TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

TERO PALONEN

DISTRIBUTED DATA MANAGEMENT OF AUTOMATION SYSTEM

Master of Science Thesis

Examiner: Professor Jarmo Harju
Examiner and topic approved by the Council of the Faculty of Computing and Electrical Engineering on 9 April 2014

ABSTRACT

TAMPERE UNIVERSITY OF TECHNOLOGY
Master’s Degree Programme in Information Technology
PALONEN, TERO: Distributed data management of automation system
Master of Science Thesis, 49 pages
June 2014
Major: Communication networks and protocols
Examiner: Professor Jarmo Harju
Keywords: automation system, automation application, data management

Automation system produces enormous amounts of data. Data is collected from the production process, automation equipment and several other functions related to production. Data is used in reporting, product quality control, maintenance of the automation equipment and improving the production process.

Several important issues have to be taken into consideration when planning and implementing data management. Properties of data quality are for example data accuracy and availability. Different kinds of automation applications must be able to access the data and applications also must be able to trust the data. In case of faulty data, reports produced by the application are incorrect. The consequences can be expensive; in the worst case, the whole production line must be shut down when trying to find the source of the incorrect information.

In this thesis, the data management of various automation applications was studied. The applications included in the study are used in automation systems in different industries. The objectives for the application study were finding out what kind of data the applications need, what kind of data the application produce, how the collected data is stored and how the stored data is used in further data processing.

One objective for this thesis was to find out the problems in the current data management of the automation applications. One problem is the scalability of the automation application in different sized system deliveries. Many of the automation applications require commercial software products, such as operating system and databases. The license fees of these products have a big part in the total cost of a small-sized system delivery. Solution for this problem could be the utilization of Open Source software products. Many Open Source based software products would be feasible, however, there are difficult issues related to the utilization of Open Source products. The license terms of Open Source products and the lack of necessary support are often showstoppers for choosing them.

Thesis also investigates effects of new technologies, such as cloud based services and big data, on the data management of automation system. Cloud based services enable the development of new business models. The selection of the cloud provider must be done carefully in order to avoid becoming too dependent on one cloud technology. Changing the technology used in implementing cloud based services could become difficult and very expensive.

TIIVISTELMÄ

TAMPEREEN TEKNILLINEN YLIOPISTO
Tietotekniikan koulutusohjelma
PALONEN, TERO: Automaatiojärjestelmän hajautettu tiedonhallinta
Diplomityö, 49 sivua
Kesäkuu 2014
Pääaine: Tietoliikenneverkot ja protokollat
Tarkastaja: professori Jarmo Harju
Avainsanat: automaatiojärjestelmä, automaatiosovellus, tiedonhallinta

Automaatiojärjestelmä tuottaa valtavan määrän tietoa. Tietoa kerätään itse tuotantoprosessista, automaatiolaitteistosta ja lukuisista muista tuotantoon liittyvistä toiminnoista. Tietoa käytetään hyväksi muun muassa raportoinnissa, tuotteiden laadunvarmistuksessa, automaatiolaitteiston ylläpitoon liittyvissä toiminnoissa ja tuotantoprosessin kehittämisessä.

Tiedonhallintaa suunniteltaessa ja toteutettaessa on huomioitava useita tärkeitä asioita. Tiedon laadullisia ominaisuuksia ovat esimerkiksi tiedon oikeellisuus ja saatavuus. Erilaisten automaatiosovellusten on päästävä käsiksi tarvittaviin tietoihin ja voitava luottaa siihen, että käytettävä tieto on luotettavaa. Mikäli tieto on virheellistä, sovellusten tuottamat raportit ja mahdolliset muut yrityksen toimintaan vaikuttavat asiat ovat myös virheellisiä. Tämä voi aiheuttaa suuria kustannuksia ja pahimmillaan jopa tuotantoprosessien alasajoa, kun virheellisen tiedon lähdettä etsitään.

Tässä työssä selvitettiin erilaisten automaatiosovellusten tiedonhallintaa. Selvityksessä mukana olevat automaatiosovellukset ovat käytössä eri teollisuuden alojen automaatiojärjestelmissä. Selvityksessä tutkittiin, minkälaista tietoa sovelluksissa käytetään, minkälaista tietoa sovellukset tuottavat, kuinka kerätty tietoa tallennetaan ja miten kerättyä tietoa käytetään.

Työn tavoitteena oli myös selvittää, mitä ongelmia nykyisten automaatiosovellusten tiedonhallinnassa on. Eräs selvityksessä esiin tulleista ongelmista on automaatiosovellusten skaalautuvuus erikokoisiin automaatiojärjestelmätoimituksiin. Automaatiosovellusten vaatimien kaupallisten ohjelmistojen, kuten käyttöjärjestelmien ja tietokantojen, lisenssikustannukset muodostavat pienikokoisessa järjestelmätoimituksessa suuren osan kokonaiskustannuksista. Ratkaisu automaatiosovellusten käyttämien ohjelmistojen kustannusten pienentämiseen voisi olla avoimen lähdekoodin ohjelmistojen käyttöönotto. Vaikka monet avoimeen lähdekoodiin perustuvat ratkaisut ovat käyttökelpoisia, niiden käyttöönottoon liittyy kuitenkin ongelmia. Avoimen lähdekoodin tuotteiden lisenssiehdot ja puutteellinen käyttötuki ovat useasti esteenä niiden valinnalle.

Työssä selvitettiin myös mahdollisten tulevien teknologioiden asettamia vaatimuksia automaatiojärjestelmän tiedonhallinnalle. Pilvipalvelut mahdollistavat uusien liiketoimintamallien kehittämisen. Pilvipalveluiden toimittajan valinnassa on kuitenkin oltava tarkkana, ettei tule liian riippuvaiseksi yhdestä pilvipalveluiden toteutusteknologiasta. Pilvipalveluissa käytetyn teknologian vaihtaminen myöhemmin voi olla hankalaa ja maksaa paljon.

PREFACE

This work was carried out at Metso Automation, Tampere from autumn 2013 to spring 2014.

Assistance and support from many people have made this thesis possible. I would like to thank Matti Sarén for the opportunity to write this thesis at Metso Automation. I would also like to thank my supervisor Jyrki Yli-Nokari for the guidance during this project and other colleagues for the support and collaboration.

I want to thank Professor Jarmo Harju for the guidance and valuable comments during this work.

This thesis would not have been possible without the support from my family and friends during my studies.

Tero Palonen

Vesilahti, 18.5.2014

CONTENTS

TERMS AND DEFINITIONS

ACN	Application and Control Node
COM	Component Object Model
DCOM	Distributed Component Object Model
DCS	Distributed Control System
DMZ	Demilitarized Zone
ERP	Enterprise Resource Planning
HMI	Human-Machine Interface
IaaS	Infrastructure as a Service
IEC	International Electrotechnical Commission
IACS	Industrial Automation and Control Systems
IoT	Internet of Things
ISA	International Society of Automation
ISA-95	Standard for enterprise-control system integration
ISA-99	Standard for secure Industrial Automation and Control Systems
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
MES	Manufacturing Execution System
MRP	Manufacturing Resource Planning
NoSQL	Non-relational database, Not Only SQL
OLE	Object Linking and Embedding
OPC	OLE (Object Linking and Embedding) for Process Control
OPC A&E	OPC Alarms and Events
OPC DA	OPC Data Access
OPC HDA	OPC Historical Data Access
OPC UA	OPC Unified Architecture
PaaS	Platform as a Service
PLC	Programmable Logic Controller
REST	Representational State Transfer
SaaS	Software as a Service
SCADA	Supervisory Control and Data Acquisition
SCM	Supply Chain Management
UML	Unified Modeling Language

1. INTRODUCTION

A modern process plant or manufacturing company contains many information systems. The most common of them are for example an enterprise resource planning system (ERP) for business-oriented company administration, a manufacturing execution system (MES) for monitoring and controlling the manufacturing processes and automated process control systems for controlling the process equipment (DCS, SCADA). (Virta, 2010.)

Today it is very common that many different subcontractors take part in the functions of the company and they often also bring their own information systems that are implemented by different system providers. Therefore there can be dozens of different information systems that need to be integrated or at least interact with one another. Due to this, it is very important that all involved information systems have a common understanding about terminology, interfaces, protocols and data structures.

1.1 Motivation

Managing and analyzing data have always offered the greatest benefits and, on the other hand, the greatest challenges for companies regardless of the size of the company or the industry the company is operating on. Businesses have long struggled with finding the most suitable way to capture information about their customers, products, and services. In recent years many significant technology advances have been made in hardware, storage, networking, and computing models such as virtualization and cloud computing. The convergence of emerging technologies and reduction in costs for everything from storage to computers transformed the way data can be used and made the development of new business models possible. (Hurwitz et al., 2013.)

In the near future, the strongest growth in automation system solutions will be made with services and software. This is due to the rapid development in the areas of digital data transmission, new technologies such as simulation and modeling, improvements in decision making support systems and the integration of different information systems. Continuous technical development and changes in business models generate the need for the next generation automation systems. Environmentally friendly operations and energy saving functions will also have a strong role when future systems are designed. The increasing data exchange and communication between various devices and systems will significantly increase the need to utilize the information produced by the automation in both production and business level. (Tekes, 2010.)

Competition in the markets for automation equipment and process control systems is intense. There are many major vendors such as ABB, Honeywell, Metso and Siemens. Also new companies are getting involved in the automation market. In January 2014 Google acquired Nest Labs with a total purchase price of $3.2 billion. Nest Labs is a home automation company that designs and manufactures sensor-driven, Wi-Fi-enabled, self-learning, programmable thermostats and smoke detectors. (CNet.com).

Because of the intense competition on the process control system market, companies are searching for new business models in other areas of the automation domain. The true value for the customers and simultaneously better profit for the automation system vendor is now made in new services and utilization of new technologies. A good analogue of the value in data flow from low-level process equipment towards business-oriented decision making systems is a bucket brigade used in fire fighting; fire fighters passing water buckets to each other to extinguish a blaze. Even though every fire fighter in the bucket delivery chain is important and relevant for the outcome, the fire fighter in the end of the chain is the one putting out the fire and therefore the most valuable. Similarly, the highest value for the customer is generated in the end of the data delivery chain, in the business level.

The International Society of Automation (ISA) is a leading, global, nonprofit organization that develops standards, certifies industry professionals, provides education and training, publishes books and technical articles and hosts conferences and exhibitions for automation professionals. ISA standards help automation professionals streamline their processes and improve industry safety, efficiency, and profitability. (ISA.com.)

ISA-95 is the international standard for the integration of enterprise and control systems. ISA-95 consists of models and terminology. These can be used to determine which information has to be exchanged between systems for sales, finance and logistics and systems for production, maintenance and quality. ISA-95 standard defines a functional hierarchy model, which represents the hierarchy of the information systems of a manufacturing company divided into levels. (ISA-95.com.)

OLE for Process Control (OPC) is an interoperability standard for the secure and reliable exchange of data in the domain of industrial automation. One of the major reasons for the development of standardized automation software is the access to automation data in device level where number of different field bus systems, network protocols and interfaces are used. Three OPC specifications have been developed: OPC Data Access, OPC Alarms and Events and OPC Historical Data Access. With the successful adoption of OPC specifications in thousands of automation products, OPC is used today as standardized interface between automation systems in different levels of the automation domain. OPC Classic interfaces are based on COM and DCOM technology from Microsoft. The OPC Unified Architecture was born out of the desire to create a true replacement for all existing COM-based specifications. (Mahnke et al., 2009.)

Due to the publicity of the master thesis any business critical or otherwise company confidential information is not introduced in the thesis.

1.2 Scope of the thesis

The thesis work contains the study of data management in automation applications located on ISA-95 levels 1, 2 and 3. These applications are part of Metso DNA automation system and their functions include process control and manufacturing operations and control. The thesis does not include the data management issues related to industrial field buses.

1.3 Objectives

This thesis is a study for how data is managed in current automation systems and applications. The objective of the automation application study is to find out what kind of data there is in automation system, how data is collected from the automation system, how captured data is stored and how the data is used after capturing. Thesis also investigates effects of new technologies, such as cloud based services and big data, on the data management.

