Vaasan yliopisto
UNIVERSITY OF VAASA

Daniel Hummel

Multi-domain maturity model for AI and analytic capability in power generation sector

A case study of ABB PAEN Oy

School of Technology and Innovations

Master's thesis

Industrial Digitalisation

Vaasa 2021

VAASAN YLIOPISTO
School of Technology and Innovation
Author: Daniel Hummel
Thesis title: Multi-domain maturity model for AI and analytic capability in power generation sector : A case study of ABB PAEN Oy
Degree: Industrial Digitalisation
Supervisor: Professor Mohammed Elmusrati
Evaluator: Professor Timo Mantere
Year of graduation: 2021 **Number of pages:** 109

ABSTRACT:
As more smart devices and smart meters are available on the market, industry actors offer AI and analytic suites and platforms where the data streams can be contextualized and leveraged in pre-made industry specific templates and model, together with self-serving machine learning environments. How can a traditional EPC company, use its domain knowledge in offering these AI and analytic suites. The assumption made is that there is no inherent value in the AI and analytics suite without data. How should this assumption be incorporated in projects executed before the operation phase where data from operation is non-existent.This thesis investigate which elements provide a value proposition in the AI and analytic suite and map this against the domain knowledge of the EPC company. The findings is a novel design in where both operational data is integrated into design for new projects. A survey is also conducted on the data utilization in the power generation sector based on the same elements. The findings is that while the granularity is low, the quality is good, with an overall maturity between managed and proactive data utilization, which indicate that there are few automated data streams, but that the data is available structurally and in a defined way.

Keywords: Digitalisation, Energy, Data processing

Contents

1 Introduction

1.1 Background and motivation

In Vaclav smils book Energy Transitions: History, he lays out the history of energy over its continuously changing transitions, with the energy evolution starting with using only animate energy as a source to produce energy in the form of food and crops, to using machines and tools and eventually harvesting fossil fuels on a global level, while always keeping a positive energy balance allowing for growth and advancements (Smil (2010)). This can be seen also by comparing GDP to Energy production, a graph that portray high correlation (Murphy and Hall (2011)). Another factor to consider while observing this is to note that energy intensity has declined at the same time, which is a measure of how efficiently energy can be converted to GDP (Smil (2010)). Power generation is a sector of the energy industry focused on generating useful power, often heat or electricity. district heating is often physically restricted due to piping and energy storage capability of fluids . Electricity is in turn a better source of energy transportation. (Kirschen and Strbac (2004))

The electricity market consist of electricity generation, transmission, distribution and consumption. During the late 70's there began actions to deregulate the electricity market to allow for more competition. This eventually broke up state owned vertical integrations and allowed for more horizontal restructuring together with the emergence of electricity wholesale markets in the grid. (Mäntysaari (2015))

Focusing to Europe, there has been observed reductions in CO2 already from the 90', which have been partially contributed to energy efficiency and liberalized energy markets, but also highly linked to GDP changes, both positive and negative (EEA (2014)). Energy efficiency improvements in mature technologies is something that is frequent in historical energy transitions (Smil (2010)). The European union have made commitments to further reduce CO2 pollution and to increase the penetration of renewable electricity in the ongoing energy transition (CEER (2021)). There is evidence that the future of the electricity market is going to be extended with more distributed production, more prosumers, and energy storage units (EC (2020)). Subsidies has fuelled the energy transition toward this model and at the same time put pressure on the traditional power plants operating on long term contracts for baseload power (CEER (2021)). These conventional power plants, excluding hydropower and nuclear power, often have a combination of high pollution while also being sensitive to fluctuation in the price of fuel. Even if conventional power plants receive some subsidies, mostly for social welfare, it is still difficult to see the long term profitability of these plants, when comparing to the low operation cost of renewable energy combined with energy storage possibilities (EC (2021)).

This new evolving paradigm (EC (2020)), Have been proposed to consist of sufficient renewable capacity in both generation and consumption while also requiring a energy ecosystem consisting of intelligent devices that coordinate consumption, production and energy storage, so that real time scheduling and energy buffering is conducted on both distribution and transmission levels. (Hussain, Narayanan, Nardelli, and Yang (2020)), (European University Institute. (2017)). This ecosystem could perhaps fall as a subgroup to the umbrella-concept Internet of Things that is often the centrepiece in digital transformation as the vision of fully interconnected devices, sensors and other actuators that operate without human interface but integrated as a part of day-to-day activities. The term Energy of Internet, or Energy Internet is the term used to describe this sub-ecosystem of IoT, that should handle and facilitate the intelligence in distribution and production through energy routers. (Tsoukalas and Gao (2008)).

To facilitate such an ecosystem, the devices have to be available in some type of platform until there are devices that are able to operate independently beyond this (Hussain et al. (2020)). There are a multitude of IoT platforms, allowing the integration of devices, sensors and other networked devices to visualize, control or otherwise utilize the data for some purpose (Alberti et al. (2019)). There are also industrial IoT platforms, as well as AI and analytics suites that offer an environment, which often is more comprehensive and allow for more computation power to execute analytics, artificial intelligence on the available data (Kalabin (2018)). Some data is generated during the engineering phase, some in operation. Inherently there is not much value in these platforms without data, a purpose or a problem statement. What these platforms inherently offer instead, is the flexibility, both in business model and technological architecture.

Traditional engineering-to-order, Engineering procurement construction companies (EPC) often exit projects or takes a smaller role according to service agreements as the plant enter into its operation and maintenance phase. This is where the IoT-, or AI and analytics suite platforms should enter into operation as this is eventually the point at which data is analysed and utilized. There have been studies into specific asset performance management applications (Lappi (2019)), machine learning solutions for maintenance (Kalabin (2018)), specific digitalized asset management (Kare (2019)), production scheduling automation Tan, Hou, and Zhang (2016) and other problem driven solutions. However, the objective would be to decouple from specific problems and have the ”problem-solver” platform inherently included in the design close to domain knowledge.

Many actors in the power generation industry are currently offering digital value adding services with descriptions echoing flexible business models such as subscription based and outcome based to the extent of offering anything as a service? Assuming that the intent here is to offer the customer a service and thus build recurrent revenue streams through more service-provider oriented approaches it also assume that the customer understand the potential or their problem statement. These platforms are marketed as offering predictive, self-learning, and other functionalities, but with little specifications on requirements to use these functionalities.(Gartner (n.d.)) The industrial actors add the term industrial, to position the platform closer to industry applications, perhaps by preconfiguring industrial connectors or domain-knowledge into the platform (*ABB Ability™ Genix Industrial Analytics and AI Suite* (n.d.); *EcoStruxure: IoT – Internet of Things* (n.d.); *Predix Platform — Industrial IoT Application Platform — GE Digital* (n.d.); *Siemens — MindSphere* (n.d.)).

This thesis tries to build the conceptual framework for how a traditional EPC should position itself while offering built-in digital platform with energy router access point solutions without specific inherent problem statement. Traditional power plant control system and automation has been designed for isolated environments with high focus on deterministic behaviour, reliability and robustness, while not necessarily considering factors relevant for complex external energy scheduling commands, which would require knowledge on system dynamics, system performance and asset condition. This conceptual model is made by determining a multi-domain maturity model that can relay the capability for analytics and artificial intelligence needed to reach a state for automatic scheduling. The domains that comprise the model consist of the power pool, the power units, the process and the assets and their respective information contents leveraged through networking and communication applied in industry 4.0 concepts. This maturity model is used to create the framework for how a EPC company can find value proposition within current domain knowledge.

1.2 Research Problem and research question

The research problem can be comprised into the following, there are a multitude of AI and analytic suites today that offer the interaction and representation of digital twins. The issue is understanding how to construct a value proposition from these platforms given the assumption that the platform hold no inherent value without data. It is assumed that an EPC company have three distinctive channels to the industry:

- Existing projects
- New projects
- Spare parts

These three channels have a different set of information and data available and it is difficult to determine how and what can be implemented in an AI and analytics platform. Constructing a general maturity model could help the different channels to include digital value proposition during the sales process. It is also assumed that the overall hierarchy of a customer acting in the power generation industry to have similar overall data requirements, such as resource planning, maintenance and operations planning tools, as well as a physical process and physical assets.

The research question can be constructed as follows:

> ***RQ1:*** *What are the domains needed for a value proposition with an AI and analytics suite?*

> ***RQ2:*** *- Can these domains be used to create a maturity model to assess data utilization in the power generation sector?*

1.3 Research objectives

The research question are broken into multiple research objects presented below.
RO1 - The Conceptual framework - Definition of the dimensions, domains, their sub-

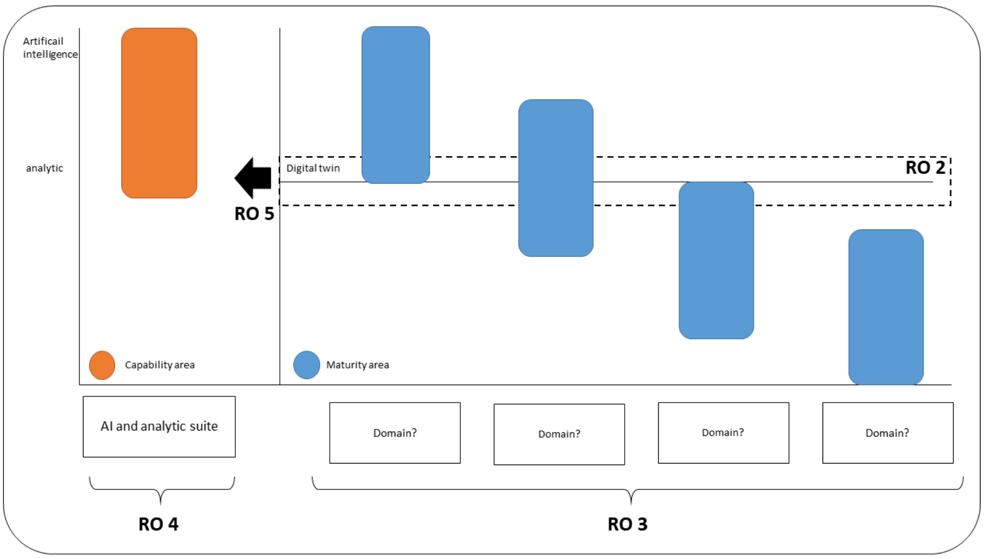


Figure 1. Research objectives.

domains relative to the thesis scope as well as their individual levels of maturity. **RO2**

- Determine the definition of a digital twin such that a reference level can be obtained based on the data requirements. **RO3** - Determine the maturity levels of the domains relative to the digital twin reference **RO4** - Determine the capability levels of analytics and artificial intelligence. **RO5** - Create a maturity model out of the domains mapped onto the capabilities of analytic and artificial intelligence. **R06** - Validation of maturity model

1.4 Frame, scope and methodology

This is done from a grounded theory approach by trying to describe the abstract relation of how these concepts are related through inductive methods of literature review, empirical knowledge and a survey, but with and emphasis on that all decisions must be based on assumptions and explanation to remove risk for personal insight. Further limitation come from the broad topic and is need to be limited to be able to properly answer the research questions.

The survey is purposely directed to actors in the power generation sector through a case study later presented. To be able to assess the operations in the power generation sector some limitations are made on the domains included in the maturity assessment. The Enterprise architecture will be based on the ISA-95 architecture which has a focus on the manufacturing hierarchy and involve enterprise functions related to the operation of na physical process. The second limitation is regarding the convergence of information technology (IT) and operational technology (OT), where the level of IT maturity inside the enterprise is not considered in the thesis. Only OT is considered and the infrastructure and cyber security addressed is limited accordingly.

These limitations create a boundaries which allow for the result to be focused closer to the research questions. Another mentionable factor is the fact that this is not a maturity model for self-assessment or to audit the inner functions of an enterprise or power plant. IT will be used to determine the capability of existent information to be leveraged in an AI and analytic suite and does therefore not consider sub-domains that relate to the governance, management, training of personnel and so on.

As there are also multiple maturity models assessed separately over different domains, literature that compare previous maturity models are used heavily since the objectives are not to evaluate the best maturity model for each domain but find a validated model for the application in this thesis.

Characteristics of the energy industry are not presented in detail, neither is the operation of a power generation plant, but the general concept and the factors around are presented to provide a foundation relevant for the objective. The survey was conducted in autumn 2021.

2 Literature review

This section will address the literature related to the topic such that it falls within the frame laid out in previous section. It should present the concept of the different topics as well as their characteristics and applications.

2.1 Energy industry and power generation

As history has progressed we have become more and more dependent on energy, especially in developed countries. At some point in time the only work done was animate, in other words, people and animals did work, and the fuel for this was the food, which often was the product of their labour, otherwise, they would not have survived. The first law of thermodynamics state that energy can be transformed from one form to another, but can neither be created or destroyed. (Smil (n.d.)) The scientific unit of energy is *joule*. one joule, is equal to a object with a mass of 1 kg with an acceleration of 1 metre per second over the distance of one meter, and is defined as work (Nm). If this is considered over a duration of time Nm/s, it will represent the energy flow, denoted as power using the SI unit watt (W).(Katsaprakakis (2020))

When only animate power was used for farming, hey had to consider that the work done to harvest the crops had to be lesser than that of the energy in the crops produced to allow for growth. If the crops had a high energy density, less crops would be needed to feed the same population, but if the power density of the area was poor, they would have been strained by the availability of energy per area, such that they would have had to walk long distances to find wood for heating the food. Food can be converted as follows, one calorie is the amount of heat needed to raise the temperature of $1cm^3$ of water by $1 \circ C$ and equals 4,2 joules. This also advocate for the use of high energy density crops, otherwise the consumption of a large quantity of food would decrease the efficiency in the energy conversion, since more energy itself would be required by the internal metabolism and heat, which would cause the energy return to become less. This process would describe the transformation of chemical energy to kinetic energy.(Smil (n.d.)).

Energy use and applications has evolved during history as different types of tools were invented to leverage and increase power density, better energy sources were found, or created, to increase energy density. Energy efficiency through most energy conversions has been observed to increase as the lifetime of technologies evolve, which together with the earlier mentioned, has contributed to a higher energy intensity, called the cost of energy. The demand for energy has steadily increased from the burning furnaces for tool making using only wood and phytomasses, to the exponential growth of energy demand following the invention of electricity that provided a form of transfer of energy and allowed for electricity production close to the source of primary fuel and manufacturing close to resources or consumers. (Smil (n.d.))

The potential for electricity was almost immediately understood. It began with ”the battle of the systems”, the race between DC and AC technology and was followed with Franklin Roosevelt's ”New deal” and Lenin's quest for a communist state comprised into the slogan ”Communism equals the Soviet power plus electrification” and this electrification was mostly done through hydroelectric projects, as it also was for the US during that time (Smil (n.d.)). Electricity can be defined as secondary energy, in which it is produced using a primary fuel that generate kinetic power to rotate generators that produce electricity. Primary electricity is production without any fuel combustion, such as for example a wind turbine or hydroelectric power.(Smil (2010)).

