overall picture of the factory, visible from the data model.

One of the benefits of an equipment model is the creation an equipment class that can then be reused in different systems, so there is no need to rebuild it but only make it an instance. This could not be utilized so much, as the equipment classes should have been built differently, more knowledge of the site would have been required. Of course, more time could have been set aside, especially at the planning stage. From the equipment model, the data user can clearly see the available information. Generally, in implementing equipment model, the planning is important, for example, if a property is added to an equipment class later and history data collection begins on equipment, it lacks previous data. And a lot of time is frequently used in thinking what data is needed but not thinking that much of how to get the data and how to construct the data pipeline. The challenge is to get the data into the system and put it into the proper equipment/data model and get the information.

It would also be a good to find out from the equipment model which third-party systems, for example, read the equipment model. It would be good to have a feature in the equipment model that would convey the end of data collection so which variable no longer accumulates data, in some way to applications.

5.6 ABB and other suppliers' services

For ABB it is important to understand the value that ABB can provide. ABB brings the knowledge of the different processes, the understanding of the process and of how can the reusable equipment model be built and this is the value that ABB should highlight. ABB understands how the equipment model brings value to the customer in terms of reducing downtime, increasing productivity and increasing improving quality. ABB should still though develop the process knowledge. Nowadays is hard to buy the process knowledge, there are many IT platforms to choose. ABB can provide a continuous solution and not only consult one problem.

Support

ABB is expected to provide training support, maintenance-related expert support, equipment model construction and maintenance, as well as system maintenance and development. It would also be good for ABB to have a more consultative approach on how to adapt the equipment model to each situation and infrastructure. There is room for development in the end-user material of ABB Ability™ History, especially material related to UI browsing needs to be developed. ABB Ability™ History is not the easiest to approach, the interface should be developed as well.

ABB's and other suppliers' expertise level

ABB should be able to contribute better what and how customers need to do to achieve their objectives. ABB has the knowledge of the challenges of the plant and the processes. ABB has the expertise also, for example, in automation challenges and industrial protocols, sensors and the digital solutions. ABB should help the customer get a better idea of the scale the system, what they need to consider and the challenges in a timely manner. ABB could also provide service for integration for the solutions and the storage and analytics, to be able to provide the complete solution. ABB Process Automation has a good expertise level on integration and data orchestration, especially process data orchestration. With the knowledge of the operations as well as IT side, it enables to deliver better solutions.

Delivering total solutions is ABB's strength, of course together with a customer. ABB has the knowledge of what should be done and what solutions are needed for it. ABB also supplies a wide range of for example systems and products. There is a lot of software on the market, which possibly makes it difficult for customers to choose. So, there is a need for people who understand how software can be utilized to solve problems. ABB should assure that the expertise is on the same level in all regions on a global level. There is a growing need for people who understand technology and business. Consulting often thinks of a more complex solution than the client needs. In the end, however, the customer wants their production to run smoothly, efficiently, and as environmentally friendly as possible. Consulting needs and solutions are based on these basic problems.

ABB would have the opportunity to develop expertise and focus on automation systems. The information could contain more added value when used from automation systems. Automation systems could also have more interface solutions that are already IoT compliant. On the other hand, data aggregation services would also be useful. ABB supplies a lot of metering and automation systems and naturally then these data are combined before it is exported to any application for use. There is not much difference between the expertise of ABB and other suppliers. Compatibility should be developed for everyone so that parts of different vendor systems can be more easily discussed with each other. This is also influenced by the fact that it is difficult to buy real all-encompassing know-how.

6 EVOLVED DATA MODELING PROCESS

The evolved process is mainly formed based on Henderson et al. (2017) and Lee (1999) processes regarding data modeling and the interview results. Many suggestions regarding the planning and maintaining phase came up from the interview results. Lee (1999) defines the scope and the requirement analysis that are used in this process. Evolved data modeling process of implementing equipment model is presented in Appendix 2. The process is described below with its sub processes. The data model refers to equipment model.