1.4 Structure of the thesis

Chapter 2 introduces automation systems in general. Both the reasons for usage of automation and problems related to automation are discussed. Distributed Control System and its components are introduced. Security related issues of automation system networks are also discussed.

Chapters 3 and 4 introduce standards related to information systems in automation domain. In Chapter 3 the ISA-95 standard and different parts of it are introduced. Chapter 4 describes the OPC and the OPC UA standards.

Chapter 5 describes the data management in general and data quality related topics. Different types of data and different heterogeneities of data are discussed.

Chapter 6 describes the automation system data management. Automation applications included in the study are introduced. Automation application data usage is described.

Chapter 7 discusses about the results of the thesis. Automation application data management problems are described. Cloud technology is introduced and discussed from data management point of view. Also the effects of big data and Internet of Things are discussed. In Chapter 8 the conclusions of the thesis work are introduced.

2. AUTOMATION SYSTEMS

Automation technology can be found in every branch of industry and also in everyday life in today's society. Embedded automation is such a natural part of life and daily functions that it is not even truly noticed anymore. For example a modern car has many computer units taking care of the driving functions and the safety of the driving. Automation systems are considered to be technology that is used in controlling production systems in process industry and manufacturing industry. In these applications, automation covers firstly the required hardware and software but also the provided services that are needed in all phases of the automation system's life cycle. (Tekes, 2010.)

Automation system may refer to a single programmable logic with a sensor and an actuator or a comprehensive plant-wide control system including several process control stations and monitoring and reporting applications. Automation system is used in automating some parts of a production phase or the entire functionality of the production. Functions of the automation system include everything from the traditional distributed control system (DCS) to the manufacturing execution system (MES) and enterprise resource planning system (ERP).

Automation system collects measurement data from the process. Based on the collected data, automation system controls the process by calculating the necessary control adjustments for the process and sending control commands to different actuators. Process data is presented to the process operator in a monitoring station. Typically the process operator is able to modify the adjustment commands by using operation station when necessary. (ABB Oy.)

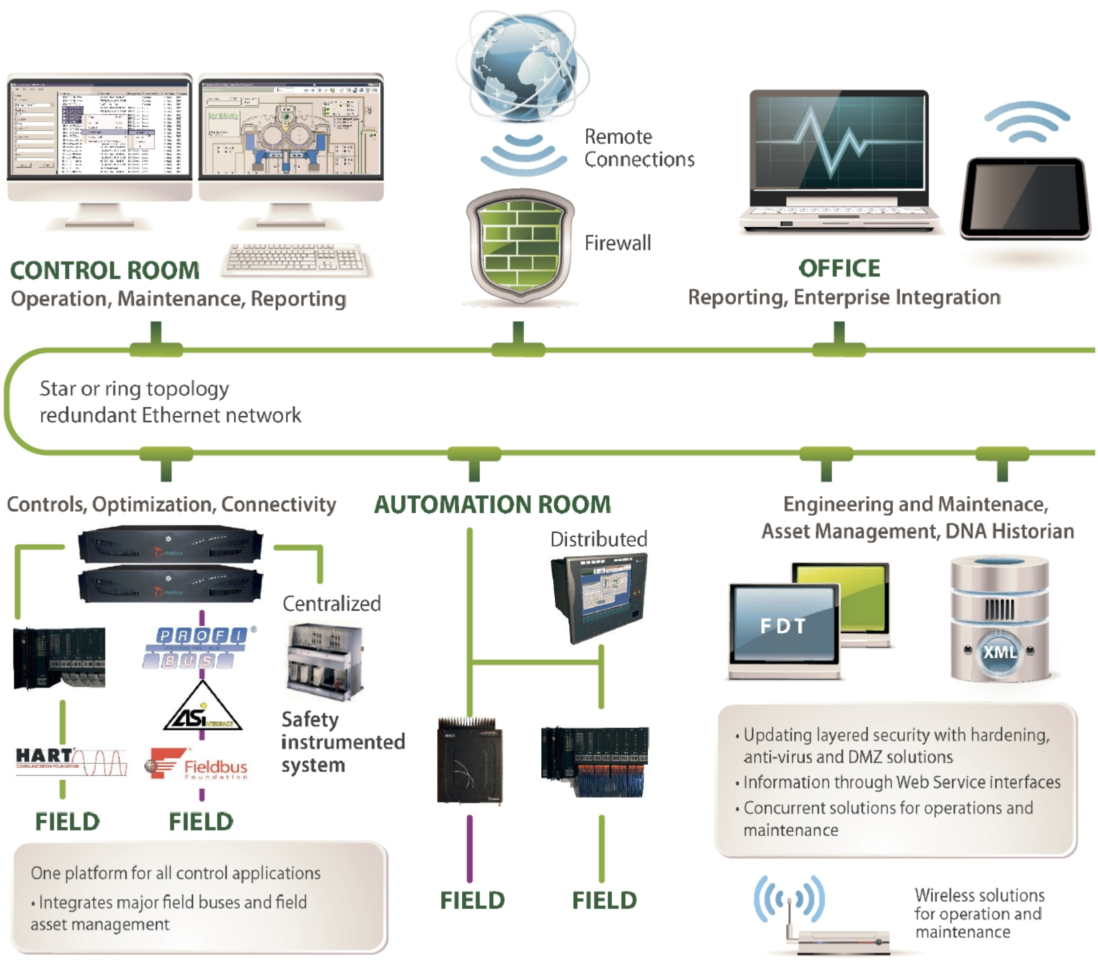

Figure 2.1. Metso DNA automation system architecture.

In the Figure 2.1 is an example of an automation system. Different functionalities of the automation system are interconnected by using automation system network. Data management functionalities are distributed in several different stations and computing units.

2.1 Reasons for automation

There are many different reasons for using automation. The following sub-sections introduce generally known and the most common reasons for automation usage. There are also mentioned some other reasons that can give an advantage compared to the competitors.

2.1.1 Better productivity

The basic idea in enhancing the productivity is to have the same or even higher level of productivity by using fewer resources. This can be achieved by using more automated machinery in the production process in order to decrease the amount of manual labor. Another way to enhance productivity is to minimize the consumption of raw materials or energy in the production process. Automation enables also more flexible working hours for production so that the expensive production equipment can be utilized also in the night.

2.1.2 Improved quality

Quality control is an important part of the production; the better the quality of the end products is, the higher price the customer is willing to pay for the products. Automation enables the homogeneous quality of the products. Typically production process has several phases and the homogeneous quality of the intermediate products enables more efficient latter production phases and therefore less loss altogether.

2.1.3 Faster production time and smaller stock size

Automation enables more flexible production. Different products can be manufactured in smaller series cost-effectively. Many fields of production have moved from inflexible mass production towards order-based production. It means that the manufacturing of the products starts only after the customer has ordered the products. This allows the minimization of the stock size and therefore less money is bound into the products waiting to be ordered and delivered. In many areas of the production industry the velocity of technical development is so fast that products become outdated very quickly. Customers are not willing to pay high prices for the outdated products and therefore the value of the stock decreases over time.

2.1.4 Information management

Automation makes it possible to have real-time monitoring of the production. Customers are able to track in which phase of the production the ordered products are. Better management of the process data also enables more effective supply chain management (SCM). A subcontractor is able to confirm the delivery time of certain intermediate products in order to keep the total production running smoothly.

2.1.5 Improved safety

Automation enables the minimization of risk and accidents in the production process. In case of an exceptional situation, automation system sets a warning and can also adjust the process back to safe state. If adjusting the process is not enough, automation system can set an alarm and shut down the whole process safely. Automation systems can monitor for example the waste water leakage, gas emissions and other important environmental issues of the production process. Automation of monotonous and hazardous work tasks releases employees to more productive and safe work assignments.

2.1.6 Traceability of the process and products

Automation enables the traceability of production process so that in case of a product callback all the delivered faulty products can be traced based on for example the batch number or the production date. This is crucial for example in the medical industry when a medicine batch is tested to be dangerous for the end users and therefore the whole batch needs to be retracted from retailers.

2.1.7 Diagnostics

Automated production systems are becoming more complex and sophisticated, which makes maintaining and repairing them more complicated. Automation systems have capabilities to identify the source of potential malfunctions and failures of the system. Status monitoring can provide information for diagnosing an existing failure and also provide data for predicting future malfunctions. Therefore the affected components can be replaced before actual failure happens and the production process can run smoothly without expensive standstills.

2.2 Problems related to automation

As automation increasingly takes its place in different areas of industry, it is often blamed for causing harm and increasing the chance of human error when failures occur, especially in high-risk industry. Usage of automation may isolate and alienate the operators monitoring the process from the process itself. Operators are sitting in control rooms watching the monitors while the automation system controls the complex process, in most cases trustworthy. At the same time the professional skills of personnel may deteriorate because of the more passive role in controlling the process. If something goes suddenly wrong, the operators are supposed to be able to control both the automation system and the process. In many cases the operators do not have enough experience for crisis situations and the operators are forced to make decisions and act quickly even without proper training.

Automation systems are implemented by using computers that execute the programs. Programs are designed to handle all the scenarios and every possible situation. Testing of the most difficult and rare scenarios and fail-safe procedures can be very demanding. Due to this, there can still be software bugs in the programs after the testing phase. If a crisis happens afterwards in the production phase, the consequences can be drastic. There are several standards for safety of the automation systems for requirements and guidance for automation system vendors.

2.3 Distributed control system

Most of modern process control systems are distributed control systems (DCS). A DCS consists typically of process stations, field buses, monitoring station, operation station and data management and reporting station. Modern process stations are capable of processing the measurement data, calculating the necessary control adjustments and controlling the process. This makes it possible to position the process station near the process itself. Due to the digitalization of automation devices it is possible for them to communicate with each other via computer network. Due to this, for example the data collecting application and process history database can be located in the server room and they still can capture data from all devices in automation system network.

In order to control the process, the computing unit of an automation system must be able to collect data from the process and transmit signals back to the process. The digital computer operates on digital binary data instead of analog data (continuous and analog signals) that is used in the production process. The components required to implement an interface between a process and a computer are: sensors for measuring continuous and discrete process variables, actuators that drive continuous and discrete process parameters, devices that convert continuous analog signals to digital data, devices that convert digital data into analog signals and input and output devices for discrete data.

A sensor is a transducer which is a device that converts a physical stimulus or variable of interest, such as temperature, force, pressure or displacement, into a more convenient form for measuring, usually an electrical quantity such as voltage. An analog measuring device produces a continuous analog signal such as voltage. The value of the signal varies in the same manner with the variable that is being measured. After the measurement, the output signal from an analog sensor must be converted to digital data. This is done by using an analog-to-digital converter. After the conversion, a digital computer is able to use digital values in its operations. If an output can have only certain values, then it is a discrete measuring device. Sensors can be either passive or active. An active sensor responds to the change in the quantity of the variable being measured without the need for external power (e.g. thermocouple responds to increase in temperature by generating a small voltage). A passive sensor requires an external source of power in order to operate (e.g. thermistor requires current to be passed through it). (Groover, 2008.)

In industrial control systems, an actuator is a hardware device that converts a signal from controller into a physical parameter. The change in the physical parameter is usually mechanical, such as a change in position or in velocity. An actuator is a transducer because it changes one type of physical quantity, such as electric current, into another type of physical quantity, such as rotational speed of an electric motor. Most actuators can be classified into one of the three categories: (1) electrical, (2) hydraulic or (3) pneumatic. Electric actuators are most common; they include different kind of electric motors, stepper motors and solenoids. Hydraulic actuators use hydraulic fluid to amplify the command signal from the controller. Hydraulic actuators are often used when large forces are required. Pneumatic actuators use compressed air as the driving power. Because of the relatively low air pressures involved, these actuators are usually limited to relatively low force applications. (Groover, 2008.)