Energy transitions in history describe how the primary sources of energy have been utilized in and together with different energy transformations. This usage is often determined by a number of factors such as the availability of the resource and its geographical density, social and environmental impact, How efficiently it can be converted, stored and be made practical. Energy sources usually fall into either of two categories, Renewable, or non-renewable. Renewable energy is referring to a form of naturally occurring energy source that is practically inexhaustible. Widely used forms of renewable energy include solar radiation, wind energy, hydrodynamic energy, geothermal energy. Biomass is also a

Energy Form		Role	1	2	3	4	5	6	7	8	9	10
1	Nuclear - captured energy in the heavy nuclei, released with their fusion and the creation of new lighter nuclei.	NS								a		
2	Chemical - energy released with a chemical reaction, mainly from the combustion of hydrocarbons.	NS, SF, TF							b	c		
3	Electromagnetic - magnetic energy: potential energy of electrically charged or magnetized objects inside an electrostatic or magnetic field, respectively.	NS, FF		d					e	f		
4	Gravitational - dynamic energy derived from the interaction forces between material objects, due to the magnetic field of earth.	SF, NS									g	
5	Elastic - energy of a stressed spring or a compressed gas in constant temperature, corresponding to energy storage. This form of energy can be the result of mechanical, electromagnetic, thermal, or gravitational interaction	SF									h	
6	Electrostatic—magnetic - energy transferred with electromagnetic radiation.	SF										
7	Electrical - energy of the tactical movement of electrons inside conductors.	TF		i	j			k			l	m
8	Thermal - amount of heat that can be released and transformed to other forms of energy, leading to temperature differences.	NS, SF, FF									n	o
9	Mechanical - energy derived from the free motion of a particle or an object.	NS, FF					p		q			
10	Heat - energy of the chaotic movement and the interaction of particles in macrosystems.	FF										

NS: Natural source **SF**: Storage form
TF: Transportation form **FF**: Final form

a. Nuclear fission and fusion
b. Electrochemical batteries
c. Combustion
d. Photosynthesis
e. Photovoltaics
f. Solar collectors, microwave ovens
g. Hydro power plants
h. Compressors
i. Electrochemical storage
j. Microwave ovens, medical applications
k. Capacitors, electrical magnets
l. Electrical motors
m. Electrical devices for heat production
n. Piston engines and turbines
o. Cogeneration
p. Compressors
q. Induction generators

Figure 2. Widely executed energy transformations (Katsaprakakis (2020).
)

widely used energy source, that is classified as renewable. The renewable generally have less energy density and power density compared to non-renewable energy sources. The non-renewable energy sources include coal, oil natural gas, nuclear fuels and have been the majority energy source, and currently still is. The power density can be measured in solid, liquid and gas fuels and is expressed as the energy content per unit of mass or volume. This expression is defined as the heat capacity which represent the thermal energy release in the combustion of that fuel under normal conditions, which is defined as 1 atm pressure and at a temperature of 274 K. Renewable cannot be represented in the same manner as the energy source is a variable in the potential energy of wind and solar radiation. There are methods to measure, calculate and assume the total available potential energy on earth as a whole or at a specific geographical location. (Katsaprakakis (2020))

These energy sources have to be transformed to a energy form where ti can produce power and do work. This transformation is bound to have losses and and can be illustrated using the transformation efficiency. How much of the energy that can be transformed from the input energy form to the expected form of energy. The rest is left unchanged or transformed into other energy forms, such as heat. A system of energy conversion machines are as mentioned earlier bound by the maximum transformation efficiency. The efficiency is often greatest at the point of nominal energy transformation, which is also represented at an instant. The energy converted over an energy process does not provide any information on capacity of the transformation. The following equation can thus be used, while it also should be noted that the instant efficiency η_P is dependable on if additional energy transformers are coupled in series or in parallel. (Katsaprakakis (2020))

$$\eta_P = \frac{P_{out}}{P_{in}}$$

2.1.1 The electricity grid

The electricity grid is the infrastructure that facilitate the energy transportation between production and consumption. It consist of the electricity production, transmission, distribution and consumption.

For most of the 20th century, the electricity market model was subject to a fixed amount of utilities that operated the grid, usually state-owned and highly regulated. These had different levels of vertical integration which means that they also would operate in generation, transmission and distribution. This created geographical monopoly of electricity in most areas and the consumers seldom had much flexibility in choosing distributor.((Kirschen & Strbac, 2004, p. 1)) Uncoupling the electricity from the monopoly regulation and government policy toward a market discipline were thought to lower prices and benefit the consumers. This deregulation began in the West in the beginning of 1980's followed by deregulation efforts in gas, transportation and airlines. Deregulation in Europe was done though a number of directives, beginning in 1993. This unbundling first allowed for independent power plants to enter the electricity market, followed by horizontal restructuring in both generation and distribution by creating a purchasing agent between these entities. (Mäntysaari (2015))

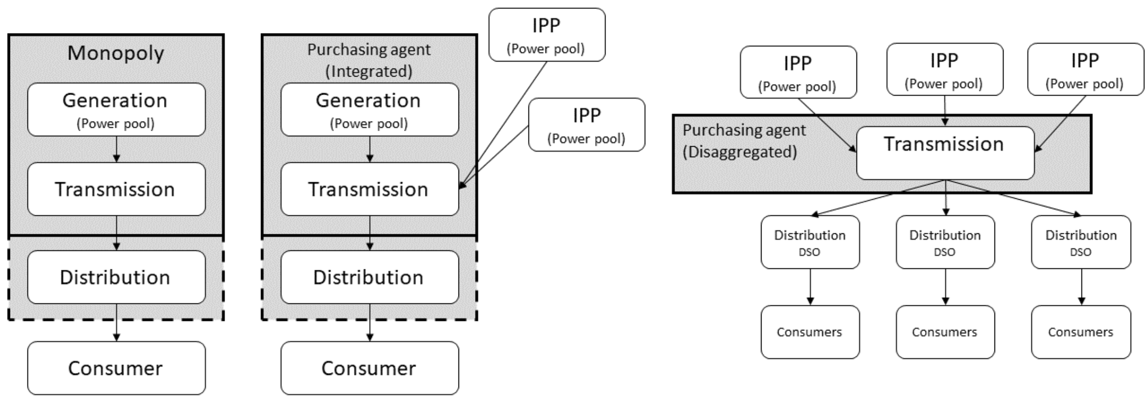


Figure 3. Electricity grid during different phases of deregulation.

The current models of competition in the electricity market are moving from a supplier-centric model, portrayed hierarchically with the producers at the top, transmission in the middle and distribution at the bottom, toward a decentralized consumer-centric model, portrayed as a star topology around a electricity market place.(Mäntysaari (2015)) This decentralized model is derived through the uncoupling of vertical integration by horizontal restructuring enabled by deregulation. The actors who operate in the electricity market is the transmission system operator (TSO), Nominated Electricity Market Operator (NEMA), National Regulatory Agency (NRA) and the distribution system operator (DSO). The TSO is responsible fo maintaining the power balance in the grid, NEMA operate and maintain the wholesale market and the NRA uphold and promote the regulatory framework. The production of electricity is done by the generating companies or Independent Power Producers (IPP) who own and or operate power pools or power units for electricity generation. The DSO, supplier or a retailer participate as a broker between the wholesale electricity market and the end-consumers, where consumer can be of varying size. It should also be noted that large consumers can bypass the retailer and participate directly, or through a separate market broker in the wholesale electricity market. ((Kirschen & Strbac, 2004, p. 3))

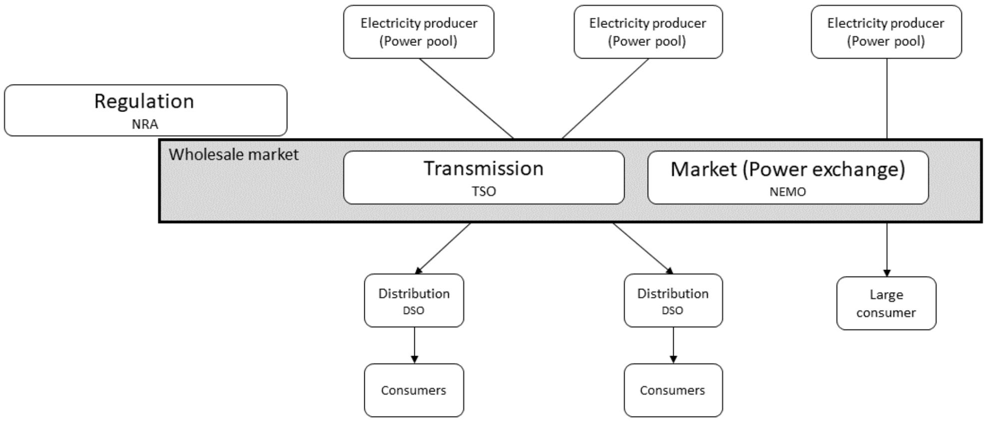


Figure 4. Wholesale electricity market.

There are two main views of this structure, the financial, which uphold the contractual structure in electricity market. This is mainly done through two different channels, either in the wholesale market or through bilateral contracts, either individually negotiated or structured contracts through financial instruments such as futures. The Physical operation differ from the financial such that the TSO is responsible for upholding the integrity of the grid and the contracts for electricity is communicated to the TSO so that the necessary balancing of the grid is done. The TSO uphold contracts with both the generation and consumption side according to access right, capacity, balancing and settlement agreements. (Mäntysaari (2015))

On the consumption side, as the demand of the consumers vary, the production units need to compensate for the change in demand by adjusting its generation units. It is important that both frequency and voltage is kept within nominal levels for the stability and safety of the grid. The voltage is usually most related to reactive power demand and frequency by the active power demand. These factors can be adjusted as long as there is potential energy available at the production plant. Energy storage is future demand-side solution that increase flexibility.

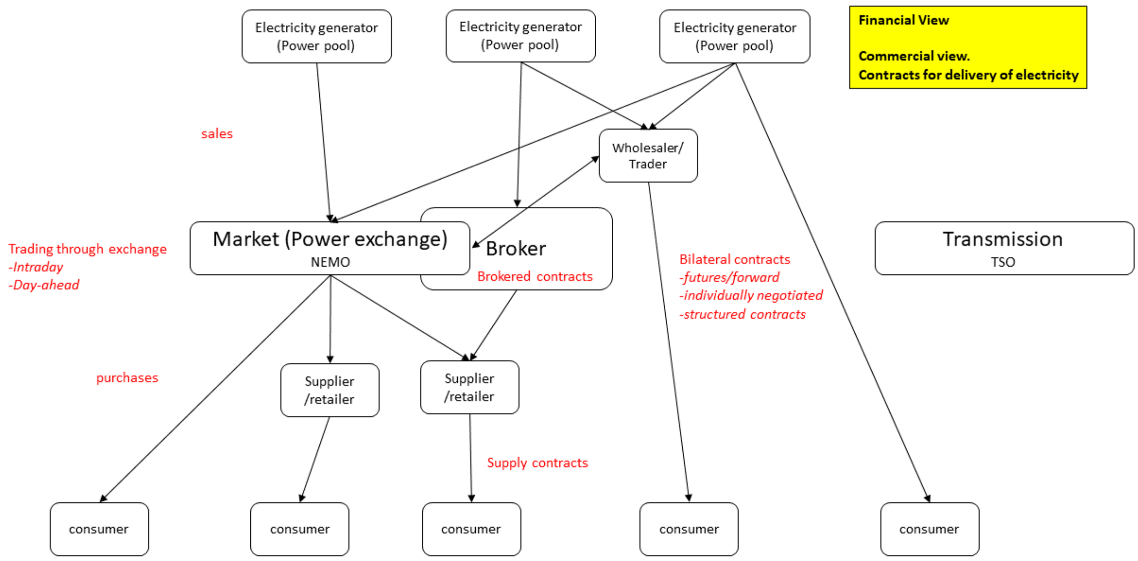


Figure 5. Electricity market from a financial view.

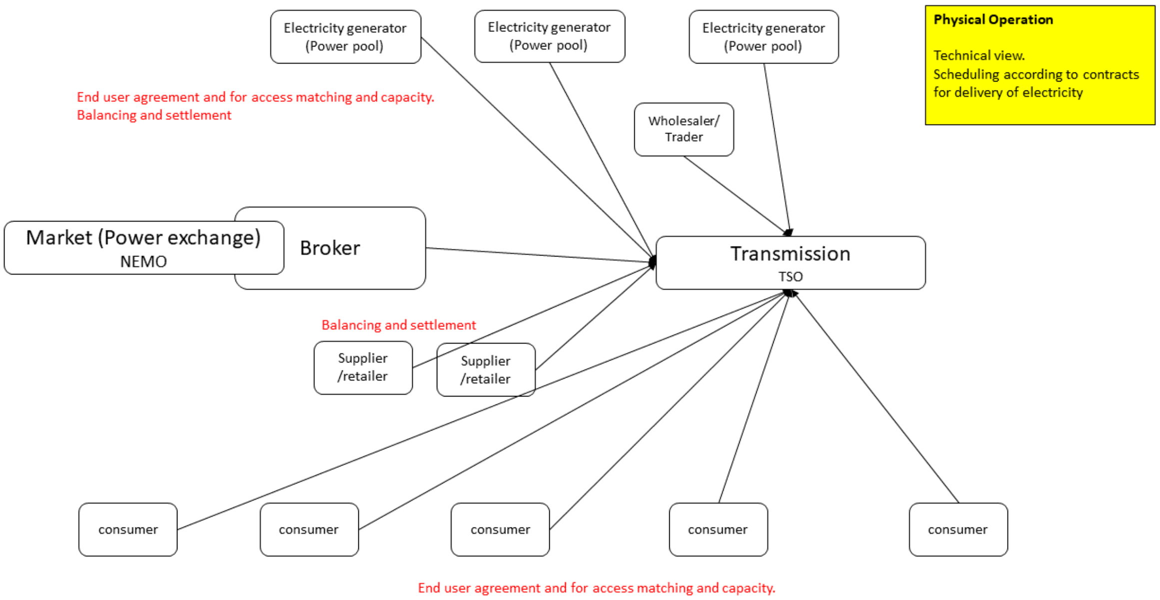


Figure 6. Electricity market from a physical operation view.

The production units can be categorized as guarantee and non-guarantee. Guaranteed power generation means that there is always, as per schedule, potential energy available at the plant and it can be operated at a moments instance to compensate for changes in the demand.

Non-guarantee power generation is where potential energy is available at stochastic periods and require penetration to the grid when this happens. It can be broadly assumed that non-renewable energy is more guaranteed, and that renewable is non-guaranteed. There is also a need for spinning reserve, to help manage the fluctuation in the demand, but also such that it can cover scenarios where demand or production drastically change in a short and unpredictable time frame. There can also be backup generators in the spinning reserve that can be started and loaded within short time frames if needed.

Many of the conventional power plants are of type guarantee power generation, as the fuel type is more suited for scheduled and deterministic operation. Wind and Solar power have become semi-predictable as it has been studied that it is possible to predict production ahead of time to some extent. However, it would also be necessary to have more storage opportunities due to the fact of this unpredictability. (Kirschen and Strbac (2004))

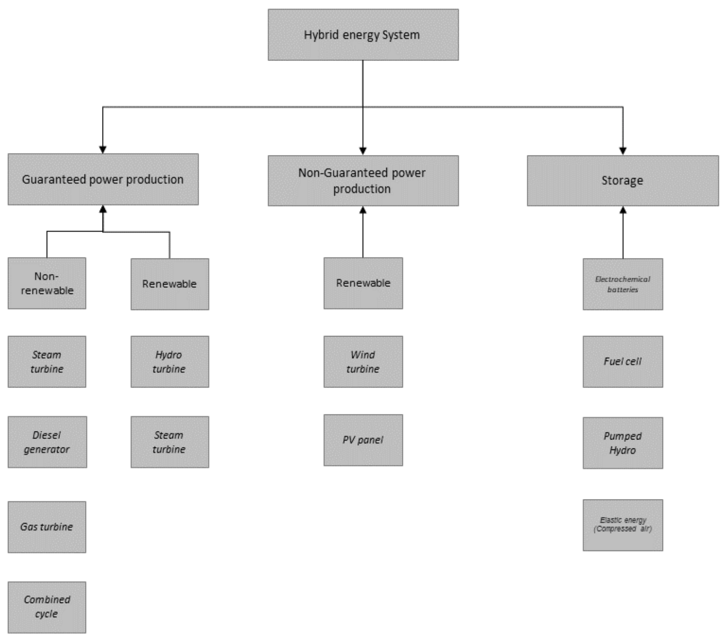


Figure 7. Energy systems.