The evolved data modeling process includes roles that are clarifying the responsibilities of each person in the process (Figure 22). The roles involved in the process are enterprise data stewards, business data stewards, data owner, technical data stewards and coordinating data stewards. The enterprise data steward is the one who have the overall picture of the data domain across business functions and is usually someone from the customer side, meaning data users side. The business data stewards are also usually from the customer side and they are defining and controlling the data with stakeholders. The enterprise and business data stewards can be combined to be one role depending of the project size. The data owner is also a business data stewards but has also the approval authority for decision about data within their domain. The business data stewards and data owner are providing and approving the access to the data and have then knowledge of the data integration of their domain. Technical data stewards, IT professionals, are the ones doing the modelling. There can be multiple technical data stewards people developing the data model and they have the knowledge of the equipment modeling tool and ABB Ability™ History platform. The coordinating data stewards are the main link between business and technical data stewards. The coordinating data stewards is aware of the situation of the process, the workload of different data stewards' teams and the overall picture of the data model. One person can have multiple roles. The roles are visible in the process flowchart.

Figure 22 Evolved process roles

The inputs of the evolved data process are existing data models and databases, signal list, data architecture, system architecture and data requirements. These affect the whole process, especially the identifying scope and requirements and planning steps. The process starts from business inputs, from customer or use cases for the data model. The main steps of the process are identified scope and requirements, connect live data, plan, build data model, review data model, develop and deploy visualization and analytics and maintain the data model.

The first step is to identify scope and requirements. Identifying the scope means specifying processes, information and constraints that fulfil the industry need. The requirements should be defined for the data model and the application scope. Some of the requirements are described during the process, this is why the enterprise and business data stewards are part of the process. Enterprise and business stewards have the best knowledge of where the data model will be used. The coordinating data stewards also participates in identifying scope and requirements to be able to forward the needed information to technical data stewards.

The next step is connecting live data. The data acquisition includes the collection of data from the devices and control systems and processing the data. This is a big difference comparing to the current process. The live data should be connected before planning phase to ensure that all data concerning the use cases or business requirements are available when the equipment model is defined in the planning phase. The business data stewards define what data is collected. The data owner gives the approval for connecting to the data source. The technical data steward has the knowledge of the interfaces and makes sure that the data abstraction interface is connected to the data source and to the storage layer of ABB Ability™ History.

Next is the planning phase (Figure 23) which consists of understanding production process, agreeing the naming convention, forming equipment model hierarchy, defining equipment classes and equipment properties and defining needed signals. Understanding production process helps technical and coordinating data stewards to have a bigger picture of the whole production process. Enterprise and business data stewards explain the production process to technical and coordinating data stewards. The stages of the production process per equipment, what is measured with the properties that require signals and whether there is identical equipment on the production line should be at least understood. Then the naming convention should be agreed. The goal is to have unified naming agreed in the beginning for multiple people easier to work at the same time. The naming convention should be agreed by enterprise, business, technology and coordinating data stewards to have everybody on the same line from the beginning. After that, equipment model hierarchy is formed by business, technology and coordinating data stewards. According to some interviews the hierarchy should be build based on real-world factory hierarchy, but it should support the business needs or use cases that the model is built. Next, the equipment classes and equipment properties are defined by the same roles as the earlier step. The equipment classes are created based on equipment that can be identical. The last step of the planning is defining needed signals which are done by data owner and technical data stewards. If, at this point some, of the signals are missing, they should be connected. Data owner is giving the approval for new signals. This is ensuring that all the possible and needed signals are connected before building the data model.

The conceptual and logical data model documents are the outcome of the planning. The main goal of the conceptual and logical data model is to have the data requirements documented and make the communication between roles easier. The conceptual data model collects high-level data requirements. The conceptual data model includes equipment classes and their properties without any other detail and relationships. Logical data model is a more detailed description of data requirements. The logical data model is mapped from the conceptual model, thus it contains equipment hierarchy, equipment classes, their properties, and data types of the properties.