Continuous analog signals from the process must be converted into digital values in order to be used by the computers. An analog-to-digital conversion (ADC) has three phases: (1) sampling, (2) quantization and (3) encoding. Sampling consists of converting the continuous signal into series of discrete analog signals at periodic intervals. In quantization each discrete analog signal is assigned to one of finite number of previously defined amplitude levels. In the encoding phase the discrete amplitude levels are converted into digital code, representing the amplitude level as a sequence of binary digits. In selecting the ADC for given application, the following factors are relevant: (1) sampling rate, (2) conversion time, (3) resolution and (4) conversion method. The sampling rate is the rate at which the continuous analog signals are sampled. The maximum possible sampling rate of an ADC is limited by the ADC conversion time. Conversion time is the time interval between the reception of the incoming signal and the determination of the digital value by the quantization and encoding phases of the conversion procedure. The resolution of ADC is the precision with which the analog signal is evaluated. Since the signal is represented in a binary form, precision is determinate by the number of quantization levels. Quantization generates an error because the quantized digital value is likely to be different from the true value of the analog signal.

The process performed by a digital-to-analog converter (DAC) is the reverse of the ADC process. The DAC transforms the digital output of the computer into a continuous signal to drive an analog actuator or another analog device in the process. DAC consists of two phases: (1) decoding and (2) data holding. In decoding the digital output of the computer is converted into a series of analog values at discrete moments in time. In data holding each successive value is changed into a continuous signal (usually electrical voltage) that is used to drive the analog actuator during the sampling interval. (Groover, 2008.)

Contact input interface reads binary data from the process into the computer. It consists of a series of simple contact information that can be either closed or open (on or off) to indicate the status of binary devices connected to the process such as limit switches (contact or no contact), valves (open or closed) and motor pushbuttons (on or off). The computer periodically scans the actual status of the contacts in order to update the values stored in its memory. Contact output interface sends binary signals from the computer to the process. The contact positions are set either on or off. These positions are maintained until changed by the computer. In computer process control applications, contact output interfaces control alarms, indicator lights, solenoids and constant speed motors. (Groover, 2008.)

A pulse counter is a device used to convert a series of pulses into a digital value. Pulse counters can be used for both counting and measurement applications. An example of a counting application is a device that adds up the number of packages moving past a photoelectric sensor along a conveyor. An example of measurement application is a device that indicates the rotational speed of a shaft. A pulse generator is a device that produces a series of electric pulses whose total number and frequency are specified by the control computer. The total number of pulses might be used to drive the axis of a positioning system. The frequency of the pulse train could be used to control the rotational speed of a stepper motor. (Groover, 2008.)

2.4 Automation system networks

Automation applications are extremely dependent on the reliability and the accessibility of the automation system. Response times of automation equipment and devices as well as the delay caused by the automation system network must be minimized. There are several requirements for automation system network:

- data rate and data latency
- physical interconnect medium
- noise immunity
- bit error rate and bus faults
- permissive interconnection length
- permissive number of network nodes
- ease of adding additional nodes
- power consumption, cost, reliability and isolation requirements.

Automation system network is usually a redundant ring topology or star topology switched Ethernet. Typically the automation system network is separated from the office network with a firewall. In Figure 2.2 is depicted one possibility to utilize a Demilitarized Zone (DMZ), a firewall configuration that uses two firewalls placed in series between the two networks. Any computer or other device that requires communication with both the office network and the automation system networks is placed between the two firewalls, within the DMZ. Each firewall can then be configured to allow the required level of interaction into the DMZ, but blocking any communication attempts from the office network directly to the automation system network and vice versa.

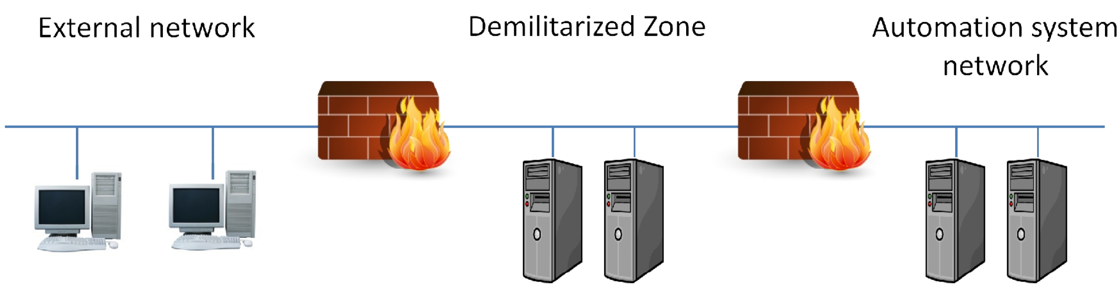

Figure 2.2. Example of demilitarized zone implementation.

In a survey carried out in the year 2013, researchers at Aalto University found a total of 2,915 automation devices in Finland that were accessible through the Internet and therefore vulnerable to cyber attacks. The automation devices targeted in the survey were associated with industrial automation systems, energy management systems, remote use of systems and building automation. For example, they control power stations, alarms and door locking. Unauthorized access to these systems could have disastrous consequences: hacking, even for experimental purposes, may cause serious damage to the system and the physical environment controlled by it, such as a factory or power plant. Remote user interfaces, in particular, have security gaps. Usage of default passwords and direct access to the login page for all Internet users in these data sensitive systems are signs of ignorance or carelessness. The search engine Shodan (http://www.shodanhq.com) that was used in the survey makes finding of vulnerable and attractive targets effortless. Of the devices encountered in the survey, 60% have publicly known vulnerabilities. (Aalto University, 2013.)

Cyber security standards are security standards that enable organizations to practice safe security techniques in order to minimize the number of successful cyber security attacks. These standards provide general outlines as well as specific techniques for implementing cyber security. Cyber security standards have been created because of the increasing amount of sensitive information that is stored on computers that are attached to the Internet.

ISA-62443 (also known as ISA-99) is an industrial control security standard created by the International Society of Automation (ISA) in order to protect SCADA and process control systems. ISA-99 offers variety of security recommendations based on the physical and logical location of the systems being protected, as well as their importance to the reliable operation of the system. In order to accomplish this, ISA-99 classifies functional areas of an industrial system into specific security levels and then provides recommendations for separating these areas into zones. ISA-99 also defines the interconnections between zones as well as how to enforce security between zones. (ISA.org.)

All ISA-62443 standards and technical reports are organized into four groups called “General”, “Policies and Procedures”, “System”, and “Component”. “General” category includes common or foundational information such as concepts, models and terminology. It also includes specifications that describe security metrics and security life cycles for Industrial Automation and Control Systems (IACS). “Policies and Procedures” category includes work products that handle issues related to establishing and operating an effective IACS security program. “System” category includes work products that give guidelines for system design and describe requirements for the secure integration of control systems. It describes also how to define security assurance levels using the zones. “Component” category includes work products that describe the specific product development and technical requirements of control system products.

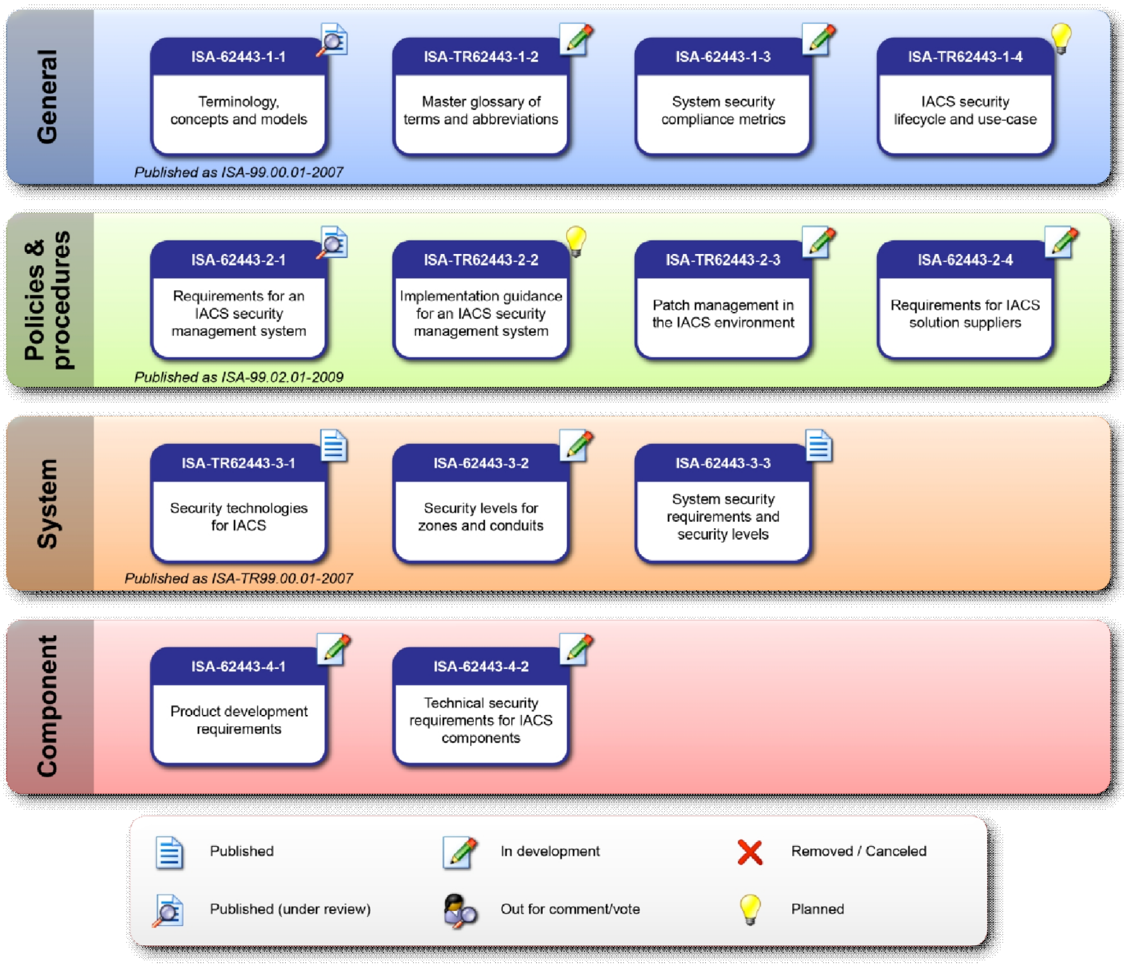

Figure 2.3. Specifications of the ISA-99 standard (ISA.org).

In the Figure 2.3 is presented different groups of the ISA-99 specifications and the status of the specifications (ISA.org). In addition to the work products of the ISA-99 standard, there are other IT security standards and practices that help organizations to integrate secure information models into software development of industrial information models.

3. ISA-95 STANDARD

The development of the information systems on the different hierarchy levels has not been concurrent. The earliest enterprise-level information systems were designed for automating the accounting and inventory management in the 1970s. Since then, many other enterprise management tools, like the resource management and the production planning, have been integrated to the enterprise-level information systems making the enterprise-level systems more dependent of the current production data. Modern ERP systems can be used to manage resources, i.e. human, material, and equipment resources, as well as sales and marketing information. However, the ERP systems themselves are business-oriented and meant to be financial management tools for the company administration, not for controlling and monitoring the manufacturing processes. This has formed a gap between the ERP and automated process control systems. Due to this gap a few problems have risen: without a close connection from the process control systems to the ERP, the timeliness and the reliability of the received data was uncertain, the amounts of data were huge and the data was difficult to analyze. (Virta, 2010.)

The continuing rapid advances in computer system capabilities and reliabilities make it possible for digital computer based systems to penetrate all aspects of process industry information and control system applications. However, such capabilities were not yet available when computer system applications first began in these industries. Due to this, automation industry is now faced with the task of interfacing and integrating the operation of the separately developed plant floor digital control systems with the very dissimilar systems used in the enterprise or business process operations of the company.

The ISA SP95 Committee has been assigned the task of preparing standards in order to enable the efficient integration of these two areas of computer applications. The objective of the ISA SP95 committee is to create a standard that defines the interfaces between different hierarchy levels and information systems. The goal is to reduce risk, cost and errors in the implementation phase of the interfaces. The standard has to define the information exchange that is robust, safe, secure and cost-effective. The information exchange mechanisms have to maintain the data integrity of each information system. (Williams, 1998.)

3.1 Functional hierarchy model

ISA-95 standard defines a functional hierarchy model, which represents the hierarchy of the information systems of a manufacturing company divided into functional levels. The functional hierarchy model is shown in the Figure 3.1. Each level provides specialized functions and has characteristics response times.