Looking closer at the production units in a power generation plant, similarities can be noted in the structure, for example most of the generation is focused on utilizing a potential energy by transforming it into electricity so it can be transported and consumed where needed. This most often a involve a main generation unit, from where the electricity then is generated. Then there are also multiple other systems of the transformation around this main unit, supporting it to hold its maximal transformation efficiency in term of cooling, heating, pressurizing, de-pressurizing, priming and lubrication. The power plant can be a combination of multiple generation units, such as thermal generators (Steam turbine, gas turbine, combined cycles and diesel generators), hydro turbines , wind turbines and solar panels. These often operate with an efficiency of between 30-50% with solar power and hydro power as outlier, solar below average at 15-20% and hydro turbines at 80-90%**cite**. This of course depend on size, configuration and the condition of the assets as well as the potential energy in the source. (Kirschen and Strbac (2004))

The combination of power plants in an electrical grid work mainly as base unit, operating with a baseload on each unit while holding a sufficient spinning reserve. Peak units are used to handle peak demands such that the base units can operate more balanced. Non-guaranteed generation units operate with a max penetration percentage of current demand. This is to ensure the stability and security of the grid. The future smart grid goes beyond the conventional setup of large centralized power plants to more distributed generation units closer to the consumer saving on extensive transmission networks. The roles become reversed as the total capacity and capability of renewable power can maintain the baseload in the electricity grid with conventional high efficiency non-renewable generation units handle demand-changes in the baseload together with advanced energy storage. Electrification and Power-to-X are factors that allow the electricity market to become more carbon-neutral due to the electrification of fuel creation from surplus renewable power generation, electrification of the transportation industry, the steel industry and other, combined with more demand tracking, forecasting toward a smaller carbon footprint. Difficult industries such as mining and cement making, can offset carbon emissions by other means until electrification is adopted. Household automation and distribution grid will require smart metering and smart grid solutions to determine patterns in both production and consumption. (EC (2020)

2.1.2 Revenue and cost

Generation facilities usually operate in a certain business model and there are some options in the electricity market, due to the combination of wholesale electricity market, forward market and the bilateral contracts and so forth.in most long term contract plans, the contracts are prepared to mitigate risk for both the seller and buyer. The contract shape and its conditions are out of the scope of this thesis, but the price structure can be as follows. fixed price is in the simplest form there is a single price per megawatt hour produced (MWh), but there is also an option to specify price according to operation stage or at different capacity levels, or even based on fuel cost or thermal efficiency and are eventually agreed on and set in the long term contracts. Variable price is as the name suggests, adjusted based on some external factors, such as a reference price or agreed mechanism. Reference price can be linked to the futures or spot prices in the electricity market, or to actual efficiency, fuel price, commodity or index price. The wholesale market price is determined by the bid and ask profile of all participators from which the NEMA calculate the most ”efficient” price for each hour in the next 24 hours. (Mäntysaari (2015))

Generation facilities can be derived from the profitability calculation and investment decision. The levelised cost of electricity can be used to make investment decision as it portray the average net present cost of electricity generation for a generating facilities over its lifetime. This present the cost as capital cost during project duration and construction and both fixed and variable cost during operation. The fuel cost is calculated separately and is subject to efficiency. Fixed costs is administration and general expenses, routine labor, materials and contract services, and variable costs is fluids, chemicals and other raw materials or products consumed during operation.(Mäntysaari (2015))

2.1.3 The power plant and its assets

Looking at the power plant, we can determine that it is a combination of energy transformations in both series and in parallel. We can also determine that the combined efficiency of the transformations equal the total efficiency of the power plant. It would therefore be optimal to have the maximal efficiency in the power plant, for both economical and sustainable reasons. Another view of the power plant is to see the transformations as assets or a combination of assets. An asset could be the diesel engine in a conventional thermal plant, a motor pumping fuel to the engine, or it could be as basic as the piping system. It could include software, network systems and control system and other embedded systems to a manageable level of granularity. (Kirschen and Strbac (2004))

Asset management is described in ISO 55000 as "*'coordinated activity of an organization to realize value from assets'*. The concept is that by managing assets over its life cycle allow for financial-, investment-, risk-, performance-, compliance-, efficiency benefits while increasing the social responsibility and organizational sustainability. The assets can be both tangible and intangible, but management on physical assets might be more applicable in a power plant but intangible assets should not be avoided. These benefits can be actualized if the asset management comply to the requirements from organizational objectives, strategy and policy, process and operations requirements while also continuously measured, monitored and reviewed. (ISO55000 (n.d.))

It can be understood that if an asset is unable to perform at the intended level, due to any factor or reason, it might require capital expenditure to replace, improve or modify the asset ro meet its requirements. This means that the asset need to provide a line-of-sight for management and organizations for all life cycle stages. The life cycle can be described in four major stage, such as The planning stage, Create and acquire stage, operate and maintain stage, and the disposal and replace stage. (ISO55000 (n.d.))

If there is transparency and alignment it is possible to determine the different specifica-

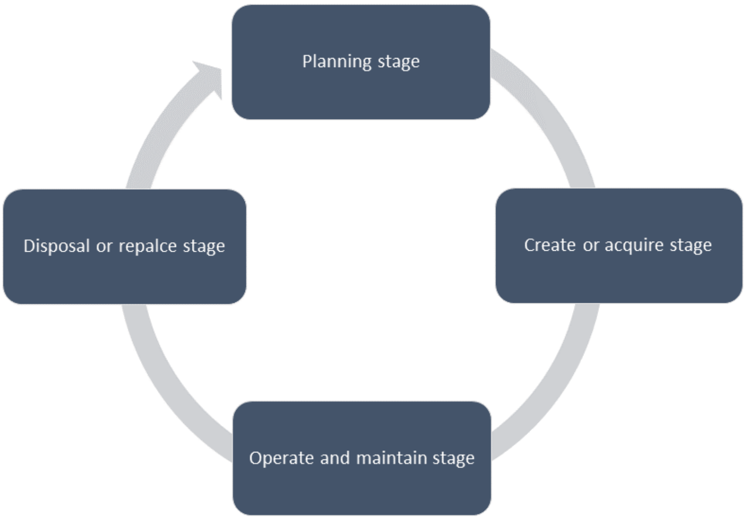


Figure 8. Asset life cycle.

tions and requirements for assets in all the different stages, and allow planning across different stages such that the integrity of the objectives are kept intact. Intra-organizationally it requires that leadership, consultation, communication, competency development and information management activities are mature enough to handle asset management. The measurement and monitoring of assets provide the foundation for these activities to handle both short- and long-term decision- and risk management. (ISO55000 (n.d.))

During the Operate and maintain stage of the life cycle, the asset should operate within its specifications and requirements set during the planning stage. As the asset is applied under different types of load, it become subject to stress. fatigue stress occur when an asset is under dynamic load such as cycles, under which the material is introduced to micro cracks. As the cracks propagate and grow, it will cause the material to fatigue, and finally cause the material to fracture. Creep stress occur when static load is applied to a material over time, under which the material will become strained and start to deform until it finally is unable to perform as intended. Creep stress can be observer over time, while fatigue can be observed over cycles. Thermal shock is another stress derived from the expansion of a section of a material due to thermal gradient that can cause cracking due to the expansion. If the crack is allowed to propagate it will finally cause failure in the structure. Wear is also a type of stress caused by the interaction of two surfaces against another such that it causes a dimensional loss in the material. (Tinga (2013))

During the operate and maintain stage, maintenance is carried out to mitigate and delay the occurrence of the stress on the assets. Different assets perform different operation and are active under different conditions which require the maintenance scheduling to avoid unplanned breakdowns that cause safety risk or and unplanned cost. assets that are not maintained can cause repetitive repairs, higher frequency of breakdowns and also increase the randomness of breakdowns. It can degrade the efficiency of the energy transformation causing lower availability and sub-par performance at higher cost. (Tinga (2013))

The above factors over the life cycle can be described as having a direct influence of the actual availability of the asset. Maintenance concepts should be applied to keep the asset in the optimal availability state for an optimal duration of time. This maintenance concept can be described as having influential factors such as accessibility, theoretical availability and maintenance strategy. Tinga (2013)

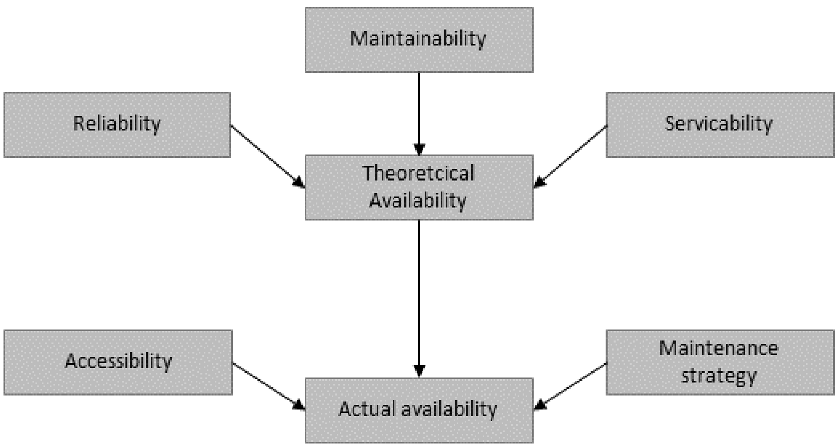


Figure 9. Asset availability (Tinga (2013)).

The accessibility is a factor of the environment in which the asset is installed in and can be difficult to control. The theoretical availability is assumed at the product development of the asset and should reference the whole life cycle. The maintenance strategy should be applied according to the failure mechanism of the asset such that the optimal availability of the the asset is preserved. (Tinga (2013))

The degradation of an asset is determined by the load on the asset compared ot its load-capacity for which it was designed. External loads in the environment, such as mechanical, thermal, chemical, electric and cosmic loads can exist in the environment around the asset and create internal loads in the asset that decrease the load to load-capacity difference and eventually lead to a failure. Typical methods to measure External and in some cases also internal loads are:

Vibration monitoring Measuring the vibrations (1 Hz to 30,000 Hz) associated with an asset to find indication of wear, misalignment, loose part.

Termography Measuring temperatures to find assets operating outside nominal temperature.

Tribology Measuring the integrity of structure of bearing/lube oil/rotor in machinery by analysing lubrication oil

Ultrasonic Similar to Vibration an analysis, often above 30,000 Hz range on compression and expansion of air, gases or liquids

Electric measuring Measuring on electrical components, such as harmonics, current- and voltage levels.

The internal load of the asset can be difficult to measure and can more directly translate into real remaining useful life without the actual degradation model of the asset. The impact is usually realized only when measured or observer manually by dismantling and disassembly of the asset itself. There are many methods to maintain assets availability.

The initial step would be to make process improvements, by modifications, redesign and change orders to improve the reliability of maintenance. An example could be to switch from manual lubrication to automatic lubrication of bearings. This would be one step further from corrective maintenance which is as the name suggest maintenance carried out when the asset has already failed. The need for corrective maintenance actions become lower as more mature methods for preventive maintenance is implemented. This is because preventive maintenance try to pre-act the asset failure such that it can be scheduled to fit into the most efficient window for maintenance with regard to cost and safety. Aggressive maintenance is maintenance scheduled such that the maintenance performance and availability is over 100% to ensure the highest availability. Preventive maintenance is not easy as if the scheduled maintenance windows are too long, the asset will fail before maintenance is carried out, if it instead is too short, unnecessary maintenance is carried out. There are multiple approaches that can be used in combination to achieve a good preventive maintenance. (Tinga (2013))

The manufacturer and designer of an asset of a system can apply model- and, or experience-based predictive maintenance policies. These are often scheduled either over time or by the usage of the asset or system. Dynamic maintenance policies are grouped in passive and proactive groups, where passive methods alert as critical conditions are reached and proactive method try instead to estimate the remaining useful life. Proactive methods can be statistical comparison of historical data against the current data, but require that the system have the identical factors and that these factors are understood. The second proactive method is to use the measured condition to construct a degradation model from where the remaining useful life can be extrapolated. This can be done using either by manually measured conditions, or by an automatic measuring device. The last method is to construct a sophisticated model where the asset or a part of the asset can be simulated until failure. This require that the dynamics of the failure mechanics and the component structure of the asset are known and understood. (Tinga (2013))

Maintenance performance can be measured by looking over the effectiveness and the efficiency. Efficiency involves the health and the availability while efficiency consider costs

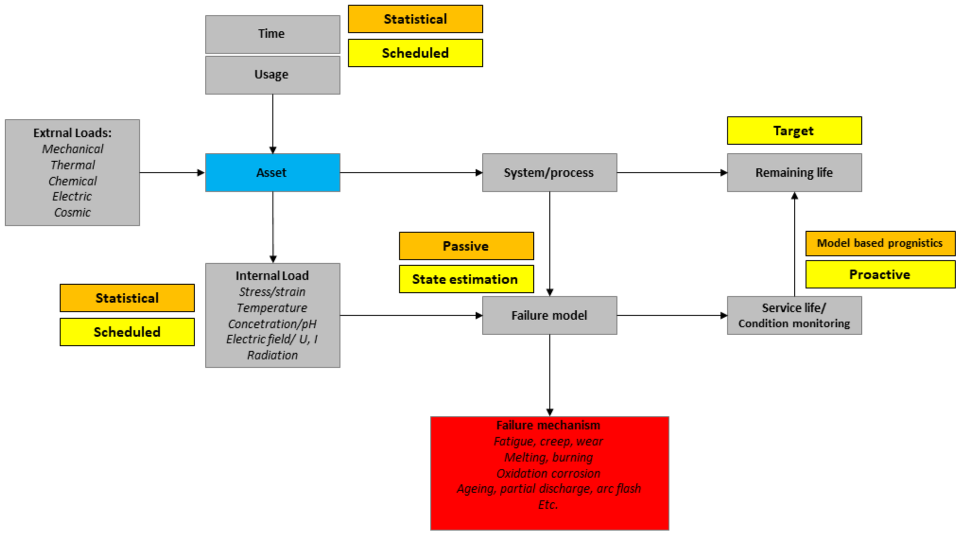


Figure 10. Asset maintenance.

and scheduling. The availability factor describe how well the asset is able to perform during the time it was intended to perform. where the length of the operation $(MBTM)$ periods is divided by the same added with the down time (DT) (preferable the downtime during times when it should have performed) and to what granularity. The availability can also indicate if too much maintenance is carried out, for example if the availability is calculated to show a greater availability than required. The health factor describe how well the asset actually perform the function it was intended to perform. This factor would like to use the Remaining Useful Life (RUL) variable, which is hard to assess, but can be attained objectively through conditional monitoring practice or subjectively though expert opinions. In heory, it is the planned RUL divided by the actual RUL, where the planned RUL is the RUL at the end of the operational period subtracted from the planned service life (SL) divided by number of periods (N). Cost is also evaluated by taking the planned cost divided by the actual cost, where the planned cost is the value of the asset at the end of its life $(C_{residual})$ subtracted from the initial investment divided by number of periods. Cost also has the variable α which can be used to describe the fraction of depreciation planned for maintenance. (Tinga (2013))

Scheduling should consider how much planned maintenance is carried out against the total mainteancnce as well as work orders performed in time over all work orders (n_{wo}) and the number of spares requested by spares available (n_{ri}). The measure of maintenance performance can according to this calculation get a value between 0-150% where 100% is the anticipated performance and a value above or under indicate over or underperformance. This equals the maintenace equation for performance (MP) as ((Tinga, 2013, p.174)):

$$MP = P_{availabilit} * P_{health} * P_{cost} * P_{scheduling}$$

where:

$$P_{availabilit} = \frac{MTBM}{MTBM+DT}$$

$$P_{health} = \frac{\Delta RUL_{planned}}{\Delta RUL_{actual}}, \text{ where } RUL_{planned} = \frac{SL-RUL_{end}}{N}$$

$$P_{cost} = \frac{C_{planned}}{C_{actual}}, \text{ where } C_{planned} = \alpha\frac{C_{investments}-C_{residual}}{N}$$

$$P_{scheduling} = 1/3(\frac{\Delta t_{planned}}{\Delta t_{total}} + \frac{n_{wointime}}{n_{wo}} + \frac{n_{riinstock}}{n_{ri}})$$

2.1.4 Dynamic systems and Control theory

A dynamic system is a system or process where the state is other than static. The change in state of a dynamic is dependent on a single or multiple variables over past and future values. Another characteristic of a dynamic system is that it stores the these past and future values in the system itself. If the system and its components can be identified, it can be mathematically represented such that a model can be constructed.