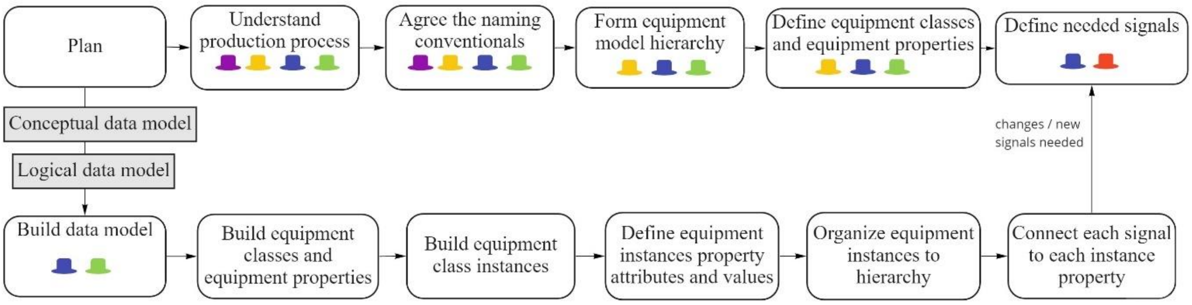


Figure 23 Evolved process, planning and building the data model

Then follows the actual building of equipment model (Figure 23). Technical and coordinating data stewards are mainly building the data model. Coordinating data stewards will contact the business data stewards if, for example, some information, assistance or approval is needed. The building starts with equipment classes and equipment properties, according to what was defined in the planning phase. Then the equipment class instances are built and their property attributed and values are defined. After this the equipment class instances (equipment instances) are organized to agreed hierarchy. After the equipment class, equipment properties, equipment class instances and hierarchy are ready, the signals are connected to each instance properties. If still some signals are missing, they need to be connected. The physical data model is an outcome of building the data model. The physical data model is the actual implementation solution. It includes equipment classes, equipment class instances and the properties and all the relationships and inheritances.

After the data model is built, the data model is reviewed by business, enterprise and cooperating data stewards (Figure 24). Once the data model is reviewed, it is approved by business, enterprise data stewards and data owner. Ensuring that the data model meets the scope and requirements from all roles. If the data model is not approved, it goes back to building phase. Next, the data model is used for visualization and analytics in applications.

Figure 24 Evolved process, reviewing the data model

The last step of the whole process of implementing the equipment model is the maintaining. The maintaining process is presented in Appendix 3. When the data model should be changed, the parts of applications that will be affected should be identified and documented using the API listing. The API listing is a list of the signals that are used by applications. From the list, it is possible to filter the applications that are using the specific signals that is going to be changed. The data owner, technical, business and cooperating data stewards should be part of identification and documentation. The documenting is done to change documentation that is available to all roles and people that are part of the implementing.

The possible steps in maintaining the data model are modifying equipment instance, modifying equipment class, removing equipment class and modify hierarchy. Some of the steps have own activities and then, after the change is done, it is noted in the change documentation. Then the API listing is again checked to have the information which application should be changed. Next, the application is updated by doing the changes in the application, according to the data model changes or automatically.

When modifying equipment class (Figure 25), the properties can be created, updated, removed. When updating or removing, the effects on the applications should be checked with API listing. That is done by data owner, technical, business and cooperating data stewards. Then, the acceptance is requested from business data stewards. Next, when updating the property, the property is updated or left as it is if the change is not accepted. After acceptance of removing the property, it will be removed and if it is not accepted the property is left as it is.