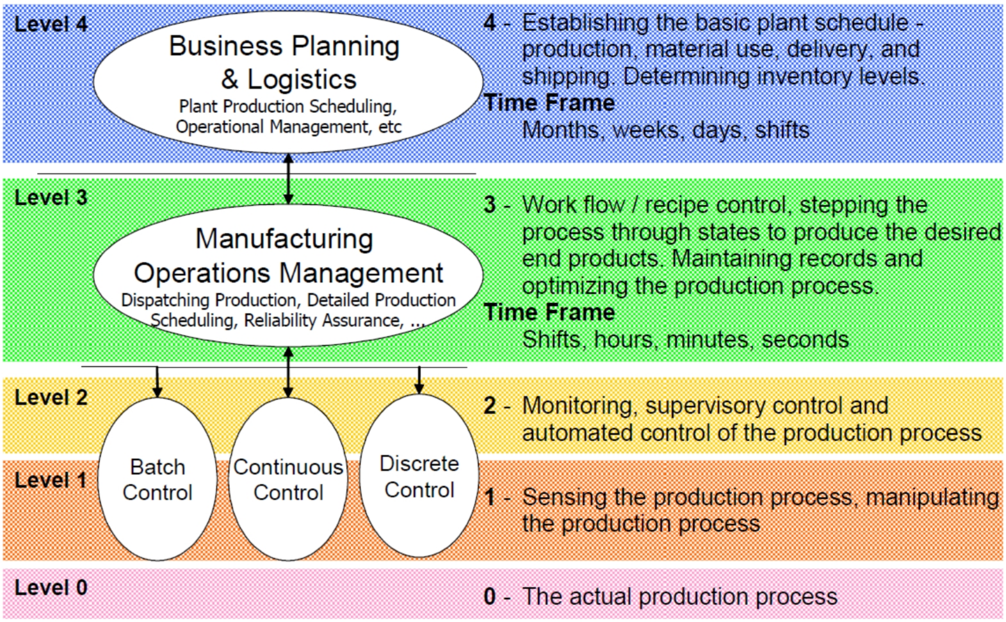

Figure 3.1. Hierarchy levels of the ISA-95 standard (Brandl, 2005).

The level 4, "Business planning and logistics" includes the enterprise-level functionalities which are often represented by an enterprise resource planning (ERP) system. An ERP system provides tools for managing the business-related activities of a manufacturing company. The most important function group of an ERP system is the finance management (FIM) which is used for accounting, sales and financial functions. Other important functions are the supply chain management (SCM), the human resources management (HRM) and the supplier relationship management (SRM). The SCM functions are used for planning the medium-term and long-term material and energy requirements and the logistics involved in the manufacturing process. The HRM functions are used for managing the personnel of the company and the SRM for managing the supplier relationships, i.e. customer information which is important for the sales, marketing and purchasing functions of the company. The time horizon in the enterprise domain processes is days, weeks, months or even years. (Virta, 2010.)

The level 3 is the production management layer, which contains the manufacturing execution functionalities such as work flow and recipe control. Optimization of the process is also a functionality of this level. The time frame of the level 3 processes range from days to seconds. (Virta, 2010.)

Levels 2 and 1 contain the process control functions. The level 2 contains functions for monitoring and controlling the process and typically operates on time frames of hours, minutes, seconds or even less. The level 1 contains functions carried out by sensors and actuators and typically has a time horizon of seconds or less. (Virta, 2010.)

The ISA-95 hierarchy model can be applied to the equipment, people, and information systems in a facility. The model defines levels that can be mapped to systems and to networks. However these systems and networks need to work together in a safe and secure manner.

Implications of the ISA-95 standard for Manufacturing Execution System (MES) supplier have been examined in Jussi Lautala's Master thesis. In the thesis the advantages of the ISA-95 standard are considered from the point of view of paper making business MES supplier. In his thesis, Lautala argues that if the ISA-95 standard would have been utilized more in the development phase of the MES system, certain application parts of the MES system could be replaced with other vendor's product with less effort. (Lautala, 2010.)

3.2 Different parts of the ISA-95 standard

There are five parts of the ISA-95 standard. Part 1 consists of standard terminology and object models that can be used to decide which information should be exchanged. Part 2 consists of attributes for every object that is defined in Part 1. The objects and attributes of Part 2 can be used for the exchange of information between different systems, but these objects and attributes can also be used as the basis for relational databases. Part 3 focuses on the functions and activities at level 3 (Production / MES layer). It is an excellent guideline for describing and comparing the production levels of different sites in a standardized way. Part 4 handles the object models and attributes of manufacturing operations management. Part 5 focuses on the transaction between business and manufacturing functions. (www.isa95.com.)

The first part of the standard, ISA-95.01: Models and terminology, provides standard models and terminology for describing the interfaces between the business systems of an enterprise and its manufacturing operations and control systems. This standard provides a consistent set of concepts and models for integrating control systems with enterprise systems that will improve communications between all parties involved. The planned benefits of the models are:

- reduce users times to reach full production levels for new products
- enable vendors to supply appropriate tools for implementing integration of control systems to enterprise systems
- enable users to better identify their needs

- reduce the costs of automating manufacturing processes
- optimize supply chains
- reduce life-cycle engineering efforts.

The ISA-95.1 part of the standard is intended for those who are involved in designing, building, or operating manufacturing facilities, responsible for specifying interfaces between manufacturing and process control systems and other systems of the business enterprise or involved in designing, creating, marketing, and integrating automation products used to interface manufacturing operations and business systems. It is not the intent of this standard to suggest that there is only one way of implementing integration of control systems to enterprise systems, to force users to abandon their current methods of handling integration or restrict development in the area of integration of control systems to enterprise systems.

The second part of the standard, ISA-95.02: Object model attributes, further defines formal object models for exchanging information described in part 1 using UML object models, tables of attributes and examples. The models and terminology defined in this part:

- emphasize good integration practices of control systems with enterprise systems during the entire life cycle of the systems
- can be used to improve existing integration capability of manufacturing control systems with enterprise systems
- can be applied regardless of the degree of automation.

The third part of the standard, ISA-95.03: Activity models of manufacturing operations management, defines production activities and information flows. Within production areas several activities are executed and a lot of information is exchanged. This part provides reference models for production activities, quality activities, maintenance activities and inventory activities. With these models e.g. the following questions can be answered:

- which department is responsible for which activities?
- which automation systems are responsible for which systems?
- what is the focus of various projects within the company and are there gaps or overlapping responsibilities?
- which functionality is represented in more than one information system?
- what are the boundaries between our automation systems?

Usage of the standard can give e.g. the following advantages:

- suppliers can develop suitable tools for production activities
- end Users can identify their needs more easily
- end Users can compare the solutions of different suppliers.

The fourth part of the standard, ISA-95.04: Object models and attributes for manufacturing operations management, defines object models that determine which information is exchanged between MES activities. The models and attributes from part 4 are the basis for the design and the implementation of the interface standards and make sure of a flexible cooperation and information-exchange between the different MES activities.

The fifth part of the standard, ISA-95:05: Business to manufacturing transactions, defines transactions in terms of information exchanges between applications performing business and manufacturing activities associated with ISA hierarchy levels 3 and 4. The exchanges are intended to enable information collection, retrieval, transfer and storage in support of enterprise-control system integration. Transactions occur at all levels within the enterprise and between enterprise partners, and are related to both required and actual activities, but the focus of this standard is the interface between enterprise/business systems and manufacturing systems.

4. OPC STANDARD

OPC (Object Linking and Embedding for Process Control) is the interoperability standard for the secure and reliable exchange of data in the domain of industrial automation. It ensures the seamless flow of information among devices from multiple vendors. The OPC Foundation is responsible for the development and maintenance of this standard. (OPC Foundation, 2014.)

The OPC standard is a series of specifications developed by industry vendors, end-users and software developers. These specifications define the interface between clients and servers, as well as servers and servers, including access to real-time data, monitoring of alarms and events, access to historical data and other applications. When the standard was first released in 1996, its purpose was to abstract PLC specific protocols (such as Modbus, Profibus, etc.) into a standardized interface allowing HMI and SCADA systems to interface with a middleman who would convert OPC read and write requests into device-specific requests and vice versa (see Figure 4.1). (OPC Foundation, 2014.)

The OPC Classic specifications are based on Microsoft Windows technology using the COM/DCOM (Distributed Component Object Model) for the exchange of data between software components. The specifications provide separate definitions for accessing process data, alarms and historical data.

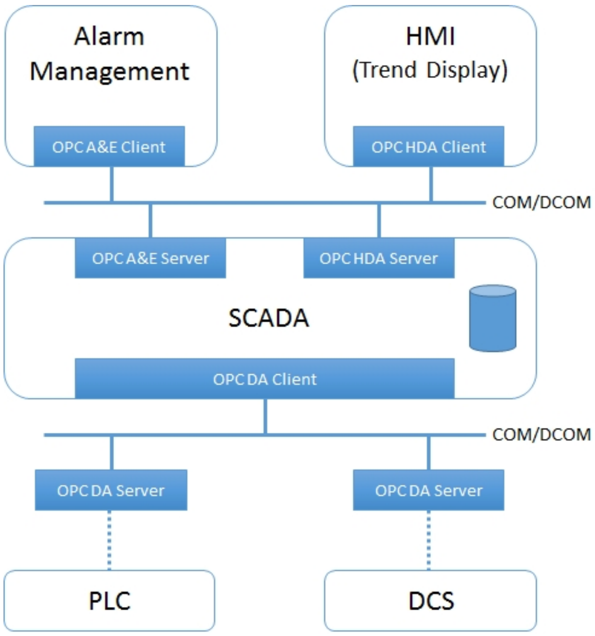

Figure 4.1. Typical use case of OPC Servers and Clients.

The OPC Classic specifications have served the OPC community well. The OPC Foundation has listed over 1,500 OPC-based products in its product catalog containing only products from OPC members. The total OPC market has over 2,500 vendors providing over 15,000 OPC-enabled products. However, as technology evolved, so did the need for OPC specifications. The Chapter 4.4 introduces the OPC Unified Architecture specification, the successor to OPC Classic.

4.1 OPC Data Access

The OPC Data Access (OPC DA) interface enables reading, writing and monitoring of variables containing current process data. The main use case is to transfer real-time data from PLCs, DCSs and other control devices to Human-Machine Interfaces (HMI) and other display clients. OPC DA is the most important of the OPC interfaces: it is implemented in 99% of the products using OPC technology today. The OPC DA Clients define the variables they want to read, write or monitor in the OPC server. OPC provides real-time data that may not be accessible all the time, for example in case of a communication interruption. The OPC Classic technology handles this issue by providing timestamp and quality status for the delivered data. The quality status can be accurate (good), not available (bad) or unknown (uncertain). (Mahnke et al., 2009.)

4.2 OPC Alarms and Events

The OPC Alarms and Events (OPC A&E) interface enables the reception of notifications of alarms and events. Alarms are notifications that inform the client about a change in a condition in the process. Events are notifications that inform the client about an occurrence of an event. OPC A&E interface enables also the acknowledgement of alarms. In order to receive notifications, the OPC A&E client connects to the server, subscribes for notifications and then receives all notifications triggered in the server. The OPC client can specify criteria for filtering the notifications. (Mahnke et al., 2009.)

4.3 OPC Historical Data Access

The OPC Historical Data Access (OPC HDA) provides access to the data already stored. There are three different methods of reading the history data. The first mechanism reads the raw data from the data archive. The OPC HDA client defines the desired variables and the time domain. The OPC HDA server returns all archived values for the specified time range up to the maximum numbers of values defined by the client. The second mechanism reads values of desired variables for specified timestamps. The third mechanism computes aggregate values (e.g. average, count, standard deviation, variance, minimum, maximum, range, duration) from data in the history database for specified time domain for desired variables. Values include always the quality status and timestamp. In addition, OPC HDA also defines methods for inserting, replacing and deleting data in the history database. (Mahnke et al., 2009.)