Today there are rigorous approaches to system modelling, and mathematical modelling of dynamic systems. modelling a system involves gathering the information on the behaviour, a step called system identification, which is understanding the domain in which the system can be represented. This is often not intuitive, as can be seen from famous attempts of system identification, such as the planetary system, where the initial system identification brought a model with the notion that the everything rotated around earth rather than around the sun. This notion was later disproved by Copernicus who modelled the model of the universe which has been the foundation for Newtons law of gravity and many more, namely with the sun at the center. System identification is done based on observations of data from where intuitive patterns are found. As the previous example illustrates, it is crucial to interpret the data in the correct domain for a valid system identification, otherwise the model wont represent the actual system. This comes from system theory that define a system as a black box with inputs and outputs decoupled from the physical construction of the real counterpart. If presented with an unknown system, this would be represented as a black box with input and output. The important aspect is that the right domain is understood such that the correct inputs and outputs are chosen. This data can be gathered to initialize the system identification, and as enough data is available, interpolation can be done over the data to reveal some system dynamics. if the data is sufficient enough an attempt to find a differential equation that represents the system could be done. A correct differential equation would allow it to extrapolate close to real world observations. Initial assumptions have to be reconsidered in case the model fails to follow the real system, for example if the initial inputs and outputs are taken from the wrong domain. Since the model now is represented in a differential equation, it can be called a white box model of the system, since it is transparent in how it describes the dynamics of the system. Some controllers work without a clear white box model, other controller require a white box model, but eventually it is within the domain of control theory to choose the control method that most control the system most efficiently. (Dorf and (2022) Branicky (2005))

2.1.5 Take-aways from chapter

The revenue structure of a power plant is derived from the capacity, efficiency and production output of the energy transformation. The cost structure in operation is derived from fuel cost, variable cost and fixed cost. It should be noted that there are other factors, such as the emission allowance units, subsidies, transmission costs that can occur which have not been explicitly addressed here and can impact either, or both cost and revenue.

2.2 Interface between enterprise (IT) and control systems (OT)

This chapter should describe the interface between physical process and business process with regard to function, activities and flow of information.

2.2.1 Enterprise systems

An enterprise is constructed from different organizations with specific business processes, activities and resources, with the combined effort toward the goal of the enterprise. Enterprise systems can be seen as the information system that integrate the intraorganizational efforts such that it allow for a holistic view of the company through a single information technology architecture. This enterprise system should include all business processes and operations that capture all flows of information and material in and between organizations. In today's global market, enterprise systems also allow for integration with stakeholders and business partners located elsewhere and allow for a more competitive environment by the seamless flow of information. (Xu (2015))

Two enterprises can have different views on the necessity of an enterprise system, or the benefit of information flow. This can cause the enterprise systems to built and integrated in various ways, also depending on the size of enterprise. Some general characteristics to consider is that it need to be modular and account for different disciplines, it has to hold and integrate a variety of information and data types. The adaptation capability of both system and organisation has to be considered, should the enterprise system be adapted to the organization or can the organization adapt to integrate the enterprise system. The constant development of these systems also make them multidisciplinary complex systems, which allow for specificity and higher efficiency, but at the cost of more maintenance and governance. (Xu (2015))

The Enterprise system should provide tools for business planning to business management with aspects to production, inventory or resources. This planning requires information on different time resolution at a high quality and both from internal and external sources, such that correct decisions can be made that meet and capture the customer need and requirements. This also requires the integration and collaboration on supply chain activities to increase the transparency and reduce delays in information exchanges. This also requires that the organizations within the enterprise are able to handle digitized information and are able to produce it themselves. This can often require some changes in their own structure or processes such as automation, lean operations and implementing recognized standards. (Xu (2015))

2.2.2 Enterprise Architecture

This is facilitated through the Enterprise architecture which consist of multiple domains, such as the Information domain, the application domain and the technology domain. Databases-, software- and networking architectures are used to create the information architecture which describe the flow of information within the enterprise. The interfaces between application within the enterprise is located in the application architecture and describe the connectivity of applications and their interfaces. The technology domain present the components of middleware, networking and operating systems architecture in the enterprise system.

Competitive advantages and benefits have been noted in the manufacturing business and has become more popular in the industry. It has been seen that the business has become more efficient and effective and allowed for more data analytics and more precise data activity into specific products and industries. What should facilitate this enterprise system and be the main interface with the environment is the enterprise architecture. There are variety of architectures which suite some organizations better, while some focus only on specific processes and some to specific industries, but essentially it is the conceptual or physical description of how parts of the enterprise system communicate. This enterprise architecture is actualized with eh enterprise integration, which often is carried out through realization of the architecture.(Xu (2015))

2.2.3 Enterprise integration

The integration of enterprise systems based on the architecture is driven, mainly by information and communication technologies. This can be the physical integration of devices, integration of separated applications and databases, to a common environment, or a business integration were different businesses are integrated under the same governance structure. ((Xu, 2015))

2.2.4 Enterprise systems in industry

The industry production and manufacturing process integration into enterprise systems is to allow alignment also for the physical process and its activities to the goals and targets of the enterprise.The ISA-95 standards "*Enterprise-Control System Integration*" describe this integration as a framework for the general interfaces without consideration for a specific domain. This standard describe the hierarchy, based on the "*Purdue enterprise reference architecture*" for how enterprise systems business activities and control systems should communicate in and across identifiable levels, such that Business planning and logistics (level 4), Manufacturing Operations and Control (Level 3) and process control system (Levels, 2, 1 and 0) can be separated based on their relation to the physical process. (*ANSI/ISA-95.00.01-2010 (IEC 62264-1 Mod) Enterprise-Control System Integration - Part 1: Models and Terminology* (n.d.),*ANSI/ISA-95.00.02-2018, Enterprise-Control System Integration - Part 2: Objects and Attributes for Enterprise-Control System Integration* (n.d.),) the Level 4 in the hierarchy can be called the enterprise domain and involve the business processes for planning and logistics. The general activities involve collecting and maintaining information on plant production scheduling and operational management. These activities are often focused on time in intervals in months, weeks or days. Looking at the role-based view, this domain looks at the enterprise level of production and scheduling, through physical, geographical or logical groupings of production sites. The information on production capabilities of various sites is often well-defined at this level(*ANSI/ISA-95.00.01-2010 (IEC 62264-1 Mod) Enterprise-Control System Integration - Part 1: Models and Terminology* (n.d.), *ANSI/ISA-95.00.03-2013 Enterprise-Control System Integration - Part 3: Activity Models of Manufacturing Operations Management* (n.d.),). The activities are defined in ISA-95 these and can be summarized to:

- Purchasing and inventory
- Collecting and maintaining goods, machinery and equipment
- Requirement management
- Production scheduling
- Capacity scheduling

Such that the following information can be determined:

- Production schedule
  - What to produce and when
- Production response
  - What was produced and when
- Production capability

  - How much ca be produced and when

- Production definition

  - What can be produced

The level 3 domain is defined as the manufacturing operations management (MOM), and involves the activities close to the physical process, such as production, quality, maintenance, and also inventory. This domain operates in time intervals of days, shifts hours minutes and seconds and can set in roles of site, area, production unit and unit. The site corresponds to that in domain level 4, but on different time intervals, Area is a location within the site which often corresponds to a group of production units. Production unit is a hierarchical grouping of equipment, such that it bring value to the Level 3 activities. The unit is the lowest for of equipment scheduled by the level 3 activities. The area, production unit and units have well defined capabilities and capacities that can be reported to the level 4 domain. This does not include all activities on a production or manufacturing site, only if the activity is crucial for safety, reliability, efficiency, quality or regulatory compliance it is included (*ANSI/ISA-95.00.01-2010 (IEC 62264-1 Mod) Enterprise-Control System Integration - Part 1: Models and Terminology* (n.d.), *ANSI/ISA-95.00.03-2013 Enterprise-Control System Integration - Part 3: Activity Models of Manufacturing Operations Management* (n.d.)). The activities are defined in ISA-95 these and can be summarized to:

- Production schedule
- Resource management
- Collecting and maintaining goods, machinery and equipment
- Data collection and analysis
- Optimization

- Managing manufacturing, maintenance, quality testing, material storage management

Such that the following information can be determined:

- Production execution
  - Operational commands
  - Operational responses
- Production data collection
  - Process quality, capacity and capability
  - Equipment quality, capacity and capability
  - Resource quality, capacity and capability
- Production definition management
  - Process rules, limits and targets
  - Equipment rules, limits and targets

The levels 2,1 are a subset of the Manufacturing operations and control domain (which include level 0-3), and can be referred to as the control domain, which include all monitoring, actuators and sensory equipment that is directly manipulating or interacting with the physical process. The physical process itself can be seen as Level 0. Level 1 operates in less than seconds, often close to real-time and involve the direct and continuous control, such as PLC systems or controller that control the physical process. Level 2 operates in hours, minutes, seconds and consist of the SCADA, HMI or other operator interface from where the physical process is monitored and controlled.(*ANSI/ISA-95.00.01-2010 (IEC 62264-1 Mod) Enterprise-Control System Integration - Part 1: Models and Terminology* (n.d.))

2.2.5 Take-aways from chapter

The enterprise system is used to align the strategies and goals of the enterprise between organizations both within and outside the company. There has to be known architectures of how domain such as information, application and technology interface between organizations and business processes. The integration is the practical realization of the architecture. there is a difference between a power plant and a pure manufacturing facility, but this thesis argues that the portrayed hierarchy can be exchanged between fields with minor differences.

2.3 Network systems

The communication of digitized information between computers is called a computer network. Today, computer networks allow for a multitude of applications in personal, business and industrial space. They don't only consist of traditional computers, but also other mobile, and smart devices are connected to each other, some are wired and some work wireless. Video, audio, photo or text files can be sent, drag n' dropped and arrive almost instantaneously from sender to receiver. It is the network systems that is the backbone that make this possible. Before the internet was widespread, there were only small network system that allowed computers to communicate, with limited connectivity at low speed. (McCabe and McCabe (2003))

Network systems are what enable the infrastructure of computer networks to exist and what gives functionality and performance to it.(Serpanos and Wolf (2011)) This can be described such that for the network system to work it has to comply to three factors:

- Functional requirements of the network protocol
- Performance requirements fo network traffic
- Implementation constraints of Embedded systems

2.3.1 Network systems

In the beginning of computer networks, there was an occurrence of bit errors, and low throughput which led to the development of more efficient and reliable transmission. It was quite early discovered that a centralized model for communication is not able to provide the best reliability from the single point of failure. More robust networks were developed from the notion of non-centralized communication model which allowed for data to flow in alternate paths to avoid failure. This however, at the cost of letting real-time functionality forego. Network protocols have to be distinguished form network systems such that it is the network protocol that define the communication method without regard for hardware, which also allow for better scalability and standardized reference models. The most known is the layered protocol stack from Open System Interconnection (OSI). McCabe and McCabe (2003)

”A network system is a computational system that executes protocol stacks and switches data among the protocol stacks.” - (Serpanos and Wolf (2011)) Below is an illustration of data sent between two end-systems over a network system.

This should be done without loss of data, loss in integrity of data and without degradation from valid network traffic. These factors can be quantified with consideration for admission control, fault tolerance and out of order statistics to describe the characteristics of the network application. The quantification of these allow for comparison and requirement management as they are composed in more detail. They are easier quantified in the form of:

- Throughput
- Delay
- Jitter
- Packet loss

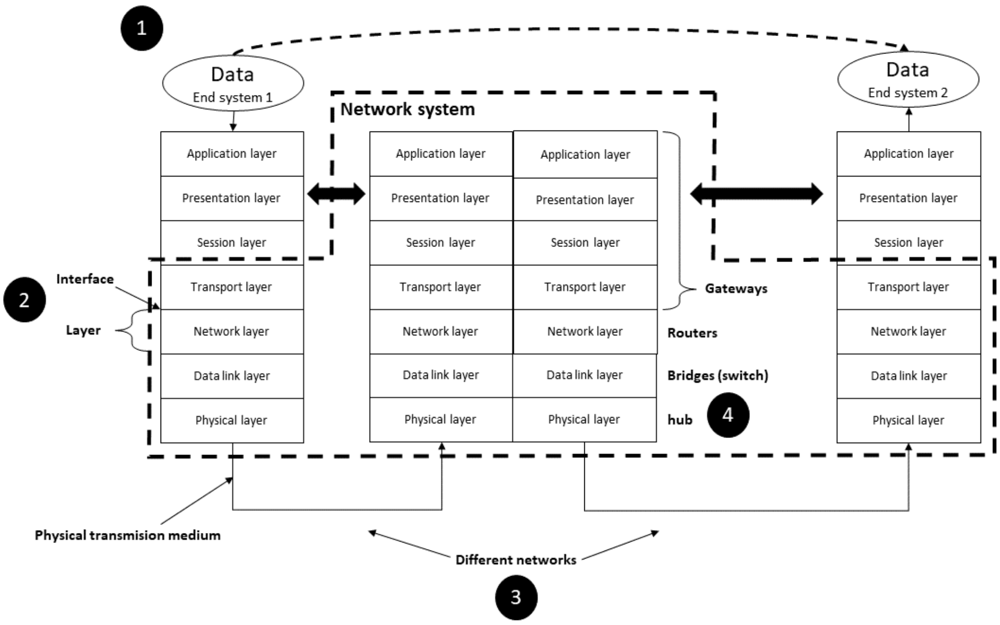


Figure 11. 1: Illustrates that data is transferred from one end system to another. 2: Different layers in the reference model, and inteface between layers. 3: Illustrate end systems on different networks. 4:Typical hardware used at different layers.