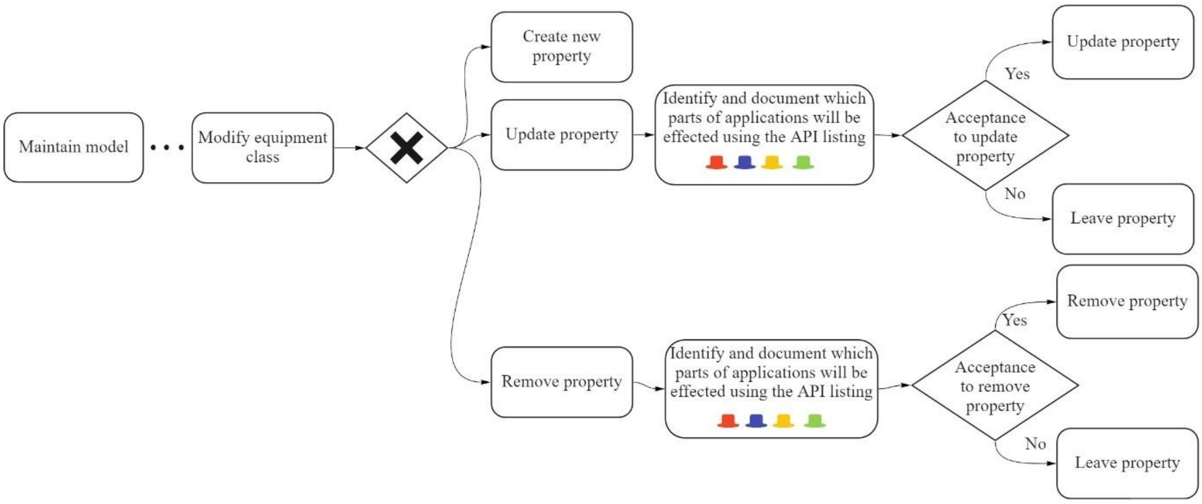


Figure 25 Evolved process, modifying the equipment class

If the hierarchy should be changed (Figure 26), new equipment instances can be created, already existing instance can be moved to other place in hierarchy or equipment instance can be removed. When removing equipment instance, the history data can be moved to new equipment instance if new equipment instance is created to replace the removed equipment instance. The history data can be also left with the equipment instance but there is no need to collect the data anymore. This happens when the history data might still be useful to use and the equipment itself is not in running anymore.

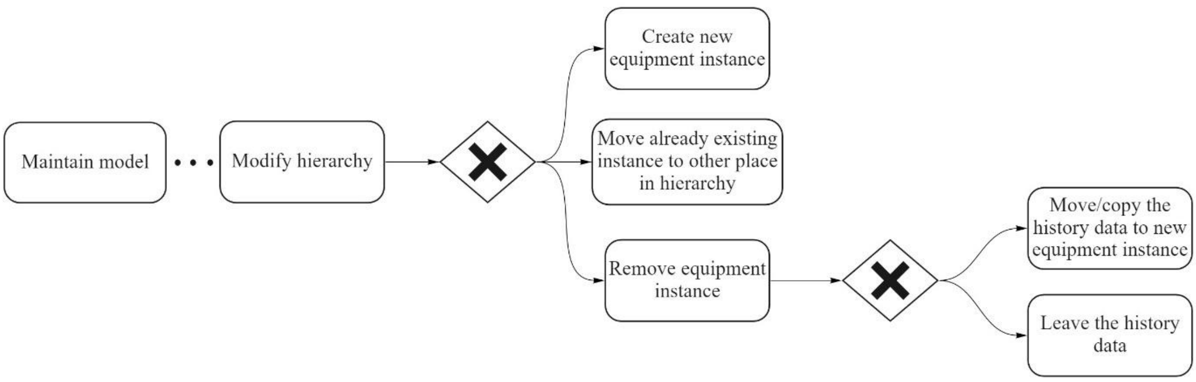


Figure 26 Evolved process, modifying the hierarchy

When the hierarchy is being modified and there is a need for creating new equipment instance, first it should be checked whether the equipment class is created, from which the equipment instance can be created (Figure 27). If the equipment class does not exist, it will be created with its equipment properties. Then the equipment instance is created. In both situation next path is added, and the signals are connected to properties.

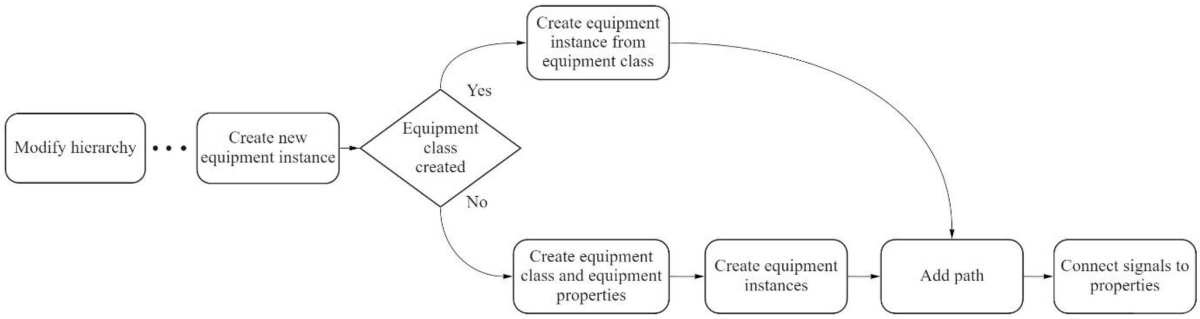


Figure 27 Evolved process, creating a new equipment instance

7 CONCLUSIONS

The main differences between the current process and the evolved process of implementing data model are defined roles, defining scope, connecting the live data before planning, more specifications done in planning, review and the change documentation integration and API listing in maintaining. Henderson et al (2017) listed seven data stewards and five of them were considered in the evolved process.