4.4 OPC UA standard

With the successful adoption of OPC in thousands of products, OPC is used today as standardized interface between automation systems in different levels of the automation domain. In manufacturing industry OPC is used also in areas where it was not designed to be used, and there are many more areas where OPC is wanted to be used. However, there are limitations to the use of OPC because of the COM/DCOM dependency. In the year 2008, the OPC Foundation released OPC Unified Architecture (OPC UA), a platform independent service-oriented architecture that integrates all the functionality of the existing OPC Classic specifications, and is backward compatible with OPC Classic. Several factors influenced the decision to create OPC UA:

- Microsoft has de-emphasized COM (Component Object Model) and DCOM (Distributed COM) in favor of cross-platform SOA (Service-Oriented Architecture).
- OPC vendors want a single set of services to expose the OPC data models, such as Data Access, Alarms & Events, Historical Data Access, etc.
- To stay competitive, OPC vendors need to implement OPC on non-Microsoft systems, including embedded devices.
- Other collaborating organizations need a reliable, efficient way to transport high-level structured data.
- Users require the ability to access OPC servers through firewalls in a secure manner.

The multi-layered approach of the OPC UA specification accomplishes these design specification goals: functional equivalence (i.e. all COM OPC Classic specifications are mapped to UA), platform independence (wide range of support for different platforms from an embedded micro-controller to cloud-based infrastructure), security (encryption, authentication, and auditing), extensible (i.e. ability to add new features without affecting existing applications) and comprehensive information modeling for defining complex information. (OPC Unified Architecture, 2014)

OPC UA is designed to enhance and exceed the capabilities of the OPC Classic specifications. OPC UA is functionally equivalent to OPC Classic, yet capable of much more:

- **Discovery**: find the availability of OPC Servers on local PCs and/or networks
- **Address space**: all data is represented hierarchically (e.g. files and folders) allowing for simple and complex structures to be discovered and utilized by OPC Clients
- **On-demand**: read and write data/information based on access-permissions

- **Subscriptions**: monitor data/information and report-by-exception when values change based on a client's criteria
- **Events**: notify important information based on client's criteria
- **Methods**: clients can execute programs, etc. based on methods defined on the server.

Integration between OPC UA products and OPC Classic products is easily accomplished with COM/Proxy wrappers. (OPC Unified Architecture, 2014.)

Platform independence is essential because of the wide array of available hardware platforms and operating systems. OPC UA functions on the following hardware platforms: traditional PC hardware, cloud-based servers, PLCs, micro-controllers (ARM etc.) and more. Supported operating systems are Microsoft Windows, Apple OSX, Android, or any distribution of Linux. OPC UA provides the necessary infrastructure for interoperability across the enterprise, from machine-to-machine, machine-to-enterprise and everything in-between. (OPC Unified Architecture, 2014.)

One of the most important considerations in choosing a technology is security. OPC UA is firewall-friendly while addressing security concerns by providing a suite of controls:

- **Transport**: numerous protocols are defined providing options such as the ultra-fast OPC-binary transport or the more universally compatible SOAP-HTTPS.
- **Session Encryption**: messages are transmitted securely at 128 or 256 bit encryption levels.
- **Message Signing**: messages are received exactly as they were sent.
- **Sequenced Packets**: exposure to message replay attacks is eliminated with sequencing.
- **Authentication**: each UA client and server is identified through OpenSSL certificates providing control over which applications and systems are permitted to connect with each other.
- **User Control**: applications can require users to authenticate (login credentials, certificate, etc.) and can further restrict and enhance their capabilities with access rights and address-space "views".
- **Auditing**: activities by user and/or system are logged providing an access audit trail.

  (OPC Unified Architecture, 2014.)

The multi-layered architecture of OPC UA provides a "future proof" framework. Innovative technologies and methodologies such as new transport protocols, security algorithms, encoding standards, or application-services can be incorporated into OPC UA while maintaining backwards compatibility for existing products. OPC UA products built today will work with the products of tomorrow. (OPC Unified Architecture, 2014.)

The OPC UA information modeling framework turns data into information. With complete object-oriented capabilities, even the most complex multi-level structures can be modeled and extended. Data types and structures are defined in profiles. For example, the existing OPC Classic specifications were modeled into OPC UA profiles which can also be extended by other organizations. (OPC Unified Architecture, 2014.)

It is expected that the OPC UA will replace the OPC Classic specifications. Because of its platform-independence and use of state-of-the-art web service technology, it is expected that the OPC UA will be applied in a wider range of industries and application, compared to the OPC Classic. (Mahnke et al., 2009.)

5. DATA MANAGEMENT

Data management is an administrative process by which the required data is acquired, validated, stored, protected, and processed, and by which its accessibility, reliability, and timeliness is ensured to satisfy the needs of the data users. The official definition provided by Data Management International (DAMA), the professional organization for those in the data management profession, is:

"Data Resource Management is the development and execution of architectures, policies, practices and procedures that properly manage the full data lifecycle needs of an enterprise."

An alternative definition provided by the DAMA Data Management Body of Knowledge is:

"Data management is the development, execution and supervision of plans, policies, programs and practices that control, protect, deliver and enhance the value of data and information assets."

Data management has to include technology advances in hardware, storage, networking, and computing models such as virtualization and cloud computing. The convergence of emerging technologies and reduction in costs for everything from storage to compute cycles have transformed the way data can be used and made new opportunities possible.

5.1 Data quality

The significance of data quality in company's decision making and operative functions will continue growing day by day as the amount of available data increases enormously. During the past decades information systems have transformed from centralized and isolated systems into networked structure. The information sharing has become more diverse between the information systems inside the company, and more often also with external stakeholders such as sub-contractors and authorities. Outcome of the overwhelming amount of data is that the quality of data deteriorates quickly without appropriate quality control in both the data processing functions and in the information sources. The diversity and distribution of the information sources enables new methods of controlling the data quality. For example it is possible for organization to prefer the information source with the best data quality, or to find errors in information by comparison of the available data from different sources.

A data quality dimension is a recognized term used by data management professionals to describe a feature of data that can be measured or assessed against defined standards in order to determine the quality of data. Batini lists the following items as the data quality dimensions (Batini et al., 2006):

- **Accuracy**: A good starting point for discussing accuracy is the entity-attribute-value model, which serves as the conceptual basis for most database implementations of real world phenomena. According to this model, entities represent real world phenomena (such as defects in the paper web), attributes specify the relevant properties of these objects (such as location of the defect or size of the defect), and values give the specific qualitative or quantitative measurements belonging to a particular attribute.
- **Completeness**: Completeness is the ability of a data set to represent every meaningful state of the represented real world system.
- **Consistency**: Consistency refers to the absence of apparent contradictions in a dataset. Consistency of data values occurs if there is more than one state of the information system matching a state of the real world system.
- **Interpretability:** Interpretability concerns the documentation and metadata that are available to interpret correctly the meaning and properties of data sources.
- **Accessibility:** Accessibility measures the ability of the user to access the data.
- **Usability:** Usability concerns the ability of the user to make use of the data.
- **Trustworthiness:** Trustworthiness of a data source measures how reliable is the data source providing the data.

5.2 Classification of data

A basic classification for different types of data is by Batini (Batini et al., 2006):

- **Structured**, when the data are distinguished in elementary items, and each of them is represented with a format that can be described by a grammar.
- **Semi structured**, when data has a structure which has some degree of flexibility. Semi structured data are also “schema less” or “self-describing” data.
- **Unstructured**, when data are expressed in natural language and no specific structure or domain types are defined.

Other classification for data can be done by the time dimension of the data. According to its change frequency, data can be classified into three categories. Stable data is unlikely to change (e.g. scientific publications), long-term-changing data has very low change frequency (e.g. addresses, currencies). The concept of low frequency is domain dependent; in an e-trade application, if the value of a stock quote is tracked once an hour, it is considered to be a low frequency change, while a shop that changes its goods weekly has a high-frequency change for clients. Frequently-changing data has intensive change, such as real-time traffic information, temperature sensor measures, and sales quantities. The changes can occur with a defined frequency or they can be random.

5.3 Data harmonization

The purpose of data harmonization is to have the same information represented in one uniform way. Data harmonization is often needed in such applications where we need to manipulate beforehand unknown data sources whose contents, structures and naming conventions are also unknown. Nowadays the data often need to be extracted from multiple data sources controlled by different applications. Because these data sources have not been originally designed to share their data, their integration is a challenge.

Semantic heterogeneity means that, in two or more data sources, data with the same intended meaning are represented in varying ways or data with different intended meanings are represented in identical ways. If information needs and data sources are known in advance, a popular approach to solving problems caused by semantic heterogeneity is to construct a global schema as a standard view over all available data sources.

Technological heterogeneities are due to the use of products by different vendors, employed at various layers of an information and communication infrastructure. An example of technological heterogeneity is the usage of two different relational database management systems. (Batini et al., 2006.)

Schema heterogeneities are principally caused by the use of (1) different data models, such as one source that adopts the relational data model and a different source that adopts the XML data model, and (2) different data representations, such as one source that stores addresses as one single field and another source that stores addresses with separate fields for street, zip code and city. (Batini et al., 2006.)

Instance-level heterogeneities are caused by different, conflicting data values provided by distinct sources for the same objects. This type of heterogeneity can be caused by quality errors, such as accuracy, completeness, currency, and consistency errors; such errors may result, for instance, from independent processes that feed the different data sources. (Batini et al., 2006.)

6. DATA MANAGEMENT OF AUTOMATION SYSTEM

Automation system produces enormous amounts of data. Data is collected from the production process, automation equipment and several other functions related to production. Data is used in reporting, product quality control, maintenance of the automation equipment and improving the production process.

Several important issues have to be taken into consideration when planning and implementing data management. Properties of data quality are for example data accuracy and availability. Different kinds of automation applications must be able to access the data and applications also must be able to trust the data. In case of faulty data, reports produced by the application are incorrect. The consequences can be expensive; in the worst case, the whole production must be shut down when trying to find the source of the incorrect information.

Automation application can mean either an application located in the levels 1 and 2 of the ISA-95 hierarchy model, or an application located in the level 3 of the ISA-95-model. Applications on the lower levels are the one controlling the automation process and producing all the process data. Upper level applications control the manufacturing functions and produce the visualization of the data from lower levels and reporting for business oriented functions of the organization. Common issues related to the data management of any automation application are security services (such as user management, access control, data security and validation), configuration services (such as device configuration, engineering e.g. version management of hardware, applications and interfaces), backup of data and configurations, audit trail and documentation services (such as product information and project documentation).

6.1 Automation applications on ISA-95 levels 1 and 2

Automation application on ISA-95 levels 1 and 2 consists typically of controller, sensors and actuators, operating and monitoring station. Controller has no storage and therefore a PC workstation is needed for storing the data. Typically controllers are running on a Linux-based operating system and PC workstations on a Windows operating system.

These applications have different kind of data. First of all there is file based data that determine the function of the application. Automation application development teams provide this file based data in from of software releases. File based data is also needed in the startup of the applications.

In current automation application, tags, i.e. identifiers for objects in the automation application, are typically stored in flat lists. The modern way is to have hierarchical object model for the tags. This way the querying of tags is more structured and access to tags is more straightforward than with a flat list. The IEC 81346-1:2009 standard, published jointly by IEC and ISO, establishes general principles for the structuring of systems including structuring of the information about systems. The reference designation identifies objects for the purpose of creation and retrieval of information about an object. SFS-EN 81346-1 standard contains a Finnish translation of the English text of the standard. The European Standard EN 81346-1:2009 has the status of a Finnish national standard (SFS-EN 81346-1, 2010).

Data management of application on ISA-95 level 1 deals with data from analyzers and automation devices. Information of analyzer measurements, events and device parameters and diagnostics are collected and stored for further use. Applications on ISA-95 level 2 provide information about process (events and alarms) and perform calculations with the data based on process models.

6.2 Automation applications on ISA-95 level 3

Different kind of reporting and information services located near ISA-95 level 3 MES applications, such as quality control system and batch control system, provide the data for user interaction services. Long-term history data is visualized for example in trends and quality maps.

Data management study for automation applications was carried out by arranging meetings with application head designers or other application experts. In the meetings, application experts introduced the applications and described the data management of the applications. The documentation of the automation applications was also studied. The following sub-sections introduce some of the applications included in the study. All the information is available at www.metso.com website.

6.2.1 Metso IQ Scanner System

Metso IQ Scanner is designed to produce accurate and stable sheet quality measurements in the hot, humid and dusty environment of a paper machine. In Figure 6.1 unique stiff beam design combines low scanner height with high mechanical stiffness.