Throughput is a broadly used term for information flow measurement. It is used in link, connection and processing to give an understanding of limitation os requirements for speed. Usual units that are used are, bits per second, packets per second or instructions per second, depending on which is most useful for the application. Bandwidth is often used with the same unit, but often refer to the total transmitted bits per second, whereas throughput is a more defined description of the valid flow of information between for example a specific layer. (Serpanos and Wolf (2011))

The flow of information is also delayed by different components. Without considering the application and only the network system, there is time latency delay in transmission-, propagation- , switching- and queuing delay in the delivery of a unit of data within the network or between links. (Serpanos and Wolf (2011))

If there are large variations in the delays, it is referred to as jitter and can occur during any of the four aforementioned processes and if there are other network disturbances such as congestion, transmission errors or system failures, packet loss might occur. (Serpanos and Wolf (2011))

To be able to sufficiently handle the different factors of communication in a network system, there need to be rules and frameworks that explain how it should work. The intitute of electronics engineers (IEEE) association has developed a multitude of standards and the IEEE 802 family is one of them and it covers Local area networks, Personal area networks and Metropolitan area networks. This family of standard defines the two lower layers of the OSI stack, physical and data link layers. The two members 802.3 (Ethernet) and 802.11 (WLAN and Mesh) are active and allow vendors that have products that operate in these layers to know the specifications and requirements to meet to be certified of the standard. This interoperability and standardization allow the use of multiple vendor devices within a network system. (Serpanos and Wolf (2011))

These two members are both creating standards for how the transmission is carried on the two lower layers over two different physical mediums. 802.3 is usually called Ethernet and address guided mediums, such as copper wires and optical fibre connections. 802.11 on the other hand address the unguided medium, such as a wireless spectrum. It also defines functions to handle network traffic at this level to prevent dely, jitter or packet loss. (Serpanos and Wolf (2011))

The network system can also be grouped based on demand such that there is a core system, acting as the highway for all data in the network system. This core system should have enough speed and bandwidth to facilitate all the network services and their protocols without compromising on throughput, delay, jitter or packet loss from changing valid network traffic of the whole system. The core system is often a combination of layer 2 and layer 3 switches, but also gateways and converters. The core system is connected to the distribution system which is capable of connecting and controlling the access to the core, through policy based switching. This is often carried out with specific multilayer switches, which combine VLAN routing and switching in the same hardware, compared to a normal layer 3 router with additional functionality for example ISP/WAN/Internet access. Access layer is where devices interface and is often constructs with layer 2 switching unless specific functionality require more. The core can collapse to distribution layer if there is no need to separate them and they can be combined to provide the bandwidth and network performance for all network services off the access devices. (Serpanos and Wolf (2011))

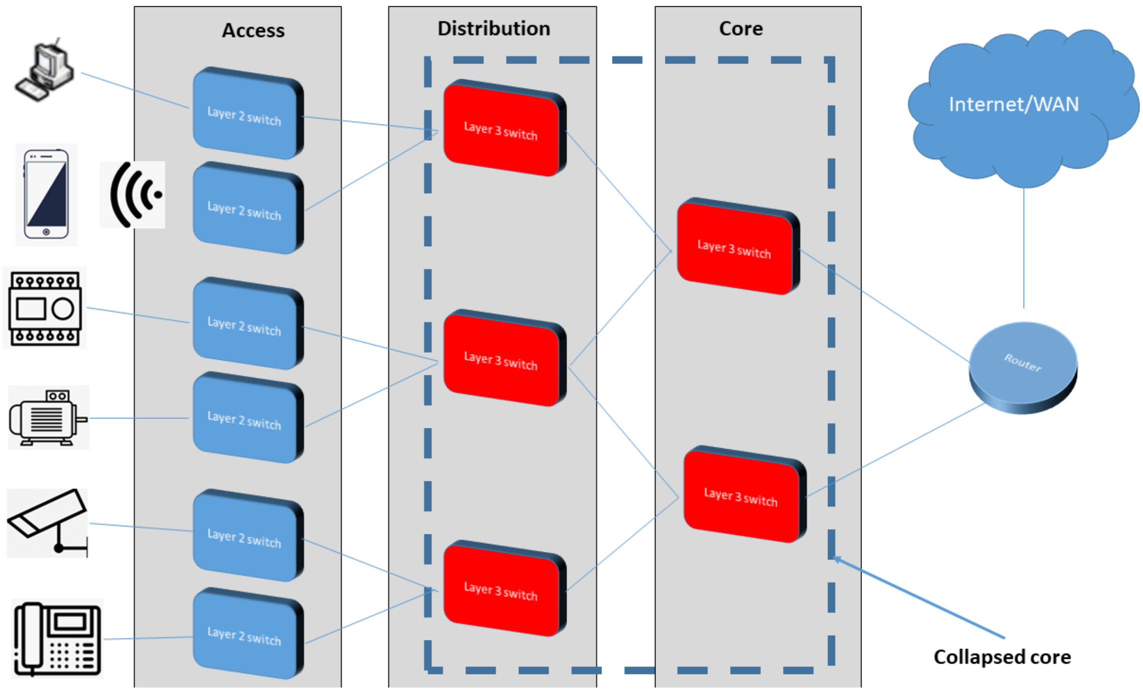


Figure 12. Network system layers.

The number of devices and users at the access layer and the desired access methods determine the requirements for the network system and also chose whether wired or wireless solutions should be used. currently wired communication is typically 1Gbit/s at the access level, and a minimum of 10Gbit/s at the distribution and core layers. The maximum defined in the 802.3 is 10/25 Gbit/s for ethernet, but can be aggregated using fiber optics to 400Gbit/s over shorter distances. It is said in the (Group (n.d.)) presented in 2020 by IEEE on industry connections bandwidth show, with large diversity, that there will be a growth in network traffic of a factor 2.3X to 55.4X from 2017 to 2025. This growth is driven by factors such as the number of users, the methods and rates of access and an increase in overall services, but do also state that 10Gbit (10G-EPON) should be sufficient for industry until 2025.

WLAN, the second member mentioned earlier, also span layers 1 and 2. WLAN is wireless transfer of data and the physical layer is radio frequency communications. The hardware transmitter generate a carrier signal, carrying the ones and zeros of information, in AC or DC format and is radiated through a antenna wave in the format of a electromagnetic wave. This is usually an isotropic radiation from the antenna that can be focused by altering construction or by changing the carrier signal characteristics. This signal operate in a 2.4Ghz or 5Ghz that are licensed for the most part. The licensed frequency are help control who can be active and operate in this area unlike other unlicensed frequency areas where anybody can transmit. ("CWNA Certified Wireless Network Administrator Study Guide" (n.d.))

Unlike the wired communication where the signal is travelling inside a encapsulated wire, wireless behaviour is less anticipated and is influenced by the environment in many ways. To begin with it shares the characteristics of wired signals such that it attenuate as it travels over distance. Other than that, it also influenced by factors that affect how it is propagated, which is how the signal spreads as it is transmitted from the antenna. These factors are absorption (Signal getting absorbed in materia), reflection(signal is reflected from materia), scattering (signal is scattered as it travel through materia), refraction(bent though air due to atmospheric conditions it passes through), Diffraction(bent through air due to atmospheric conditions it passes around). These factors can affect the frequency, amplitude and phase of the signal. This require the specific consideration of the following when designing WLAN network systems: WLAN Coverage Design (what should be covered and how), Received Signal (strength of received signals), Signal-to-Noise Ratio (background noise), Dynamic Rate Switching (switching data rates as moving in coverage area), Transmit Power (Ap optimal transmit power), Roaming Design (AP switching) and so on. ("CWNA Certified Wireless Network Administrator Study Guide" (n.d.))

Wireless communication can be established point-to-point, to bridge or connect two networks, or as an access point, in a point-to-multipoint configuration. 802.11ax (Wifi 6) allow for a maximum data rate of 1147Mbits/s over a 20MHz bandwidth (9608Mbit/s can be achieved if 80+80bandwidth can be used) in the 2,4Ghz, 5GHz or 6GHz frequency bands over an range of 30meter indoor, and 120m outside. ("CWNA Certified Wireless

Network Administrator Study Guide" (n.d.))

Both 802.3 and 802.11 are non-real time due to the Carrier Sense Multiple Access that is carried out as a media access method to avoid or detect collisions in transmission. The functionalities for collision detection (Full duplex) and collision avoidance (half-duplex) are both random access protocols and can therefore not be deterministic or have guaranteed transmission delay.("CWNA Certified Wireless Network Administrator Study Guide" (n.d.))

The header sizes and payload sizes differ for Ethernet and Wlan, with Ethernet having a payload of 1500bytes and WLAN, a payload of 2304 (MSDU). However, there are jumbo frames on Ethernet, outside 802 family, that allow for 9000 byte payload, and Wlan, in 802.11ax, have functionality to aggregate frames and wrap them into a single frame (7935bytes for A-MSDU and 11 kilobytes for A-PSDU). There are also some limitation on payload from the upper layers, for example transport layer in TCP over IP limits payload to 1500 bytes at lower levels. (Serpanos and Wolf (2011))

Cloud Networking

Cloud networking is more in the operation of the network system and what its priorities are. If the network traffic is definite and number of devices and users are known, as well as the access methods. It can be built with specific hardware without scalability. If scalability is necessary, it comes down to how elastic it has to be and how it needs to scale. Here it can be distributed or made such that services are separated and available on demand, but locally managed within the enterprise. Another version is to have and only connect the devices to access points and then outsource all networking capabilities and use it per demand. (Serpanos and Wolf (2011))

5G

Another emerging technology is 5G, which should not be confused with WLAN 5Ghz, which is as described above used at layer 1 and 2 of the OSI reference model for applications such as access points, stations and clients. 5G LTE is on the other hand used for enhanced mobile broadband (eMBB), massive machine type communication (mMTC) and Ultrareliable low-latency communications (URLLC). The standards for 5G are not fully finalized at the moment of this thesis, and there is no ”true” 5G commercially deployed. There will be requirements on backhaul bandwidth in core and distribution to carry all the data as 5G is reported to have peak data rates at 20Gbit/s and 10Gbit/s and should manage one million devices per square kilometre without falling below 100 Mbit/s. (Group (n.d.)) mention however, that there are assessments into a 800BGbit/s and 1,6 Tbit/s Ethernet standards, which could help manage the backhaul in the core for 5G. (Lei et al. (2021))

Enhanced mobile broadband is directed toward human-centric devices that require access to multimedia content, services and data. This multimedia content can include video sharing, streaming, Augmented reality and Virtual reality in high density areas. AN example is 8K UHD video with a frame rate of 50 FPS would require about 300 Mbit/s, compared ot 1080p at 60FPS which only require 6,8 Mbits/s and more extreme 24K 3D video with update rate 200 FPS it would require 3360 Mbit/s at 5ms delay. (Lei et al. (2021))

Machine type communication on the other hand is directed at communicating massive amount of sensing data directly from location to cloud storage for further processing. The sensory devices usually have own or shared processing such that they can work by directly sending all data or sending data on demand or at conditions and the data itself can also be specific or conditional data. Individual sensors might have small packets if only transmitting status of measurements, but can be increased if video feeds and other more demanding data formats are implemented. (Lei et al. (2021))

Ultrareliable and low-latency communication is a service directed toward manufacturing and, production, remote medical surgeries, smart grid, transportation safety and other applications were wireless control is required to be deterministic and reliable. (Lei et al. (2021))

2.3.2 Security

Looking at the fundamentals described earlier we can assume security also can be viewed from the same perspective as network system such that it can be summarized to Network protocols, network traffic.This would assume that there need to be some security in all three dimensions to protect the data passing in the network system. This in handled through confidentiality, integrity and availability. If the data is sensitive it could be necessary to protect it by making the data itself not viewable by unwanted systems or people, confidentiality. This can be done at different levels in the OSI reference model, by for example starting with access control at the application level, but as the data can also travel among unsupervised paths, both wired and unwired, there need to be some protection on the traffic itself also. This can be done by encrypting the data before transmission. Considering wireless communication, the wireless transmission itself need to be considered separately, and has methods for encryptions, such as WEP, WPA and WPA2. If this is done properly, any access on either wireless or wired paths, prevent data to be revealed in plain visibility, but distorted in the encryption. As methods for encryptions evolve, the idea to distort the data such that it causes confusion and diffusion. Confusion is achieved by distorting and changing the data such that there is no resemblance to the original data from the cipher data and that the method for distortion cannot be revealed by manipulating the cipher data. Diffusion is done by permutations such that the original data is ”shuffled” and done in such a way that shuffle method cannot be revealed from manipulating the cipher data. (Dzung, Naedele, Von Hoff, and Crevatin (2005))

Confusion: 01234 would become ABCDE

Diffusion: ABCDE would become ECABD

Another part when thinking of information security is the integrity of data. There are methods to determine the authenticity and integrity of information between transmission and receiving ends by digital signature. If both integrity and authentication of data is achieved, it is called non-repudiation, which means that the sender cannot dispute that the data was sent or the content of the data.

This is done using hash algorithms, and is deployed as a one-way hash onto the data to generate the digest. This creates a specific string of characters regardless of the length or size of the data. This means that by having the same data at two different locations, running the same hash algorithm on the data should output the same string of characters, and if it does not do that, it means that the data has been altered. The steps to achieve non-repudiation uses the hash which is then encrypted with the senders private key to create the digital signature. This signature is then sent alongside the message to the receiver. The receiver applies the message to the same hash functions and decrypts the digital signature using the senders public key. Now the receiver can compare the generated hash with the original hash from the receiver and non-repudiation is achieved if it is a match. (Dzung et al. (2005))

Availability is the final pillar of information security and is to guarantee that authorized users and systems have access to both systems and information at all times. This also increase the fault tolerance in systems and increase the overall deterministic behaviour as the system is expected to operate as intended. This can be achieved by designing the network system to meet these requirements. This is done by correct capacity and capability of network system, and can utilize functions such as redundancy, automatic backups, physical security of networking assets and the training and maintenance of the network systems. (Dzung et al. (2005))

2.3.3 OT infrastructure

Operation technology (OT) is technology that is directly connected to the physical process and consist of the Industrial Control system (ICS) which is a combination of processing resources, communication resources, sensing and actuating devices. The ICS infrastructure can be grouped into operational zones based on characteristics and requirements. This should however not be confused with the information hierarchy of ISA-95. Instead, these operation zones are called Enterprise, Control and Field and allow segmentation based on different requirements and characteristics in network and devices. These operation zones can also be described as hierarchical such that Enterprise resides at the top, control in the middle and Field at the bottom. Enterprise zone consist of mostly IT infrastructure, with different access points, and a variety of most IP based protocols and is often connected to other networks or the Internet and emphasises confidentiality, integrity and availability. (Colbert and Kott (2016)):

The Control zone can be described as control room devices, where such systems as SCADA resides. This zone emphasises less to no confidentiality and integrity and focuses on extreme availability in form of safety and reliability, and have requirements to be time deterministic.