The interview results showed that there is a need for more specified planning and flexible maintaining. Defining the scope and the requirements analysis was also highlighted in the literature review. They prevent the amount of needed changes later in the implementation process. The importance of documenting the requirements and changes is high, but in reality, they take a lot of time and might be not that well done. Most of the interviewees felt that there should be more documentation and planning but the time is usually a limiting factor. Other issue that came up in the interview was that it is not possible to collect the data before the equipment model is ready.

The literature review showed that time series data, PIMS and data modeling are together considering many industry trends requirements regarding data. I4.0 with cloud, OPC UA and ISA-95 already affect the mentioned areas and will affect them from many perspectives. The cloud provides many functions when it comes to IoT, especially broad network access and rapid flexibility. ISA-95 is used mainly to define data flows and interfaces between company's business systems and manufacturing control systems. The hierarchy idea of equipment model is based on equipment hierarchy of ISA-95. With OPC UA the data communication is secured and it provides mechanisms for revealing certain semantics to the data.

7.1 Answers to research questions

1. *What are the best-practice processes and workflows to implement a PIMS data model?*

The goal of the evolved data modeling process is to ensure that the unpredictable impacts of changes in data model are smaller in applications. At the same time, evolved process increases the capability to make flexible and transparent data model. The review step gives the opportunity for various roles to suggest modifications before the equipment model goes to production use. As the literature review showed the modifications are easier and cheaper to do before the data model is in use.

With the defined roles, the responsibilities are evenly divided and the activities are clear to each role. The enterprise and business data stewards are in practice usually from the customers side, and the technical data stewards are from ABB side. The coordinating data stewards can be from either side. The roles are set in the process chart to clarify which roles should be part of each step.

The live data is connected before planning to ensure the data will be saved to the database and make it faster available. This could be done, for example, with implementing the variable model first and building the equipment model based on the variable model. After connecting the live data, the planning steps assure that the scope, requirements and plans are documented with conceptual and logical data model. Review and approving process ensures that the data model meets the scope and requirements from all the roles.

The purpose of having the change documentation integrated is that the changes are straight visible from the data model for all the roles. The change information that should be visible from the property are what and why the change was made, how objects changes, when the change was done, who made the change and where the change was made. The API listing should be checked before and after the changes to ensure that all the applications related to the changes are updated and the impact risks are known beforehand.

2. *What are the requirements for the software product supporting the proposed best-practice process to create a PIMS data model?*

The requirements for the software product are to be able to connect the live data before building the equipment model, API listing and the change documentation integration. The software product should also support the review step so that the reviewing would be effortless for needed roles. In addition, the flexibility was highlighted in literature review and interview results, thus the software product should be development to be more flexible. The I4.0 especially requires flexibility from all parts.

Now the live data can be connected after the data model has been built. This causes challenges because then there is an urgent need to build the data model to be able to collect the data. Therefore, the live data should be connected before data model development to have time for planning the model carefully. Connecting the live data also takes time and it is useful to have all the possible data available to be able to know if more data is needed.

The API listing is a list of signals and interfaces and applications the equipment model is connected to. The interfaces are divided to the signals that they use. The API listing helps to communicate between data modelers and data users if and when changes in the applications should be done. When the changes can be identified right after the change is done in the data model the data users will not have to wonder why the application does not work later.

The change documentation integration means that when changing, for example, the property in the equipment model, the change is noted to the property. This is to ensure that when someone else is modifying the equipment model, they know what was changed and why directly from the property. The equipment model is saving the changes automatically. Then there might not be a need to check a separate document that consist of all changes. The software should still be able to have a collection of all the changes and the change information to avoid any inaccurate changes and their risks to applications. This is helps for many people to work at the same time.

3. *What are the impacts of Industrial internet of things on data modeling process?*

The impacts are mostly concerning the usability of the data and visibility of the data (Figure 9). The volume of the manufacturing data is high; thus the data model should be capable to process the such amount of data. The live data is therefore connected before the planning the data model in the evolved process to ensure that all data is collected from the beginning and that the data model is able to process all the data.

The traceability of the changes in the plant should also be visible in the data model. Therefore, the data modeling process should have a traceability mechanism. The change documentation integration helps the traceability.

The real-time capability is important in I4.0, which requires that the time spent on data modeling is not long so that the data can be used with the data model as soon as possible. The latency of each signal should be also considered, meaning how long it has been since the signal has entered the system from the moment it was originally generated in automation. Also, during the whole data modeling process, the data model should be consistent, reliable and scalable.