Figure 6.1. Metso IQ Scanner frame with sensor platform.

In addition to performance diagnostics, the intelligence built into Metso IQ Scanner also allows its scanning speed, acceleration and deceleration rates to be adjusted and controlled to suit the CD and MD control requirements. Edge to edge scanning is controlled by precise sheet edge detectors. The sensor platform employs a modular design, using standard sized slots for Metso IQ sensors. Sensors are therefore easily interchanged and new sensors can be easily added and plugged into IQ Bus, the common utility supply system.

6.2.2 Metso PQV System

The unified Metso Process and Quality Vision system is a multi-functional tool for analyzing the origins of breaks and defining the source of coarse or very fine and complex paper web faults that do not necessarily cause breaks but nonetheless have to be eliminated before they become customer problems. Errors in the paper web can be cut-out or patched. Based on the information of the error map, customer rolls can be cut on the winder based on quality criteria in order to maximize the value of the reel (tambour). Thus, the system has a dual purpose: to improve paper web runnability and to help the mill to ship the best fault-free customer quality.

Figure 6.2. Metso PQV web inspection system integrated into paper machine.

The digital matrix cameras take an instantaneous snapshot of the full field of view of each camera, so all pixels are captured at the same time. The camera captures about 95% of the light from the flashing LED light source which illuminates the sheet (see Figure 6.2). It looks continuous to the eye, but the LED light beam is actually flashing at a very fast rate, which is adjustable from 200 microseconds down to 5 microseconds. The high frequency snapshots of the paper web are digitally combined in the image processing system to eliminate overlapping of the sequential images, thereby producing a continuous view of the moving web.

6.2.3 Metso DNA Paper Quality Monitoring

Metso DNA Paper Quality Monitoring (Metso PQM) is a tool for mill-wide quality management and reporting. It can combine process data and quality data from process controls, quality controls, web inspection systems and laboratory testing. Metso PQM improves production and quality efficiency by providing paper web sheet quality information visualized with easy-to-use, online graphic user interface for operators and mill management.

6.2.4 Metso DNA Historian

Metso DNA Historian is a process management tool that records data related to materials, process conditions and quality properties in a pulp and paper mill. Information becomes available to the whole organization, enabling better collaboration and mill activity support. Metso DNA Historian collects time-series data from the production process. History process data provided by Metso DNA Historian is used in long-term trends, disturbance and alarm figures, and statistical reporting provide examples of tools that flexibly combine process information for future improvement.

6.2.5 Metso DNA Field Device Manager

Metso DNA Field Device Manager provides configuration and maintenance capabilities for field equipment. It allows for the efficient commissioning of field devices and control applications at the same time. It ensures quick access to relevant information, and helps its users to find the right solutions rapidly in device malfunctions. As such, it supports proactive field asset management and enables effective planning of future maintenance work. Configuration and maintenance information is available for both process operators and maintenance people in the office.

6.2.6 Metso DNA Machine Monitoring

Metso DNA Machine Monitoring measures and analyzes the mechanical condition and performance of machines, based on vibration measurements and other machine parameters. Metso DNA Machine Monitoring provides both protection and diagnostics tools for critical machinery, as well as condition monitoring and analyzing tools for predictive maintenance use. Online machine condition monitoring enables continuous monitoring, thus providing the fastest possible way to act on problems to secure plant availability, protect assets, provide information for maintenance planning and increase working environment safety.

6.2.7 Metso Paper Lab

Metso Paper Lab Automated Paper Testing Laboratory provides paper and board makers with more possibilities to manage the quality of the products. Paper testing function can be moved from the laboratory right next to the production line. Metso Paper Lab is suitable for testing all grades of paper and board and it offers the widest selection of industry-standard tests with over 400 reported properties.

6.2.8 Metso DNA Engineering Environment

The Metso DNA Engineering environment contains tools to engineer and maintain plant automation. A state-of-the-art engineering and maintenance tool enables scalable multi-user environment for concurrent engineering. This single tool allows life cycle management for all control applications, field buses, and field devices, as well as network document management. In addition, advanced diagnostic tools offer easy maintenance support for both applications and hardware components. The environment is designed so that the engineering server (or engineering client) is connected via the network to the backup server. The backup server is connected via the process/control room bus to the other application servers (e.g. process control servers).

6.3 Automation application data

Automation application requires data in order to function properly. Configuration data for automation applications is read from automation engineering database. Configuration data consists of for example default values for various parameters and safety limit values. During the production process, data from the process itself and produced products is captured into several databases.

6.3.1 Relational data

Automation applications handle data of various entities that have relations with one another. For example in paper making there are paper quality applications that take measurements of the paper web running at speed of 2,000 m/minute. The paper web is wound up to a parent reel that can be over 90 km long. The parent reel needs to be split into paper web sections suitable for customers. The properties of the paper (e.g. brightness, gloss, opacity, smoothness, surface strength and ink absorption) depend on the paper grade defined for the paper and the coating methods used in production. (KnowPap.) In addition to these are production related entities such as work shift, working hours, production personnel, etc.

All entities mentioned in the example above (e.g. quality measurements, parent reel, customer roll, paper grade) have relation to each other. The data and relations between the data items are stored into relational database tables.

6.3.2 Time-series data

Process history data is time-series data which is stored into a historian database. Historian databases include built-in data collection capabilities and they can capture data from process stations that has connections to sensors via field buses. Historian database can collect large volumes of real-time plant floor information from various process devices at incredibly high speeds. It is ideal for capturing data from sensors and other real-time systems because it uses manufacturing standards such as OPC, which facilitates communications by providing a consistent method of accessing data across devices.

The performance of conventional relational databases is suitable in terms of supporting simple operator queries such as viewing recent value trends in flows, temperature, or other analog values. Plant-wide historian databases provide a much better read and write performance over a relational database and a millisecond resolution for true real-time data. This capability enables better responsiveness by quickly providing the granularity of data needed to analyze and solve intense process applications.

With a relational database the maintenance alone can be a full-time job because of continually managing archives and disk space due to the lack of compression; performance can be severely undermined, even with proprietary, pre-compressed data workarounds. However, the powerful compression algorithms of plant-wide historians enable storing years of data easily and securely online, which enhances performance, reduces maintenance and lowers costs.

Naturally some information is lost due to the compression. However, the savings on storage achieved by compression are impressive. In most cases the compression savings between a relational database and a properly tuned Swinging Door Algorithm in a process historian is between 1:1000 and 1:5000, thus a 1 gigabyte process historian database is a 1 to 5 terabyte database in a relational database, which is significant. (OSI PI Blog.)

Automation system data collection interval varies from 10 ms to 1 hour. The amount of measurement points (also known as tags) is typically between 15,000 and 30,000, but can be over 50,000 in large automation systems.

6.4 Usage of automation application data

Data produced by automation applications is used in several ways in many different functions of automation system and manufacturing company. Automation system uses the data in adjusting the production process. Based on different measurements performed by analyzers and quality control applications, the efficiency of the process is optimized. The objective of the optimization is to maintain the desired quality level by using as little resources as possible. Economical use of raw materials and energy is environmental friendly and saves money for the company.

Metso DNA Operate is a front-end for process operators. It facilitates efficient alarm operations with 1 ms resolution showing the abnormalities in the production process. The sequence of events is then used in finding the roots of the disturbance in the process.

Other usage for the collected data is reporting purposes. Metso DNA Report is end-users' reporting environment with report development tools. Metso DNA Report ensures a seamless interaction between people and processes in plant, mill or corporationwide organizations. It contains all reporting, analyzing and community tools individually set up for each user. Metso DNA Report is a web portal which runs in a web browser. Metso DNA Report has basic report set: log, summary, runtime and total reports. It also has report scheduler for automated printing and e-mailing.

One specific example use for Metso DNA Report is emission monitoring for large combustion plant. Several European Union directives will require improvement in environmental performance during the next few years. Metso DNA LCP Emission Monitoring application provides the power plant environmental and operating personnel with the necessary information about the environmental compliance of power plants both in real time and in history (see Figure 6.3). In order to produce the figures reported to competent authorities, the application refines the emission measurement data into required values in expected units.

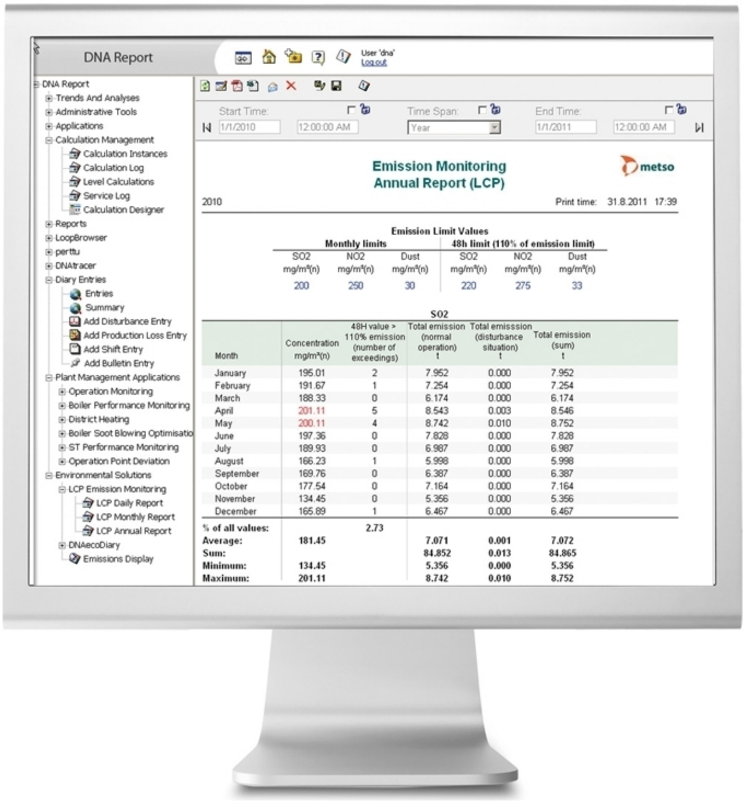

Figure 6.3. Metso DNA Report used in emission monitoring.

Another example of the DNA Report tools is Tracer, which is a process history analyzing tool. In Figure 6.4 is a screenshot of DNA Tracer that combines traditional trends with alarms and events in the same picture with a common timeline. Data and events can be for example diary entries, intelligent field device alarms, mechanical condition monitoring data or operation events. It helps users immediately to see what happened in a certain process situation and how the sequence of events proceeded.

Figure 6.4. Metso DNA Tracer automatically combines trends, alarms, operations, messages and diary entries of the selected tags in the same picture.

Metso DNA Report Designer is a graphical tool that enables users to create web reports without programming skills or knowledge of different databases. The user can change the report template or make new templates and related data queries using the DNA Report Designer.

6.5 Databases

Usage of databases in Metso DNA automation applications is diverse because of the heterogeneous nature of the automation application data. Different types of automation application data are time based data (process measurements), structured data (attributes of application specific entities) and unstructured data (diary, images, video).

6.5.1 Relational database

Relation databases are used both in process control station (PCS) level and in enterprise (MES, ERP) level. In embedded industrial controllers there are certain limitations of memory, storage and processor performance. Therefore lightweight solutions are preferred. In many cases there is no backup functionality on PCS level and the collected data might be lost in case of a power outage.

The most common solution for enterprise level application is to use Microsoft SQL Server for relational data. Microsoft SQL Server includes support for structured and semi-structured data, including digital media formats for pictures, audio, video and other multimedia data. Multimedia data can be stored as BLOBs (binary large objects). For example images from camera based paper quality control application are stored in relational database. Due to the massive amount of data storage consumed by the videos, only the link of the video file on storage is typically stored in database.

6.5.2 Historian database

Metso DNA Historian database uses proprietary software solution. DNA Historian is a high-performance data management system for collecting time-continuous (real-time) process data. DNA historian includes process interfaces for Metso specific engineering and maintenance systems such as Damatic XD, Metso DNA and maxDNA systems. For third party systems, the data collection method is OPC DA client. With DNA Historian up to 50,000 tags and all alarms and events can be collected online. The length of the history is limited only by the available hard disk space. Old data can be archived for later usage. Typical collection cycles are 10 s and 1 s, but with special applications, the collection cycle can be down to 100 ms. The small collection cycle and high amount of tags require high performance server used as hardware platform for the historian database.