The field zone is the plant or process and include the industrial automation and cyber physical systems (CPS) or Cyber physical devices (CPS). A CPS are often embedded systems such as Programmable Logic Controllers (PLC), Remote terminal Unit (RTU), Intelligent embedded device (IED). The combination of these create the essential part of the ICS, that is monitored and managed by operators from the control zone. (Colbert and Kott (2016))

The devices in the field zone have traditionally not had any applications for confidentiality and integrity and instead focus on extreme availability and strict deterministic execution derived from strict design principles, validation, testing and fault control, which can be further described as:):

- Fault avoidance — avoid faults by design
- Fault removal — reduce, by verification and testing the presence of faults
- Fault tolerance — provide correct function despite presence of faults

This is done in both hardware and software to mitigate all types of faults. Hardware is often more subject to random faults that software, and this is due to the fact that software faults are triggered from input information.(Colbert and Kott (2016))

2.3.4 OPC UA

OPC Unified Architecture (UA) is the successor ot OLE for Process control (OPC) which was established to reduce development work and need to develop driver interfaces by creating a model for the hardware and software interface. The successor offer more scalability, platform independence (by removing DCOM dependency) and higher security. (Tian and Hu (2019))

OPC UA allow for and support object oriented modelling of applications and the specific complex data types and structures. This provide a way for device manufacturers to represent the data in a unified way within the specification of organisations such as IEC, ISA and so forth. This allow for data to be unified between devices and structured in a standardized manner. (Tian and Hu (2019))

Unlike the above mentioned solution for ProfiNET for layer 2 real time applications, OPC UA on time sensitive network (TSN),member 802.1 of the IEEE 802 family, will work above layer 2 and allow for routing capabilities and IP support in client-server, or publish subscribe communication format. If more embedded devices start to implement OPC UA, publish subscribe networks could be more suitable as due to the many-to-many type communication compared to traditional client-server, which is intended for many-to-one. (Tian and Hu (2019))

TSN offer the solution to the problems of difference in periodicity and network demand, interoperability and complexity of bus, by implementing features such as clock synch, traffic scheduling and system configuration so that flow reservation allow data frames to be communicated within a known time window between synced nodes in the network

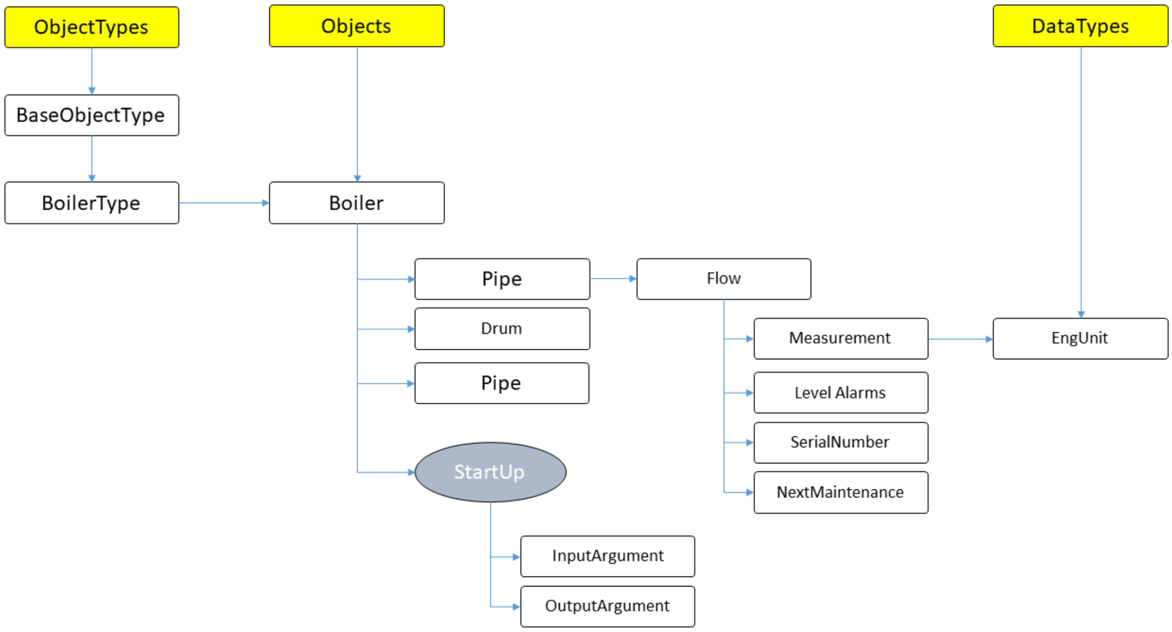


Figure 13. OPC UA example: Here is a example representation of a node model. The boiler is available in the objectType and is here represented in the address space of a device representing the boiler. This data structure can then be browsed by a client..

system, resulting in real-time communication. (Tian and Hu (2019))

2.3.5 Cyber threats

As mentioned earlier the implementation of information security is to guarantee the confidentiality, integrity and availability. Cyber security can be thought of as the way to protect against threats against any of these aspects. Threats against the confidentiality is that a third party tries to access information in a malicious way, this can be done in various ways, ranging from social engineering to password cracking. Passwords, encryption and other methods today are implemented in ways that try to prohibit these methods of malicious threats. These threats are more important in IT than in OT as information in IT is more can be personal or otherwise sensitive such that it can be exploited. The information on the OT network is often only process data, but can also be sensitive production data or other information about the process that can be analysed and give away schedules and or repetitive behaviour that can be further exploited. ((Colbert & Kott, 2016))

Threats on integrity is also more sophisticated in IT, as it can allow implementation of unknown files, application or other altering of files without knowledge to the owner. This could also include sending fake command and requests disguised as someone or something for malicious reasons. For OT it could mean that a malicious threat is able to monitor the network traffic to learn certain patterns in the traffic due to its deterministic behaviour. This could allow the threat to then issue the same commands over the communication line, such as opening breakers in a grid, shutting down processes or starting processes.((Colbert & Kott, 2016))

Availability in IT is having enough maturity in the network system to manage changes to maintain availability. As discussed earlier, packet loss, delay jitter can be consequences for network congestion and a malicious threat can try to increase or manipulate the traffic such that the devices in the network system fail. It can mean shutting down or breaking software and or disabling breaking hardware. In OT, Availability is traditionally the most important aspect and both hardware and software is optimized to have the highest level of safety and reliability. There is also less leeway in control or field network systems as they are often time-critical and the implemented hardware is deterministic becoming a target for malicious threats that would try to disrupt this. prevention methods include minimizing the attach surface. ((Colbert & Kott, 2016))

2.4 Industry 4.0

Industry 4.0 is reordered as the next industrial revolution, as we have come to a point were most new devices are connective and most businesses have digitized their information. Digitized information allow for digitalisation of processes. It is the maturity and availability of big data, cloud computing, IoT-devices and artificial intelligence that will propel the 4th industrial revolution (Landeta Echeberria (2020)).

2.4.1 Digital twin

The notion of Digital Twin was first introduced by NASA as they created a physical copy of a space craft to be able to troubleshoot and test remotely. Today digital twins are referred to by many different name, and the concept is more related to the Product life cycle management, such that the digital twin should be created at the initial step of creation and then be part of the product throughout its life cycle. It is also classified as a bidirectional transfer of information between the physical and digital asset. To be bi-directional will imply that the physical asset is able to receive and hold information and parameters. Otherwise it will be a digital representation of the physical asset, not a digital twin. The more information available for the digital twin, the better the representation will be. Advanced digital twins should also allow for animations and simulations which can be beneficial in for example training and product design.(Armendia, Ghassempouri, Ozturk, and Peysson (2019), Singh et al. (2021), Barricelli, Casiraghi, and Fogli (2019))

2.4.2 Digital transformation

Digital transformation is the actualization of the industry 4.0 concept. It can be assumed that businesses have information available to create digital twins of its assets, and can leverage these digital twins with big data using artificial intelligence. This could perhaps be best illustrated using a basic example from the transportation industry. A number of trucks are transporting material from a common site A to various sites B. As they depart from site A, they are interconnected and are collectively calculating the optimal route for transportation that is most beneficial with drag, weather conditions, vehicle capacity, traffic and other factors. Data collection from cameras, engines and the environment is all the time fed to cloud computing resources that calculate and return the most optimal configuration of engines chassis, to minimize consumption or maximize delivery speed. The destination sites are aware of the routes, locations and the estimated time of deliveries based on environmental conditions and the feedback from the truck itself and is optimizing for its arrival and departure. The energy provider is also participating and is aware of the window for potential energy need as the trucks arrive at different locations. Value is created through utilized information and data.(Siebel (2019) Abolhassan (2017) Saldanha (2019))

2.4.3 Internet of Things (IoT)

Internet of Things seem to be a more holistic term that cyber physical devices (CPS) in literature. Internet of things often refer to any device that securely can send any information to any other device. Overall the term is used to refer to modern devices with more advanced sensors and more complex data. This could be for example a wireless vibration sensor combined with a thermographic sensor sending vibration data and thermographic video in real time to a cloud hosted application. Cyber physical devices often refer to a physical device with embedded computation processing and sensors that is able to actuate on an output by itself. This could for example be a valve that is measuring flow in the proximity of the flow meter but at the same time is also receiving real time data from a nearby temperature sensor and can act accordingly by communicating with near by CPS's over machine-to-machine communication or with a parent system. The driver is network systems and network protocols. (Mandler et al. (2016))

2.4.4 Big data

The big data concept is derived from the exponential growth in users and devices connected over the internet and within local networks. This connectivity and interaction with devices generates data in different formats, both directly as in a photo taken by a phone, but also indirectly as metadata generated in the photo, such as GPS data. This can be illustrated by the growth of data which was measured ot 600 MB in the 1950's and reached 100 000 000 000 MB in 2010 (Balusamy, R, and Gandomi (2021)) and estimated to be about 650 000 000 000 000 000, or 65 Zettabytes in 2020. In the beginning of computing, Structured data was prevalent and the database management systems were constructed accordingly. Today, most data is unstructured or semi-structured. This require data mining to find the useful information within the unstructured data. Data mining should not be confused with big data. Big data refer to data characteristics, such as volume velocity and variety. Data mining is the process of where large datasets are processed to find underlying knowledge or value.(Balusamy et al. (2021))

The big data characteristics, volume, velocity and variety refer to the dimensions of data where volume represent the size of data, usually in above terabyte unit. Velocity refer to the speed of generation, processing and rate of analysis on data. Variety refer to the format of the data, which can be either structure, semi-structured or unstructured and the source of the data can be human-generated or machine generated. Traditionally data was structured and its use case was determined before it was stored. This allowed it to be configured and organized into columns and rows, called records for rows and attributes or fields for columns, in a structured form, thereby the name.(Balusamy et al. (2021))

Structured data is stored in relational database that has great properties for consistency, isolation and allow for more privacy and control. Popular types of relational databases are MySQL, oracle and DB2, and are all able to handle data according to a specific scheme. Because of the construct of relational databases and the fact that it checks if the incoming data support the scheme of the data base, will restrict the velocity. large relational databases, billions of rows, need efficient indexing to be able to handle volume as to not have a database query to do full row-lookup. (Balusamy et al. (2021))

Schemaless databases (NoSQL) are non-relational and store a higher variety of data and does not require the table structures to be determined before it can receive data. This means they can easily be scaled horizontally, something that is restricted in relational databases. a big data version of schemaless database system is Hadoop which allow any data or data to be written, and is after that preprocessed and prepared for analysis.(Balusamy et al. (2021))

Another type of data that is more difficult to handle is data streams. Data streams contains sequential data flowing from one node to another continuously without defined format. Data streams can be unlimited or unbound, making them almost impossible to store permanently in storage due to its compounding volume, the velocity of the flow require high capacity processing, which is also limited by variety as the data stream can have different characteristics such as infinite length, changing concept and evolution. (Last, Kandel, and Bunke (2004))

2.4.5 Cloud computing

Cloud computing is the extreme form of demand-driven computation and networking compared to the traditional home computer. In the beginning most computational power was local and computers were used to run only local applications and services. As the internet became more widely adopted, website hosting became more popular, but visitors was limited by the hardware and software on the local computer. This then evolved to computer storage where individuals could lease or bring physical hardware and maintain it themselves, a form of computer hotel. This then evolved further into server architectures that became more widely available to buy or lease. The problem with having fixed server hardware for a website is that it is still not scalable. The reliability, efficiency can be improved, but not scalability because fixed hardware does not allow for capacity planning. This means that the website could only handle a fixed amount of visitors, and if there was sudden increase, it would be overloaded. If the host of the website had resource, they could anticipate demand and install more resources, or install resources as it reaches its full capacity, or even try to match the demand by slowly increasing resources. This of course required additional cost in personnel, operating cost, security and administration. (Erl, Puttini, and Mahmood (2013))

Cloud computing provide flexibility to this by providing the IT environment with scalable resources, such that the resources will automatically increase as demand increases. With cloud computing the website host would only use resources according to demand, such as 100 resources during peaks, and less during lower demand. An IT resource could be a Physical server, virtual server, software, service, storage device or network device. IT resources can be provided in different formats, such as an on-premise IT resource as in the earlier example where the hardware is located within the premise of the organization. If the on-premise resources are migrated to the cloud, they become cloud based IT resources and vice-versa. Hybrid or redundant deployment means that the IT resource exist both on-premise and on cloud. It should also be noted that an IT resource can exist in the cloud and not be accessible remotely. Scaling of IT resources can be scaled vertically or horizontally, Scaled in or scaled out. Horizontally, refer to adding or removing resources. Vertical scaling is replacement of resources based on capacity, scaling up is replacing for higher capacity and scaling down is replacing for lower capacity. (Erl et al. (2013))

The payment plans are often flexible and short term, providing granularity in configuration and extensive scalability while not having software or applications locked in infrastructure. These characteristics are the enabled by technological innovations such as clustering of IT resources for increased reliability and availability which is further extended by grid computing, allowing pooling of IT resources that are heterogeneous and geographically dispersed. Another innovation is virtualization which allow software to be decouple from the underlying hardware by adding a virtualization layer between. This virtualization layer can be used to host multiple virtual images of a IT resource on a shared physical IT resource.

The provider of the cloud can deliver almost any IT resource or combination of IT resource as a service, the most common solutions are Infrastructure-as-a-Service, Platform-as-a-Service and Software-as-a-Service. more specialized variations could be Storage-as-a-service, communication-as-a-Service or Testing-as-a-Service.

IaaS is offered to cloud consumers that require a high level of control over the cloud-based environment, and is often offered non-configured IT resources such as virtual servers where the hardware requirements, such as processing, memory and networking capacity is defined. The scope of environment offered can also include other raw, non-configured IT resources according to specification.

PaaS can be seen as a level above IaaS, and offer the cloud service as a ready configured environment of IT resources. This is often used by cloud consumers to scale on-premise environments for substitution, elasticity or to offer cloud services themselves. This can be due to offset economical incentives or other factors such as missing technical capability to manage an environment themselves. The PaaS offer less control over the lower levels of IT resources compared to IaaS (Erl et al. (2013))

SaaS is where a IT resource or a combination of IT resources create a final solution offered as a product to cloud consumers. These have very limited control and is often offered as usage based services where the cloud consumer only has access to the front end.(Erl et al. (2013))

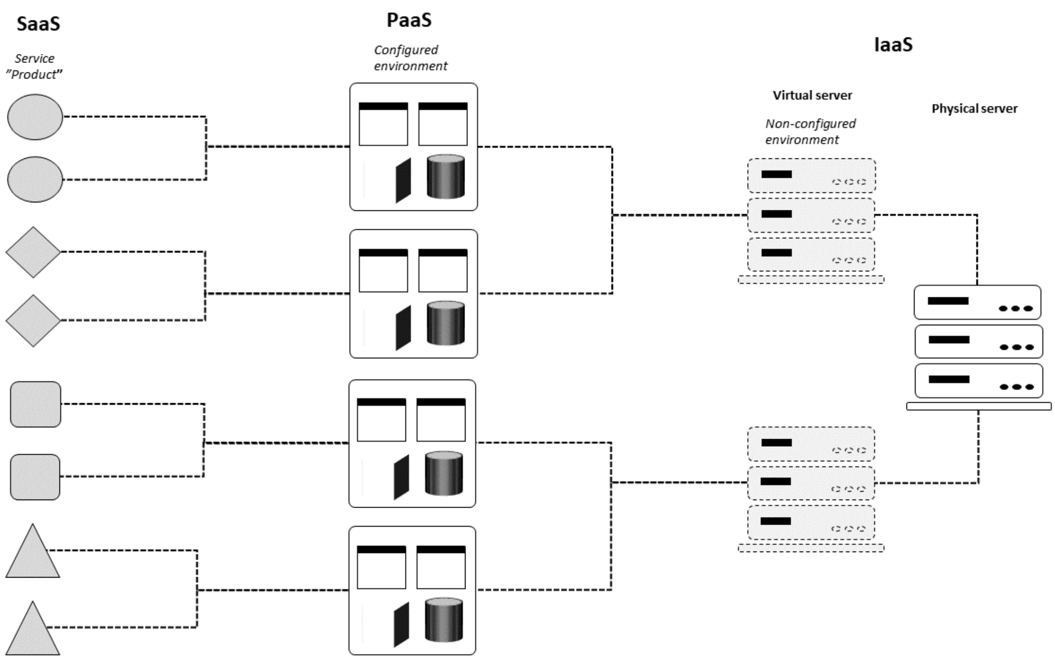


Figure 14. cloud computing structure..