7.2 Limitations and further research

This thesis included interviews of eight ABB employees and two ABB’s customers representatives. While the interviews results offered a lot of information, interviewing more people from different roles from ABB and more customers could give more comprehensive results. The opinion of the interviewees influenced quite much the results since there was not that many of them. This should be considered when interpreting the results.

The system topology affects the data model a lot and it should be planned well before starting to develop the data model. Therefore, the system topology and should be studied. The industry trends might also affect the system topology.

The roles were defined in the evolved data modeling process but the data governance model could be considered more. Thus, it should be considered what the approval chain is, when the changes are done and when there is a need for the approval chain. The data modeling process should be also considered from the inheritance aspect. That how to use the inheritance to optimize the process.

REFERENCES

ABB. 2020. ABB Annual Report 2020.

ABB. 2021. ABB Ability™ History. [internal online document]. [cited 4.5.2021].

Cavalieri, S & Salafia, M. G. 2020. Insights into Mapping Solutions Based on OPC UA Information Model Applied to the Industry 4.0 Asset Administration Shell. *Computers.* Vol 9, pp. 1-28.

Chen, D. 2005. Enterprise-control system integration – an international standard. *International Journal of Production Research.* Vol. 43, No. 20, pp. 4335-4357.

Du, N., Tang, L., Liu, X. & Chen, X. 2018. Research and Application of a Developing Approach for PIMS Software. Proceedings of the International Conference on Mechatronics and Intelligent Robotics (ICMIR2017). Recent Development in Mechatronics and Intelligent Robotics. Vol. 1, pp. 83-90.

Gilchrist, A. 2016. Industry 4.0: The Industrial Internet of Things. Bangken, Nonthaburi, Thailand. Apress Media. pp. 250.

Graube, M., Hensel, S., Iatrou, C. & Urbas, L. 2017. Information Models in OPC UA and their Advantages and Disadvantages. 22nd IEEE International Conference on Emerging Technologies and Factory Automation.

Hall, T. 2020. The Role of Data in Industry 4.0 [cited 25.5.2021]. [online article]. Available at: https://industrytoday.com/the-role-of-data-in-industry-4-0/

Halpin, T. 2001. Information Modeling and Relational Databases: From Conceptual Analysis to Logical Design. San Francisco. Morgan Kaufman Publishers. pp. 792.

Henderson, D., Earley, S. & Sebastian-Coleman, L. 2017. DAMA-DMBOK. Data Management Body of Knowledge. Basking Ridge, New Jersey, USA. Technics Publications. pp. 624.

ISA-95. 2021. [web page]. [cited 25.4.2021]. Available at: https://isa-95.com/isa-95-enterprise-control-systems/

Johnsson, C. 2004. ISA-95 – how and where can it be applied. Conference ISA Expo. Huston, TX, USA.

Khan, M., Wu, X., Xu, X. & Dou, W. 2017. Big Data Challenges and Opportunities in the Hype of Industry 4.0. IEEE ICC 2017 SAC Symposium Big Data Networking Track.

Lee, B., Kim, D.-K., Yang, H., Oh, S. 2016. Model transformation between OPC UA and UML. *Computer Standards & Interfaces.* Vol. 50, pp. 236-250.

Lee, Y. T. 1999. Information modeling: From design to implementation. Proceedings of the second world manufacturing congress. International Computer Science Conventions Canada/Switzerland. pp. 315-321.

Leighton, B., Cox, S. J., Car, N.J., Stenson, M.P., Vleeshouwer, J. & Hodge, J. 2015. A best of both worlds approach to complex, efficient, time series data delivery. *International Symposium on Environmental Software Systems.* pp. 371-379.

Llusà Serra, A., Vila-Marta, S. & Escobet Canal, T. 2016. Formalism for a multiresolution time series database model. *Information Systems*. Vol. 56, pp. 19-35.

Muza, P. 2005. Process Information management System at Mount Isa. *First Extractive Metallurgy Operators' Conference Proceedings 2005*. The Australasian Institute of Mining and Metallurgy (The AusIMM) pp. 179-187.

Namiot, D. 2015. Time Series Databases. Conference: XVII International Conference «Data Analytics and Management in Data Intensive Domains» (DAMDID/RCDL'2015), Obninsk, Russia, October 13 – 16.