7. RESULTS AND DISCUSSION

Data management of automation applications is diverse. Key issues related to the data management of applications near the process are sorting the data to the user and optimization of performance and data storage. Sorting of the data means that required data can be located and accessed quickly and with minimum effort.

7.1 Data management problems

One of the objectives of the automation application study was to find out the problems in the current data management of the automation applications. The problems were found in many levels of the automation system.

7.1.1 Performance vs. storage usage

In the optimization of the storing the data, some kind of tradeoff between processing performance and storage usage must be done. In order to save storage space, higher compression is used in storing the data i.e. data is stored in lower frequency when applicable. However, performance of processing the data is lowered at the same time. This is due to the time-consuming interpolation of the data points in order to fill in the gaps between stored values when the data is retrieved from the storage. Also with higher compression ratio, the accuracy of the statistical aggregations calculated based on the compressed data values, such as mean value and standard deviation, is controversial.

With lower compression ratio, more storage is used in storing the data. The processing the data is quicker because the interpolation of the missing data points is not needed. The statistical aggregations calculated based on the higher data amount are more accurate.

7.1.2 Scalability

Scalability of the automation application in different sized system deliveries can be an issue that is hard to tackle. Automation applications can be designed in a way that specific operating system or certain database is needed. These proprietary software products have certain requirements for the hardware they are installed on. For example historian database implementation typically needs to have a high-performance server hardware and redundant functionality such as more than one power supplies and hard disks. The process history data collecting should not become disturbed in case of some hardware failure.

In a small-sized automation system delivery the additional hardware products have a big part in the total cost. Because companies want to win the bidding competitions, all the additional costs need to be minimized in the offers. Therefore it would be beneficial if the automation applications would function in different scale environments. In some cases not all the functionalities of the automation applications are needed and therefore some functional limitations would be acceptable.

7.1.3 Costs of commercial software products

Currently available Metso DNA automation applications on the MES level typically use Microsoft SQL Server as a relational database management system. SQL Server also includes an assortment of add-on services that either run as a part of some SQL Server component or out-of-process as Windows Service and present their own API to control and interact with them. Because of this, Microsoft Windows has been chosen to be the operating system in many automation applications. The license fees of these products have a big part in the total cost of a small-sized automation system delivery. Utilization of Open Source software products, for example Linux operating system and databases could be a solution to decrease the software costs.

7.1.4 Lack of information sharing between projects

In the automation application study it was found that information sharing between different automation application development teams has some weaknesses.

In some cases several application development teams have developed very similar solutions to the same kind of problems without knowing the other team's efforts. This kind of action spends the resources of the development teams and results in more difficult maintenance because of slight differences in the outcomes.

On the other hand there are cases where application development teams have developed diverse solutions to the same kind of problem. The different solutions use different databases and interfaces. This is also demanding for the maintenance personnel because they need to be able to handle maintenance issues in very heterogeneous software field.

7.1.5 Version and revision control

In some of the automation applications included in the study, there is a need for better version and revision control system. Without proper version and revision control, some important information might be lost due to the overwriting data in the application. An example of such situation could be a change in fuel recipe. If parameters of the recipe are changed without controlling the versions of recipes, the outcome is that the previous version of the recipe is not stored into the version control system. Therefore it is difficult to reproduce the process and have the exact same product. Another related item for improvement is the traceability of changes made in application. It means that some kind of tracking and recording of application user's actions would be useful.

7.1.6 Usability

The appearance of Windows client applications is a little outdated and the user interface could be more flexible. Web based services could provide more flexible customization of the front-end of operating and monitoring applications. The utilization of web based services would also enable wider selection of suitable platforms and operating systems. It would not matter if an operator is using a laptop with Mac OS or a tablet computer with Linux for monitoring the production process.

7.2 Open Source

Open Source software products have been utilized widely during the past few years. Many Open Source based software products would be feasible, however, there are difficult issues related to the utilization of Open Source products. The license terms of Open Source products and the lack of necessary support are often showstoppers for choosing them. The most widely used Open Source license, the General Public License (GPL), requires that companies which base their software on Open Source software licensed under the GPL must make available the source code of the company's software to the company's licensees and to permit such licensees to modify and redistribute the company's software without charge to the third parties.

There are also doubts about critical issues such as performance, reliability and scalability of the Open source products. In many cases the insufficient information and the lack of time prevent data management experts to perform an in-depth research on Open Source alternatives. (Riesco, 2006.)

7.3 Cloud based services

This thesis also investigated the requirements for automation system data management coming from possible future technologies. Cloud based services enable the development of new business models. From data management point of view there are many demanding issues that need to be considered when utilizing cloud technology.

Cloud computing provides convenient, on-demand network access to a shared pool of configurable computing resources including networks, servers, storage, applications, and services. These resources can be rapidly provisioned and released with minimal management effort or service provider interaction. This way cloud service customers do not have to purchase the hardware equipment and software from the store, set the computing environment up and organize its own IT administration to maintain the resources.

A cloud can be public, private or hybrid. A public cloud infrastructure is owned by a cloud provider and cloud services are sold to the customers. A private cloud infrastructure is operated solely for a specific customer. It may be managed by the customer or by a third party; it may exist on premise or off premise. Hybrid clouds consist of private and public clouds that remain unique entities but are bound together by standardized or proprietary technology that enables data and application portability.

Communication between heterogeneous systems has many problems. One of them is the serialization of data i.e. how the data is exchanged between different systems and services. Early implementations used ASN.1 (Abstract Syntax Notation) format for representing the data exchanged. More commonly used format today is XML, which encodes data into ASCII-formatted documents. In the recent years, the JSON (JavaScript Object Notation) has grown its popularity. JSON is an open standard format that uses human-readable text to transmit data objects consisting of attribute-value pairs. Despite its name, JSON is a common data interchange format and code for parsing and generating JSON data is available in a large variety of programming languages. Although JSON is intended solely for data serialization, its design as a non-strict subset of the JavaScript scripting language poses several security concerns. These concerns center on the use of a JavaScript interpreter to execute JSON text dynamically as embedded JavaScript. This exposes a program to errant or malicious scripts. This is a serious issue when dealing with data retrieved from the Internet.

The fact that cloud service is available via internet connection is beneficial for the automation systems and applications located in rural environment. With a relatively small cost of a satellite modem or a cellular data connection, the applications without any fixed internet connections are able to access cloud based services. Example of such application is a mobile crusher that can operate in distant location with only cellular network data communication possibilities. In Figure 7.1 is a Metso LokoTrack LT106 crusher in its demanding and rugged working environment.

Figure 7.1. Metso Lokotrack is market-leading solution for mobile crushing and screening.

7.3.1 Cloud computing stack

There are currently three cloud based service models. Software as a Service (SaaS), where the end-user uses an application, but does not control the operating system, hardware or the network infrastructure. In this situation, the user executes applications over the network. Next is Platform as a Service (PaaS), where the end-users host an environment for their applications. The end-users control the applications, but do not control the operating system, hardware or network infrastructure, which they are using. Finally, there is Infrastructure as a Service (IaaS), where the end-user accesses computing resources such as CPU, memory, middleware and storage. The end-user controls the resources, but not the cloud infrastructure beneath them. (Antonopouos et al., 2010.) The roles and work share between cloud provider and company's own IT administration have to be clear in order to have a working collaboration.

Cloud service providers try to provide simplified software installation, maintenance and a centralized control over the software used. The end-users can access the cloud-based services anytime and from anywhere. Naturally, this type of access is based on the bandwidth that a user has over the Internet and therefore poor interconnections mean that the use of cloud-based resources is not viable. (Antonopouos et al., 2010.) Even though most of current cloud-based systems use different APIs and protocols in the internal implementation of the cloud architecture, sharing and co-operating within cloud and between internet services by using REST (Representational State Transfer), an http application protocol based API, has become the norm. In Figure 7.2 is depicted the cloud protocol stack with REST layer on top of http application protocol.

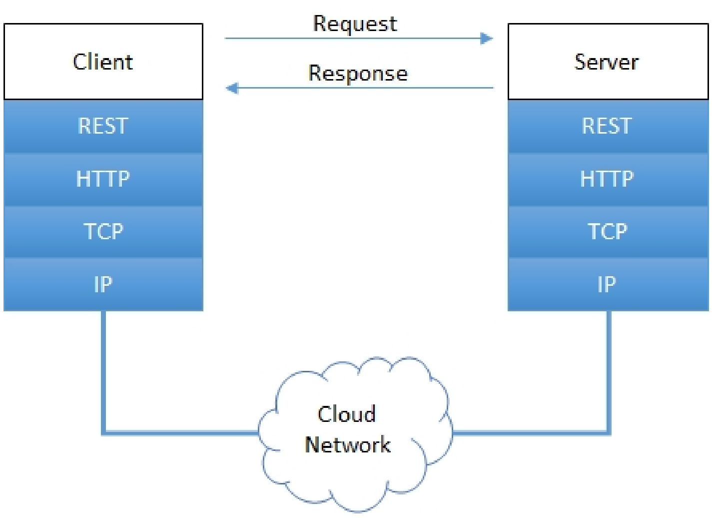

Figure 7.2. Cloud protocol stack.

7.3.2 Cloud security

Organizations have traditionally been uncomfortable with the idea of storing their data, and applications on system infrastructure and services they do not control. On the other hand, today increasing number of medium-sized and large organizations are relying their business critical data to cloud providers with email cloud services of Microsoft and Google showing the way. In addition, a shared infrastructure increases the potential for unauthorized access and exposure of sensitive data. In case of a business application service, company who is providing the cloud based service has their customer's data to be concerned. Cloud based systems need to secure procedures concerning authentication, identity management, compliance and access-related technologies, which are becoming increasingly important. Within the cloud security model, the end-user needs to be able to trust the cloud provider's security model. In addition, it may be a case that an end-user wants to use sensitive data on a cloud based system, in which case it would be useful to be able to encrypt the data so that it is safe and cannot be stolen by other users. (Antonopouos et al., 2010.) The cornerstone of succeeding in cloud based business services is to be able to ensure the security and trust of the customers. Organizations spend a lot of resources maintaining their in-house built IT solutions with infrastructure and software updates and taking care of the customer data. Cloud based solutions provided by external operator could in fact be more secure solution.

The shift towards cloud based services is driven by many factors including ubiquity of access (only web browser is needed to access many cloud services), ease of management (effortless configuration, application level backup mechanisms due to the data distribution), and less investment (affordable enterprise solution deployed on a pay-per-use basis for the hardware and software provided by the cloud providers). Furthermore, cloud computing offers many advantages to vendors, such as easily managed infrastructure because the data center has homogeneous hardware and system software. (Antonopouos et al., 2010.)

7.3.3 Cloud based SCADA

Cloud computing is relatively new functionality for SCADA (supervisory control and data acquisition) applications. By moving to a cloud based environment, SCADA providers and users can significantly reduce costs, achieve greater reliability, and enhance functionality. In addition to eliminating the expenses and problems related to the hardware layer of IT infrastructure, cloud based SCADA enables users to view data on devices like smart phones and tablet computers, and also through SMS text messages and e-mail. (Automation.com.)

Cloud computing can support SCADA applications in two fashions: (1) the SCADA application is running on-site, directly connected to the control network and delivering information to the cloud where it can be stored and disseminated, or (2) the SCADA application is running entirely in the cloud and remotely connected to the control network. The first method is the most common. The control functions of the SCADA application are entirely isolated to the control network. However, the SCADA application is connected to a service in the cloud that provides visualization, reporting, and access to remote users. (Automation.com.)

With cloud technology the scalability of the provided service is dynamic and inexpensive because the resource management does not involve the purchase, deployment, and configuration of new servers and software. If more computing power or data storage is needed, users simply pay more for additional resources. On-demand resource capacity can be used for better resilience when facing increased service demands or distributed denial of service attacks, and for quicker recovery from serious incidents. The scalability of cloud computing facilities offers greater availability. Companies can provision large data servers for online historical databases, but only pay for the storage they're using. (Automation.com.)

7.3.4 Costs of cloud based services

During the application study, several meetings with cloud service providers were arranged. In these meetings various topics related to automation systems and automation application data were discussed and considered from cloud technology point of view. Rough estimates about the massive amounts of cumulative data and pricing of different cloud based services were used to calculate the total costs of an example cloud based service.