The agreement between the cloud service provider and the cloud service consumer is set in the Service-level agreement (SLA). The SLA defines the behaviours, limitations and quality of service that the cloud provider should be able to meet. For an IaaS it should define the requirements for hardware, such as processing and memory capacity, failover protections, availability percentage and pricing plan. For PaaS, the hardware is undefined, so the SLA should define the Application and database mangement systems (DMBS) requirements. Failover is handled in the IaaS, so the Paas SLA can define the elasticity requirements instead. SLA for SaaS can be held more basic, and often define the response time of the cloud service. The pricing plans also have different focus. Even if the IaaS is unconfigured per say, there are running costs of the hardware, such as management storage and energy, which often transfer to the SLA as running cost over time with an additional cost for data transfer to and from the cloud. Paas has the same characteristics, with a running cost over time within limits or according to levels of activity. SaaS is often usage based and pricing plans are often determine cost by the number of requests to the service. (Erl et al. (2013))

The deliver of the cloud service can also be deployed in various formats such as a public cloud, community cloud and hybrid cloud. Public clouds are as the name suggest open publicly accessible environments owned by various third party cloud vendors. Community clouds are very similar to public clouds but focused to a certain community that define and manage the cloud to its consumers, both within or outside the community. Private clouds are clouds where organizations are both the provider and consumer of the cloud and where the physical servers are located on-premise. It is still defined as a cloud if it is remotely accessible to cloud consumers. There are also hybrid deployment models where the different types are combined, for example to differentiate on security. (Erl et al. (2013))

2.5 Analytic applications

Analytics is applying statistics on a dataset to find and discover useful value adding patters. These statistical methods are applied on a small population extracted from the total population with he intent of making inferences. The datasets are usually restricted by what has been proven to be sufficient for the applied model, but usually range between 20-50, but can also be within 5-200 observations. The small population should mirror the properties of the population from where it was extracted so that the results form the model can be statistically significant. The datasets used in statistics for analysis are usually structured and have been formatted to run with ”out-of-the-box” classical models. traditionally models susch as ANOVA, T-tests, chi-squared tests and F-tests were applied on data based on what result to analyse, while now the data is explored prior to the selection of model. The latter is usually referred to as Exploratory data analysis (EDA). This concept is focused on letting the data guide the model selection by looking at it visually. This is done by plotting raw data, histograms, probability plots. Statistical plots such as mean plots, box plots and standard deviation plots are utilized. Human natural pattern recognition can be used by plotting positioning plots to see if anything is revealed. EDA is also applied on small data, because of the processing requirements, but also since large datasets can attain asymptotic properties as those in infinite datasets.(Ratner, Day, and Davies (2011))

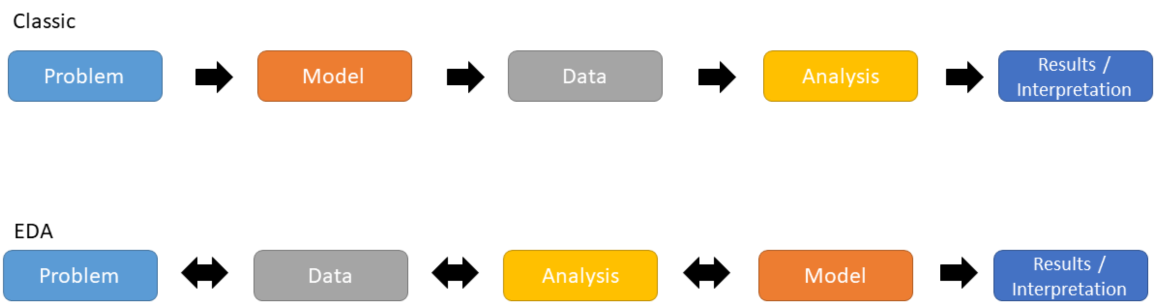


Figure 15. Analytic methods.

EDA has benefits over the classical approach since it is guided from the data and allow the problem, data and analysis to be intertwined. Due to this it allows for more flexibility, practicality and simplicity. The concern with EDA is also that due to the flexibility and universality, it can also be a restricting factor to determine what to use and the validity of the choice. Problems of EDA is often that too large populations are used, which can make it unstructured, unspecific and in the same time ill conditioned with missing values and outliers.(Ratner et al. (2011))

Different functions and method can be applied to data to ”correct” for features such noise, skewness and non-linearity by for example smoothing, log transform and ladder of powers. The data can arise from different concepts and they need to be understood correctly. These depend on the origin of measurement and can be nominal, ordinal, interval and ratio. These have specific attributes that need to be considered. To be significant, the end result has to be both reliable, valid and reliability does not necessarily apply validity, since it is possibility to reliable and wrong. (Ratner et al. (2011))

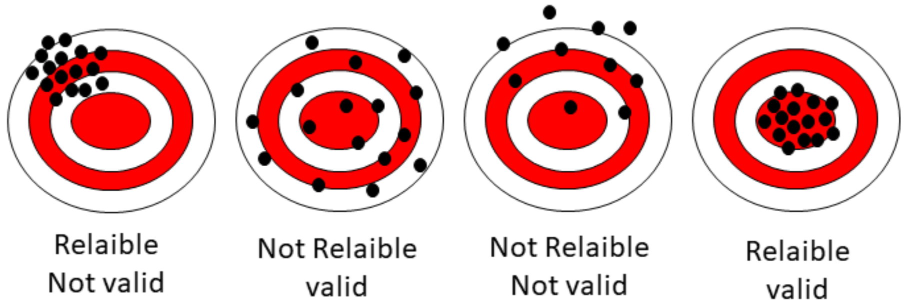


Figure 16. Analytic methods.

2.6 Artificial intelligence applications

Artificial intelligence is an umbrella term for technologies that impose the notion of conceived intelligence. There are different types of disciplines in artificial intelligence based on the input data and objective. The disciplines widely used today is Machine learning, Natural language processing, computer vision and robotics as well two more abstract disciplines, such as knowledge representation and automated reasoning. Traditional statistical analytics try to prove a hypotheses by using a sample population to draw an inferences to the target population. Some relation can be seen in how branches of artificial intelligence is applied to predict a dependent value on the basis of features in independent variables. The main difference is that artificial intelligence allow for unconventional and unstructured data formats to find the most probable dependent variable and does not care how it is found. Rather the objective is that it will have a high accuracy and validity in its predictions and that the model is able to achieve these results in a production environment. This is the most basic practical application of artificial intelligence and can be adopted to most of the disciplines. (Russell and Norvig (2022))

Control theory, what can be known as stochastic optimal control evolved around the notion of cost functions and optimization. Traditional control theory is derived from continuous linear or non-linear dynamic systems with a varying degree of freedom. For good control, it would need to be made linear time invariant, such that impulse responses eventually can be identified, and the system and system characteristics are found and a sufficient control can be created. As the degrees of freedom and non-linearity increase, the systems become difficult to mathematically model and simulate. AI work well in environment were these characteristics exist, if the input and output data is available such that the environment (system) is fully observable. (Russell and Norvig (2022))

The conceptual objective and or motivation for artificial intelligence can partly be derived from the test Alan Turing constructed in 1950 to answer the question, ”Can a machine think?”. Turing himself limited his own proposal to a set of written questions that if answered by a machine, would have intelligence if the answers couldn’t be distinguished from answers given by a person. The total Turing test is a concept extending this thought experiment by including the interactions in the real world, in the form of physical simulation. (Russell and Norvig (2022))

Achieving this involves understanding numerous of different domains and theories, to understand languages for communication, knowledge representation and decisions, reasoning for reaching conclusions, adaptation to changing circumstances through extrapolation, perception of the world through vision and speech, and finally movement through robotics. (Russell and Norvig (2022))

On a general term, AI has reached a high level of maturity in generative applications for text, audio and images to a such high degree that synthetic and non-synthetic outputs are hard to distinguish, industrialization of computer vision in applications such as object detection, natural language processing is reaching greater economical impact as more mature models are adopted into everyday applications from tech giants such as both Google and Microsoft. Machine learning taking large steps in healthcare and biology, providing patter recognition and desiece. (for Human-Centered Artificial Intelligence (n.d.))

Machine learning consist of three distinctive methods, supervised learning, unsupervised learning and reinforcement learning, which can be applied as deep learning methods neural networks. Machine learning is essentially observing a environment through a large set of data in which a hypothesis exist, and using the data of the environment to determine if the hypothesis is true.

Supervised learning is the practice of mapping outputs to inputs to learn the map-function. Outputs are called labels and can be of different variations, but is an representation of the object to be mapped to the input. this is done through utilizing features from the environment space. This practice is often combined with EDA since it can be important to get prior knowledge on the environment. This analysis of the dataset is also important to understand which model could give the best result, given the features, label and their respective characteristics. Supervise learning rely highly on correctly labelled data to achieve good performance. when the model is applied with the dataset, it is split into training-, validation- and test sets. The model is then initialized with some starting parameters and trained using the training dataset, this result is evaluated with the validation dataset, where parameters are tuned. This can be done multiple times or until a wanted error rate. The model can then be evaluated with the test data. if the model is able to perfectly predict all labels in the training data, but poorly in the test data, it has interpolated the data in the training set, and thus is overfitted. The contrary is undefitting, when data is unable to depict the actual model of the environment and is missing crucial information. The loss function computed from the error rates is a representation on how well the model is working. Hyper-parametrization can be utilized as an option to the initial parametrisation of the model during training as a mean to improve the performance, this is a way to use a structured approact to choose the optimal parameters for the model. (Russell and Norvig (2022))

Unsupervised learning is the practice of learning without any labels. This is usually done to reveal patters, anomalies, clusters of data, or to determine important features in the environment. Semi-supervised learning methods are a combination of the above two methods in which both are used, but with more focus on unsupervised, but uses some labels to boost performance. (Russell and Norvig (2022))

Reinforcement learning is a method to an agent in the environment that acts in the environment itself to gain rewards for ”good behaviour”. If the agent is able to freely act in the environment, it will eventually learn how to act as to achieve the highest amount of reward points, which in theory should equal the wanted outcome. (Russell and Norvig (2022))

Deep learning and neural networks are sometimes used interchangeable, but deep learning is techniques where the environment is changed into a complex network of algebraic circuits in where the connections are configurable on strength and amplitude. The name neural network comes from the superficial resemblance of actual activities of neurons in the brain. (Russell and Norvig (2022))

Natural language processing, Robotics and Computer vision are fields in where the ”natural” interactions are converted into a sequence of zeros and ones such that the knowledge in the natural format can be used in machine learning. This involves the conversion of images, speech and text into strings of binary data that can be used in machine learning to extract useful information. This requires an amplitude of factors but two important ones are understanding the domain in which the original data is collected from and secondly, understanding how to choose the right model for the data, and what to extract from the data. (Russell and Norvig (2022))

For text, it is understanding language models, grammar, parsing, semantics and other factors subject to the text, needed to convey the original content. Speech has the same constraint, but from other factors such as mispronunciations, dialects, difference in voices , such as pauses and emphasises. (Russell and Norvig (2022))

Robotics involves quantification of perception and action, where to large problems problem are partial observability and stochasticity in real continuous high-dimensional states.

(Russell and Norvig (2022))

Computer vision is understanding the image formation in terms of camera placement, depth, direction, flow and other characteristics the human eye can perceive from an image. (Russell and Norvig (2022))

If the labels on the data is sufficient with what we want to learn, we need to only find the best model to learn the same. Often the above characteristics are hard to label, and there are fields of practice dedicated to understanding them. However creating a model(Russell and Norvig (2022)) , require the following steps:

1. Data collection - data sources, availability, storage and privacy.
2. Data preparation - analytics, preprocessing and splitting of data.
3. Train a model - Choosing algorithm, Regularize, tune parameters
4. Evaluate - result evaluation, Feature importance and cost
5. Serve mode - set in production
6. Retrain - evaluate integrity of old values.

This creates a counterpart to the white box model, namely a black box model without any transparency in how the output is linked ot the input. This is necessary when the purpose is to create a model that operate in an environment which has more rules that are applicable in normal programming. There is a growing field of Explainable AI (XAI) which is focused on increasing the transparency of black box models. (Russell and Norvig (2022))

2.7 Maturity models

One of the first maturity models designed is the Capability Maturity Model drafted in 1990, This maturity model was designed for those who wants to better understand, identify, transfer or improve practices for software design within a organization Paulk, Curtis, Chrissis, and Weber (1993). A maturity model should relay how well a process is defined, managed, measured, controlled, and its effectiveness. The explicit levels allow for identification and growth, as well as provide a common measurement for comparison Paulk et al. (1993).

The purposes above can be comprised into three distinctive types of maturity models, descriptive, prescriptive and comparative. As the names suggest, descriptive is applied as a diagnostic tool to identify the current capabilities. The prescriptive maturity model should identify the desirable level of maturity with guidelines on how to achieve it. A comparative maturity model is used as a benchmarking tool. Pã (2011)

The purpose of the maturity model is considered during the design phase of the model. Pã (2011) propose a framework for the design of a maturity model called the design principle framework. This framework consist of a *Basic design principles*, *design principles for descriptive purpose of use* and *design principles for prescriptive purpose of use*. They purposefully leave out the comparative model.

A good maturity model should be constructed in such a way that the Application domain, purpose, target group, class of entities are clearly defined, as well as how it differ from other maturity models and its extent of empirical validation. Specific maturity and maturation aspects have to be defined such that levels, maturation paths, dimensions and granularity is understood and the theoretical foundation is robust and underpin evolution and change. This is also in reference to the application domain with target-group documentation.

Descriptive maturity models are more advanced and require a basic model that is empirically evaluated and hold strong intersubjectively verifiability. The prescriptive model is also complex and require that the improvement paths are understood and provide basis for the best practices for each maturity level.

3 Creating conceptual framework for value proposition and maturity model

The conceptual framework is the basic construct of elements used in an AI and analytic suite in the power generation sector. This conceptual framework can be used to create value propositions and maturity models.

3.1 The big picture, leveraging AI

The AI an analytic suite is a versatile platform, highly dependant on the underlying network capabilities, connections as well as the computation capabilities of the hosting environment. Artificial intelligence (AI) itself is dependant on a hypotheses, and the environment in which a hypothesis exist and its observability. By leveraging different types of AI and machine learning practices, a black box model of the environment is created, in which the hypotheses is tested or constructed. This hypothesis might origin from a problem statement.

Here, the environment is the business process of a power plant, or a pool of power plants. It is assumed that there is an information model implemented, used to integrate the environment of the physical power plant to the business operation of a company, enterprise or organization. However, it cannot be assumed that the business operation information model provide a fully observable environment.

There are multiple business models for how a power plant can position itself in the electricity grid, but it is assumed that in the majority of situations the underlying concept is that the original capital cost for the construction of the power plant is repaid through profit from revenue generated in sold electricity.