Neubert, R. 2016. IIoT offers economic, operational benefits. *Plant Engineering.* pp. 9-10.
OPC Foundation. 2021. Unified Architecture: Interoperability for Industry 4.0 and the Internet of Things. [online document]. [cited 18.03.2021]. Available at: opcfoundation.org/wp-content/uploads/2016/05/OPC-UAInteroperability-For-Industry4-and-IoT-EN-v5.pdf.

Pauker, F., Fruhwirth, T., Kittl, B. & Kastner, W. 2016. A systematic approach to OPC UA information model design. 49th CIRP Conference on Manufacturing Systems (CIRP-CMS 2016). Procedia CIRP. Vol. 57, pp. 321-326.

Pfrommer, J., Grüner, S., Goldschmidt, T. & Schulz, D. 2016. A common core for information modeling in the Industrial Internet of Things. *Automatisierungstechnik.* Vol 64, no 9, pp. 729-741.

Pras, A. & Schönwälder, J. 2003. On the Difference between Information Models and Data Models. *The Internet Society.* Information Models and Data Models. pp 8.

Saldana, J. 2011. Fundamentals of qualitative research. New York, Oxford University Press. pp. 200.

Satyanarayana, S. 2012. Cloud Computing: SaaS. *Computer Science and Telecommunications.* Vol 36, No. 4, pp. 76-79.

Sendler, U. 2018. The Internet of Things. Berlin, Germany. Springer-Verlag GmbH Germany. pp. 278.

Simsion, G. & Witt, G. 2005. Data Modeling Essentials. Amsterdam, Boston. Morgan Kaufmann Publishers. pp. 561.

Stich, V., Schmitz, S. & Zeller, V. 2017. Relevant Capabilities for Information Management to Achieve Industry4.0 Maturity.18th Working Conference on Virtual Enterprises (PROVE), Vicenza, Italy. pp. 28-38.

Trunzer, E., Cala, A., Leitao, P., Gepp, M., Kinghorst, J., Luder, A., Schauerte, H. Reifferscheid, M. & Vogel-Heuser, B. 2019. System architecture for Industry 4.0 applications. *Production Engineering.* Vol 13, pp. 247-257.

Unver, H. O. 2012. An ISA-95-based manufacturing intelligence system in support of lean initiatives. *The International Journal of Advanced Manufacturing Technology.* Vol 65, pp. 853-866.

Vegetti, M. & Henning, G. 2014. ISA-88 Formalization. A Step Towards its Integration with the ISA-95 Standard. Proceeding of the 6th Workshop on Formal Ontologies meet Industry, Rio de Janeiro, Brazil.

Zhap, L., Sakr, S., Liu, A. & Bouguettaya, A. 2014. Cloud Data Management. Springer International Publishing Switzerland. pp. 202.

Appendix 1. Structure of the interviews

Interview questions

Industrial Digitalization

- What view you have on industrial digitalization?
- How do you believe or see how the industrial digitalization affects the business?

Process data

- What do you think are the characteristics of process data?
- Where is the process data used?
  - How about in the future, where could it be used for? What are the barriers to wider use?
- What requirements does process data impose on data parsing? How should process data be classified and/or organized, to make it more available?

Data modeling

- Is the ABB Ability™ History variable model (Variables) or Equipment model familiar? Which one have you used and for what purpose?
- What are the benefits of the variable model? What about the Equipment model?
  - Where else could you use the data models? What about in the future?

Data modeling life cycle

- How are data model requirements defined before building?
  - How much do the requirements change during the building process?
- How often has there been need to modify the data model since it is put into production use? How does it differ from modifying the data model before it is put into production use?
  - Which of these situations is more challenging?
- How well does the data model adapt to process changes and changed business needs?
- Have there been any challenges as the hierarchy of the equipment model has changed?
- How the above changes affect existing applications and the availability of data in them?

- What other potential challenges have been encountered in data modeling? What about the potential benefits of data modeling? What benefit is expected?

Industry trends

- Is Industry 4.0 or the industrial internet of things already familiar?
    - Have Industry 4.0 practices been generally considered in your operations? What about in data modeling?
- Will cloud have an impact in implementing new software?
- How the standard OPC UA is considered in implementing process information management system and data models?