One concern that was identified as a result from the meetings was that the selection of the cloud provider and cloud technology must be done carefully in order to avoid the vendor lock-in i.e. becoming too dependent on one cloud provider or technology. By assigning business critical processes and data to a cloud service provider with certain technology, it may get really difficult and expensive to attempt to dislodge from the arrangement if it is needed to change the cloud service provider to another with different technology. Naturally cloud service providers are trying to persuade customers to use various services that they provide, such as

- on-demand instances of virtual computing environments
- different kinds of database services
- usage of data processing services (e.g. MapReduce distributes the computational work across a cluster of virtual servers in the could)
- monitoring services for measuring statistics of cloud resources and applications
- load balancing functionality
- transferring the data into cloud provider's server free of change (however, transferring the data away from cloud provider's server is typically chargeable).

The ease of utilization of these services and relatively low costs makes them even more appealing for the customer to purchase. A good thing with cloud technology is that the evaluation of a cloud based service can be executed in small scale. The evaluation system can be built incrementally by adding more functionalities piece by piece.

During the evaluation improvement ideas of the service architecture can be identified. Based on the evaluation results the decision of developing and launching the full-scale version of the service can be done confidently. If the evaluation results promote against the development of full-scale version, then the evaluation setup can be decommissioned.

7.4 Big data and Internet of Things

Big data is becoming one of the most important technology trends. Big data has the potential for dramatically changing the way organizations use information to enhance the customer experience and transform their business models. Many companies are experimenting with techniques that allow them to collect massive amounts of data to determine whether hidden patterns exist within that data that might be an early indication of an important change. Implementing a big data solution requires that the infrastructure be in place to support the scalability, distribution, and management of that data. (Hurwitz et al., 2013.)

Customers who have invested in an expensive automation system would like to have their production process as optimized as possible in order to maximize the efficiency of the system. Automation system vendor could use big data solutions and cloud computing to analyze the process history data from all of their customers. When optimization items have been discovered, the vendor could offer to advice concerned customers how to adjust their processes.

Improvements in network speed and reliability have removed other physical limitations of being able to manage massive amounts of data at an acceptable pace. In addition, the impact of changes in the price and development of computer memory, it is now possible to imagine ways that companies can leverage data that would have been inconceivable only five years ago. (Hurwitz et al., 2013.)

The Internet of Things (IoT) is a computing concept that describes a future where everyday physical objects will be connected to the Internet and be able to identify themselves to other devices. Internet-enabled things will bring various benefits to both organizations and individuals by facilitating or simplifying environment sensing, proximity triggering, automated sensing and actuation, all of which can be utilized in various application domains, ranging from automated home appliances to smart grids and high-resolution asset and product management.

In terms of business, IoT represents a tremendous opportunity for various types of companies, including IoT application and service providers, IoT platform providers and integrators, telecom operators and software vendors. According to some estimates, M2M communications alone will generate approximately 714 billion euros in revenues by 2020, and many IoT vertical segments are expected to experience a double-digit growth in the upcoming years. Among the most prospective vertical application domains are consumer electronics, automotive, and healthcare, as well as intelligent buildings and utilities. (Mazhelis et al., 2013.)

In case of a production defect, data on unstructured human interactions along with transaction data and tag information from the device and relating them with process measurements, data from line devices, operators, materials, quality data, operator and maintenance activities can all be incorporated and searched so that it is possible to see anything that happened in the production phase during that defect. For example process operator could have written in textual form in the production diary or blog about some weird noise coming from certain machine in the production line. This kind of unstructured information is difficult to include in traditional root cause analysis.

Another example of IoT technology is to have the ability to feed local weather forecast information into the energy management application controlling the company's premises so it can start cooling the building down or heating it up a day in advance.

7.5 NoSQL databases

The relational database model has been dominating since the 1980s, with Relational Database Management System (RDBMS) implementations like Oracle databases, MySQL and Microsoft SQL Servers. In an increasing number of cases the use of relational databases leads to problems both because of deficits and problems in the modeling of data and constraints of horizontal scalability over several servers and big amounts of data.

NoSQL databases (Not Only SQL) refers to a diverse and increasingly familiar group of non-relational data management systems. NoSQL databases are not built primarily on tables, and generally do not use SQL for data manipulation. NoSQL database management systems are useful when working with a huge quantity of data and when the nature of the data is not relational.

NoSQL systems are distributed, non-relational databases designed for large-scale data storage and for massively-parallel data processing across a large number of servers. They also use non-SQL languages and mechanisms to interact with data. There are two trends that bring out these problems to the attention of the international software community. First, the exponential growth of the volume of data generated by users, systems and sensors, further accelerated by the concentration of large part of this volume on big distributed systems like Amazon, Google and other cloud services. Second, the increasing interdependency and complexity of data accelerated by the Internet, social networks and open and standardized access to data sources from a large number of different systems. (Moniruzzaman and Hossain, 2013.)

8. CONCLUSIONS

This thesis studied the data management of automation applications. During the study, several problems in the data management of current automation applications were noticed and a few improvement ideas were discovered. Thesis also investigated the impacts of new technologies, such as cloud based services and big data, to the data management of automation system.

Many of the automation applications are built on Windows platform. Therefore computers that are running Windows operating system are needed at customer's site as well. However, additional license of Windows operating system is needed for every computer intended for automation application use. The utilization of Open Source based operating systems, such as Linux, would decrease the expenses of hardware.

Another reason for using Windows operating system is the Microsoft SQL Server based database solutions used in several automation applications. There are many applicable Open Source database products that could be used instead of Microsoft SQL Server.

The usability of a little outdated Windows client automation applications could be improved and modernized by utilizing web based services for producing the front-end user interface. The utilization of web based services would also improve the portability of the automation applications to other operating systems and platforms such as laptops and tablet computers with Mac OS or Linux operating systems.

Improving the information sharing between automation application development teams would beneficial. Innovations and solutions made inside one development team would be useful for other teams. This would prevent inventing the wheel again, in other words wasting resources in solving the same problem again.

In some automation application improvement was needed in version and revision control system. A proper version and revision control system is needed for better traceability of changes made to application and production process. Version and revision information is needed for example in production reports.

As new technologies, such as cloud computing, big data and Internet of Things, are being utilized, new business models are available for companies that react to the demand in time. Information has value in business only if the information is applicable. Data mining mechanisms and methods are useful also for industrial data. Modern automation systems provide massive amounts of data and that data is being stored in organization's many information systems.

Popular solutions of the social media could be useful also in the automation domain. Customers could for example report abnormalities in their process in web-based maintenance portal. Maintenance personnel could then give advice and instructions for solving the problematic issues. Based on reported customer support cases, similarities in the problems could be found and improvement ideas for the particular product could be found.

In the future, data management challenges will be focused on exploiting the automation application data more efficiently. This can be achieved by providing new cloud based services for customers in order to optimize their process performance. Utilization of the new technologies, such as big data and Internet of Things, can provide significant competitive advantages and improve the profitability of the data management functionalities of the organization.

REFERENCES

Aalto University. Thousands of automation devices in Finland vulnerable to cyber attacks. [WWW]. [Accessed on 4.5.2014]. Available at:
http://www.aalto.fi/en/current/news/2013-03-20/

ABB Oy. TTT-käsikirja 2000-07, Luku 24. Prosessiautomaatio. [WWW]. [Accessed on 30.1.2014]. Available at:
http://cna.mikkeliamk.fi/Public/JormPekk/ABB/TTT-KIRJA-2000/PROSESSIAUTOMAATIO.pdf

Antonopoulos, N., Gillam, L. 2010. Cloud Computing: Principles, Systems and Applications. Springer-Verlag London Limited. 379 p.

Automation.com. Cloud Computing for SCADA - A White Paper from InduSoft. [WWW]. [Accessed on 5.5.2014]. Available at:
http://www.automation.com/pdf_articles/Whitepaper_CloudComputing.pdf

Batini, C., Pernici, B., 2006, in IFIP International Federation for Information Processing, Volume 214, The Past and Future of Information Systems; 1976-2006 and Beyond, eds. Avison, D., Elliot, S., Krogstie, J., Pries-Heje, J., (Boston; Springer), pp. 51-62.

Brandl D. The IT Implications of ISA 95 and ISA 99, 2005. [WWW]. [Accessed on 5.2.2014]. Available at:
http://www.brlconsulting.com/Files/The%20IT%20Implications%20S95%20and%20S99.pdf

CNet.com. Google closes $3.2 billion purchase of Nest. [WWW]. [Accessed on 6.5.2014]. Available at:
http://www.cnet.com/news/google-closes-3-2-billion-purchase-of-nest/

DAMA. Data Management International. [WWW]. [Accessed on 6.5.2014]. Available at:
http://www.dama.org

Groover, M.P. 2008. Automation, Production Systems, and Computer-Integrated Manufacturing. Third edition. London, Prentice-Hall International. 831 p.

Hurwitz J., Nugent A., Halper F. and Kaufman M. 2013. Big Data For Dummies. New Jersey, John Wiley & Sons, Inc. 339 p.

ISA.com. The International Society of Automation. [WWW]. [Accessed on 8.2.2014]. Available at:
http://www.isa.org

ISA-95.com. Technology ISA-95. [WWW]. [Accessed on 5.2.2014]. Available at:
http://www.isa-95.com/subpages/technology/isa-95.php

KnowPap. KnowPap - e-Learning Environment for Papermaking and Automation. [WWW]. [Accessed on 5.5.2014]. Available at:
http://www.knowpap.com

Lautala, J. ISA-95 -standardin ja siihen liittyvien suunnittelumallien hyödyt valmistuksenohjausjärjestelmävalmistajalle. Master of Science Thesis. Tampere 2010. Tampere University of Technology, Faculty of Automation, Mechanical and Materials Engineering. 82 p.

Moniruzzaman, A. B. M., Hossain S. A., 2013. NoSQL Database: New Era of Databases for Big data Analytics - Classification, Characteristics and Comparison. International Journal of Database Theory & Application August 2013, Vol. 6 Issue 4, p. 1-13.

OPC Foundation. The Interoperability Standard for Industrial Automation. [WWW]. [Accessed on 24.4.2014]. Available at:
http://opcfoundation.org

OPC Classic. [WWW]. [Accessed on 24.4.2014]. Available at:
https://opcfoundation.org/about/opc-technologies/opc-classic

OPC Unified Architecture. [WWW]. [Accessed on 24.4.2014]. Available at:
https://opcfoundation.org/about/opc-technologies/opc-ua

OSI PI Blog. [WWW]. [Accessed on 1.5.2014]. Available at:
http://osipi.wordpress.com/tag/compression-ratio-of-a-process-historian

Mahnke, W., Leitner, S-H., Damm, M. 2009. OPC Unified Architecture. Springer-Verlag Berlin Heidelburg. 339 p.

Mazhelis, O., Warma, H., Leminen, S., Ahokangas, P., Pussinen, P., Rajahonka, M., Siuruainen, R., Okkonen, H., Shveykovskiy, A., Myllykoski, J. 2013. Internet-of-Things Market, Value Networks, and Business Models: State of the Art Report. Computer Science and Information systems Reports. University of Jyväskylä, Department of Computer Science and Information Systems. 95 p.

Riesco B., N., Navon C, J., 2006, in IFIP International Federation for Information Processing, Volume 214, The Past and Future of Information Systems: 1976-2006 and Beyond, eds. Avison, D., Elliot, S., Krogstie, J., Pries-Heje, J., (Boston: Springer), pp. 107-118.

SFS-EN 81346-1. Industrial systems, installations and equipment and industrial products. Structuring principles and reference designations. Part 1: Basic rules. Helsinki 2010. Suomen Standardisoimisliitto SFS. 1+157 p.

Tekes. 2010. Automaatio liiketoimintaprosessien tukena. Tekesin katsaus 271/2010. Helsinki. 113 p.

Virta, J. 2010. Application integration for production operations management using OPC Unified Architecture. Master's Thesis. Espoo. Aalto University, School of Science and Technology. 59 p.

Williams, T. 1998. Enterprise integration in the process industries. Presented at the World Batch Forum 1998.