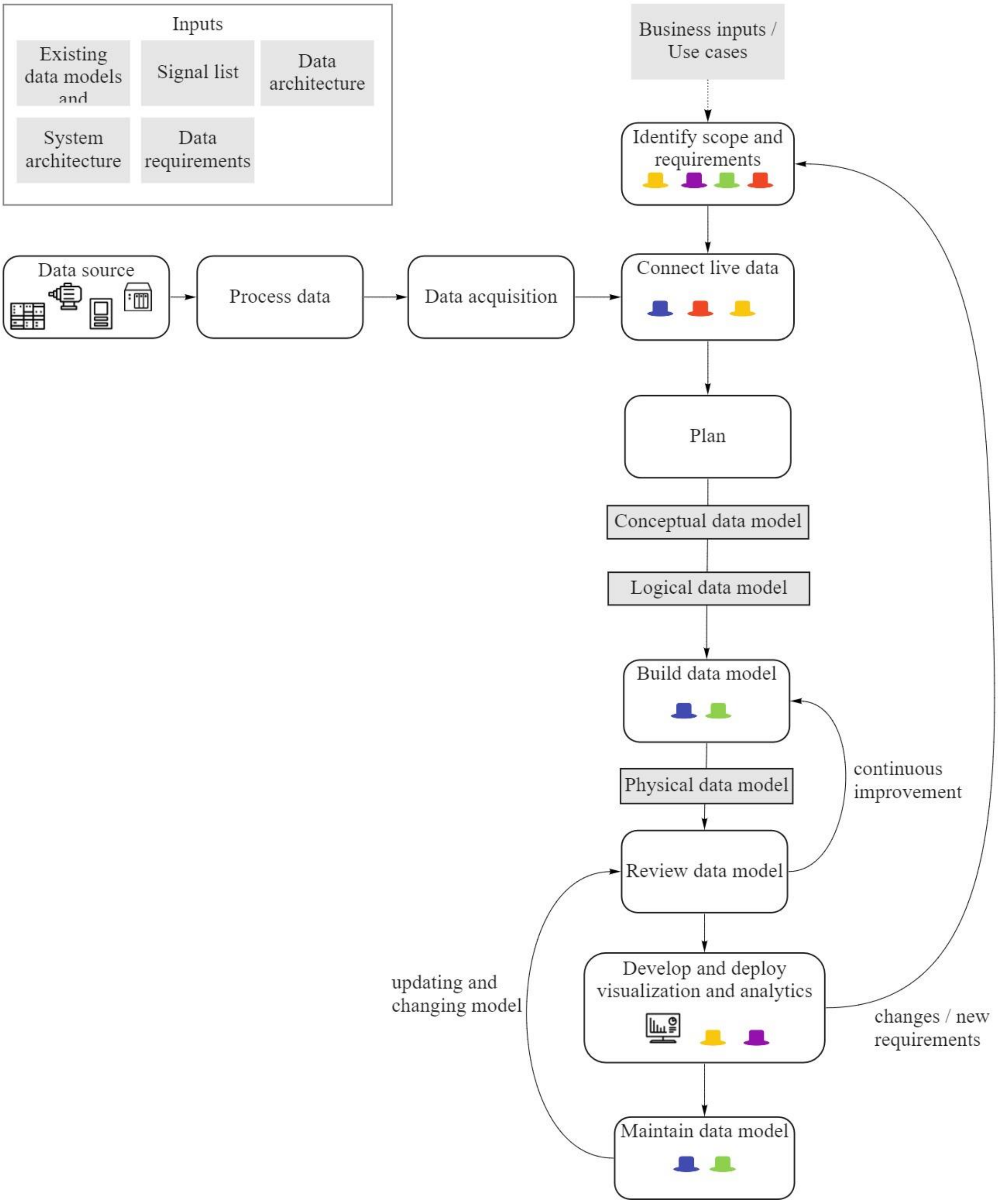
Implementation of equipment model

- What has caused challenges in implementing equipment model?
- How were the use cases built? How were they considered in the equipment model?
    - How the overall picture of the factory was considered when modeling the use cases?
- How was the hierarchy of equipment model constructed? How were the equipment classes formed?
- How much time was planned to implement of the equipment model?

Services of ABB and other suppliers

- What kind of services ABB and other vendors should be able to deliver regarding the discussed subject?
- What is the level of expertise of vendors? What should be developed?
    - What about the level of expertise of ABB? What should ABB develop?
- What is hard to consult and what kind of consultation would be needed?
- What king of service is hard to buy?

Appendix 2. Evolved data modeling process of implementing equipment model



Appendix 3. Evolved data modeling process, maintaining the data model