The power plant is subject to fixed costs, such as administration and general expenses,

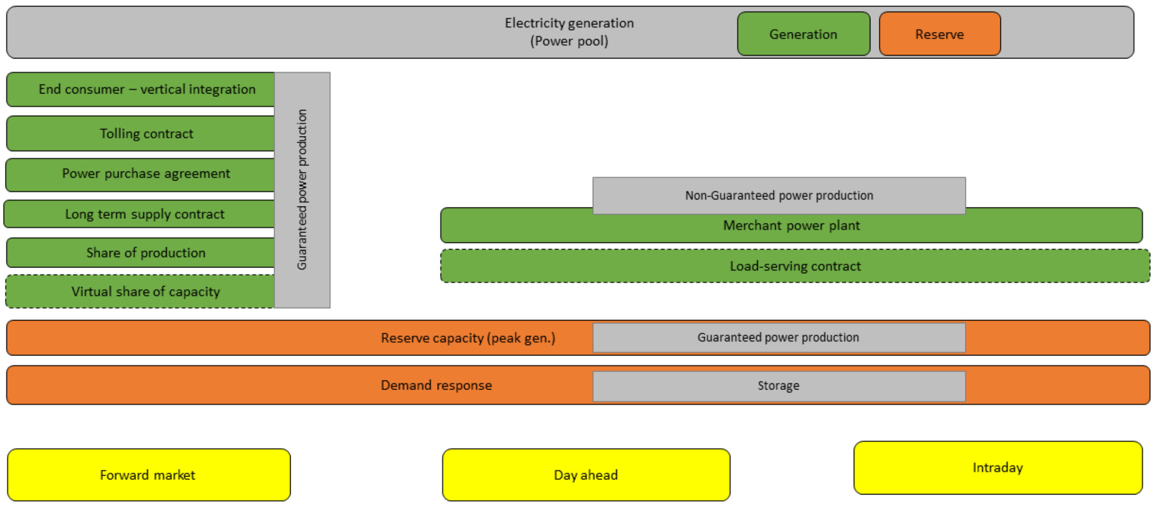


Figure 17. Business models of in power generation.

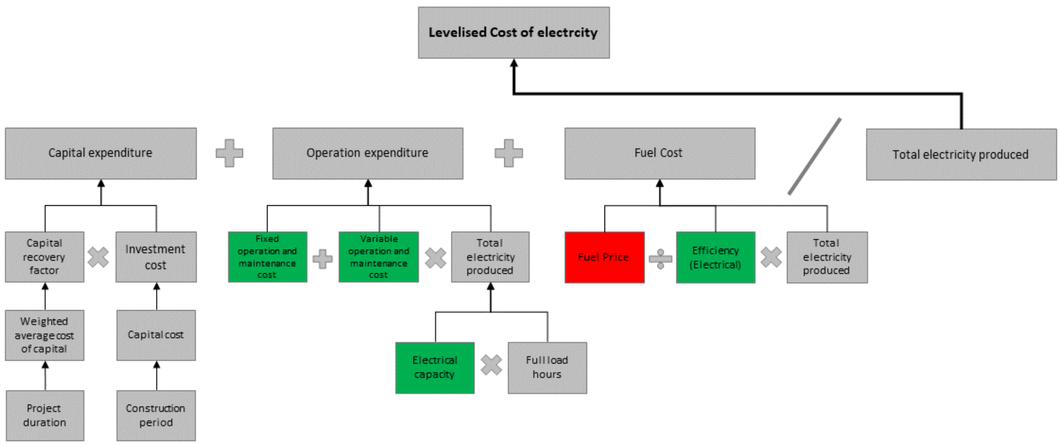


Figure 18. Levelized cost of electricity show a typical cost structure, where fuel cost is separated from operational expenditure. The elements in green can be assumed to be controllable to a certain degree, both in design phase and in the operation phase, the red element is more unpredictable and subject to most external factors..

routine labor, materials and contract services, and perhaps also, to variable costs of fluids, chemicals and other products consumed during operation. The variable costs can include fuel related expenses or be calculated on a non-fuel related basis, where the fuel cost is related to the efficiency of the energy transformation in the plant. The activities and functions that carry out activities related to revenue and cost can be located in the level 3 and 4 domains referenced with the ISA-95 enterprise system hierarchy. The actual factors of influence to cost and revenue are then seen in domain 2 and below. These would be such as the health and availability of assets, as well as the efficiency of processes.

This can be comprised into the following elements for digital representation, which also can be called the value propositions as they are the indicative of the performance of the customer business.

- Capital Cost - *costs to engineer, procure, construct, and commission all or some equipment within the plant facility fence line*
- Variable Cost - Operation and maintenance cost that does vary with generation or consumption
- Fixed Cost - Operation and maintenance cost that does not vary with generation or consumption
- Production or consumption revenue - revenue generated through generation or consumption of electricity.

Applying this format to a level of maturity for the AI and analytic suite could be done by expanding the conceptual framework on a fundamental level to incorporate a wider target state than merely profit tracking. This target state would represent the electricity grid as the hypotheses environment in which the power plant is acting as a rational agent, maximizing its objective through real-time multi-domain decision making according to the real status of both the environment and the power plant capability. Underpinning this concept is that both electricity grid and power plant exist in fully observable environments. This wider target state allow to create the differentiation between levels and functionalities in the AI and analytic suite.

The maturity level required should, at baseline, be a digital representation of the power plant and should span the hierarchy of level 3 and 4, from the perspective of the above assumed elements for digital representation as a monetized measurement in each function or activity (1). Below this level, the representation of the environment is not complete, and a only limited hypotheses exploration can be done using analytics in smaller and fixed populations. A fully observable environment should allow the digital twin representation by integrating a fully observable environment of the energy transformation in hierarchy level 2 and below, where the asset health, availability and process efficiency is actualized (2). The final level based on the wider target state would be the digital twin operating as an agent in the electricity grid and is acting on the controls of various functions in the power plant such that the operation is optimized according to some objective (3).

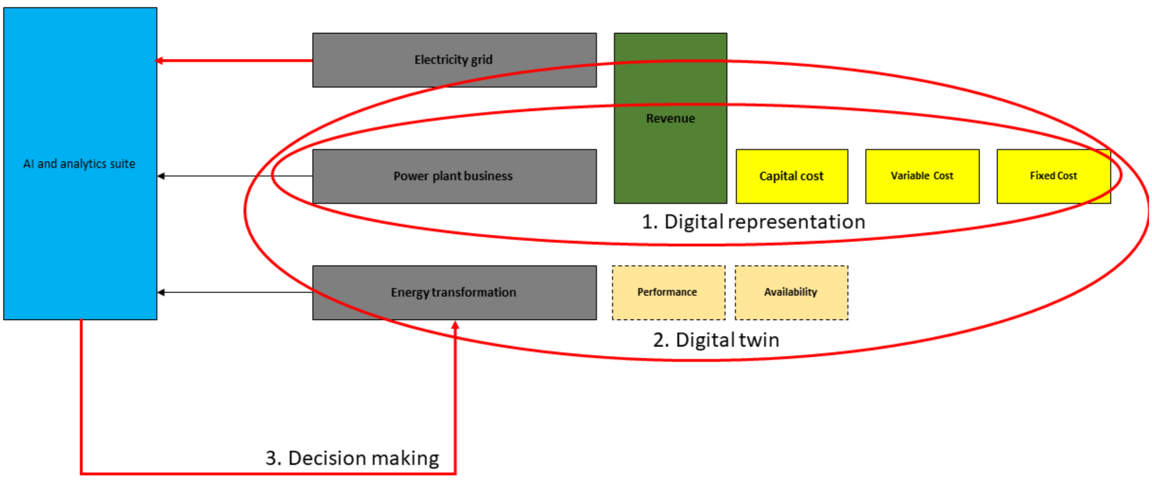


Figure 19. Value proposition elements integrated into the AI and analytic suite with the purpose of optimizing an objective.

This conceptual framework can, based on this define the minimum class of entities that should be included for a digital representation based on the functions and activities in hierarchy level 3 and 4 of ISA-95, which would include:

- Purchasing and inventory
- Collecting and maintaining goods, machinery and equipment
- Requirement management
- Production scheduling
- Capacity scheduling
- Resource management
- Data collection and analysis
- Optimization
- Managing manufacturing, maintenance, quality testing, material storage management

The digital twin should incorporate the energy transformation, which hold the availability, health and process efficiency data (health and efficiency can be referred to as performance). This require addressing the convergence of IT and OT since the sub-level 2 domain comprises of operational technology. Decision making require that the digital twin has an observable environment in which it operates, this require a representation of the electricity grid.

3.1.1 The central domains

The central domains required for the AI and analytic suite in the conceptual framework is therefore:

- Electricity grid *Production or consumption revenue*

- The power plant business *Capital Cost*, *Production or consumption revenue*, *Fixed Cost*, *Variable Cost*
- Networking and hosting hardware. *Convergence of IT and OT*
- The energy transformation *Performance*, *availability*

Electricity grid should have the information necessary to relay the current status of the electricity grid such that the possibility to penetrate and generate revenue, either by production or consumption can be revealed in real time. This information should mainly originate from the TSO, NEMO, BRP and BSP. The information is more likely available if there is a legal obligation to publicly provide or reveal it in a uniform way. A factor that speak against unstructured data in this domain is the fact that the electricity grid has been highly regulated, which implies a more strictly monitored, structured and anticipated environment. This combined with the notion of a de-regulated electricity grid with complex distribution and distributed production in the energy of internet should increase the granularity and resolution of data.

The power plant business is support functions around the procurement, operations and maintenance of the power plant allowing it to function with the highest availability and performance. There are today standards that define how the information should be handled and distributed in organizations and enterprises. The origin of this information is dependant on the maturity of both the implementation and execution of these standards. There can be specific information models for manufacturing or production, or it could be a business process focused information model. However, as the goal of the enterprise in general is to generate an income, it can be assumed that relevant information is located within the IT infrastructure of that enterprise.

The energy transformation, which consist of the process with physical assets, should in theory contain the necessary information to determine both performance and availability, in past, present and future states. The difficulty to attain the true states is more complex toward future states, as the past and present is measured, while the future is a forecast. The ability to measure is indicative on how well it actually represent the actual state over the time-frame. The forecast is instead dependant on a true measurement of both past and present actual states. The ability to measure the true availability is dependant on the understanding the capability and capacity of the process during all states, as well as the wear and condition of assets during all states. Traditionally most of the data from the process is structured, but there is great potential to use unstructured data methods here, for example thermal imaging and sound capturing for measuring condition and wear (Russell and Norvig (2022)). Process efficiency, can be easier to measure and can be improved by optimization, replacement, upgrade or re-design. (Katsaprakakis (2020))

The networking of assets and the process data can be separated from the electricity grid and the power plant business data since these exist in a more mature information technology (IT) environment, where confidentiality, integrity and availability (CIA) is more rule than exception. Accessing this data can however be over various access points, though VPN, or DMZ zone with more limited throughput capability which can limit the use of unstructured data with higher data velocity, volume requirements. Real-time capability is not a necessity in these entities and is often secondary by functionalities that prioritize availability. The network in process and asset entities is often isolated and subject to reliable and deterministic behaviour in close to real-time time frames. The convergence of this OT and IT require more from the OT side, than from the IT side. It would not be practical to introduce the IT data to the OT network, since the OT network is inherently unsecure an designed at a throughput to which it can provide the most deterministic behaviour. This is true for network traffic, protocols and embedded systems in the OT network. The integrity of the operation technology should not be compromised with regard to its deterministic behaviour or by introducing unnecessary lag in the feedback control loop.

The hardware model on which the AI and analytics suite is hosted on is depending on the purpose, but allow for multiple different models according to need. It is quite difficult to estimate the processing needs for training AI models. The different types of AI functions

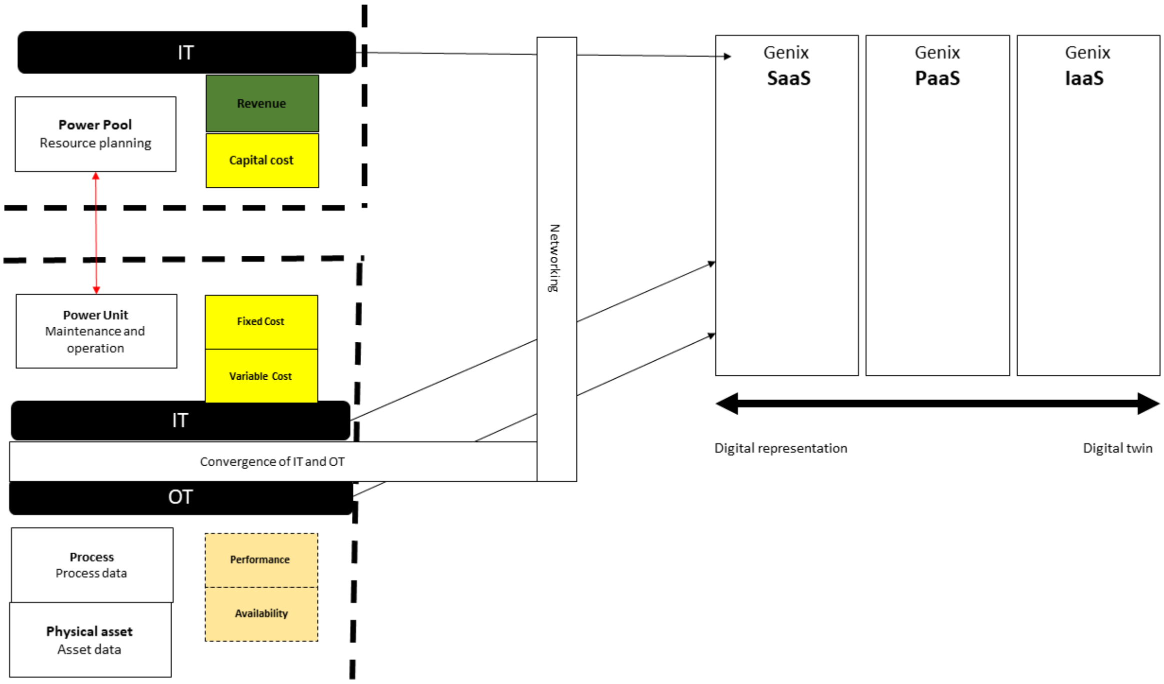


Figure 20. Value proposition elements integrated into the AI and analytic suite with regard to business architecture and networking.

can be described as earlier mentioned as digital representation, Digital twin and Decision making. The deployment format (SaaS, IaaS or Paas) can be done according to the functionality, or available data. The digital representation is mostly descriptive, with assumed low variation in computation requirements. Predictive functionality which rely on forecasting can be achieved using simple autoregressive functions to more advanced deep learning models such as the recurring neural networks such as LSTM. computer vision, Natural language processing can be good digital representations and allow for a higher observable states that a single point measure. This would require the maturity of the IT/OT network to support this.(Russell and Norvig (2022))

3.2 Multi-Domain maturity model to assess Analytics and AI suite in Power generation industry

Given the above class of entities, the main domains to be included can be described as the power pool, representing the data related to the revenue conditions and capital cost.

The second domain would be the power units, where variable cost and fixed costs should be represented. The energy transformation consisting of process and physical assets. should be the last domain. This could be more describing according to the power plant structure rather that business structure. The energy transformation need to consider its maturity in networking as well due to the IT/OT convergence. Big data can be relevant in the power pool and power unit, as an indicative meter on how well unstructured data is adopted. Even if unstructured big data would be applicable in the energy transformation, it is computationally most likely modelled and processed elsewhere.

The maturity model is based on the construct for the conceptual framework, by actualizing the cost and revenue factors into concrete function and activities in the power pool and power units. The maturity model is an extension of the conceptual framework and should reveal the overall data utilization. Each class of entity is therefore described by its current capability for measurement, forecasting and decision making with reference for the AI and analytics suite.

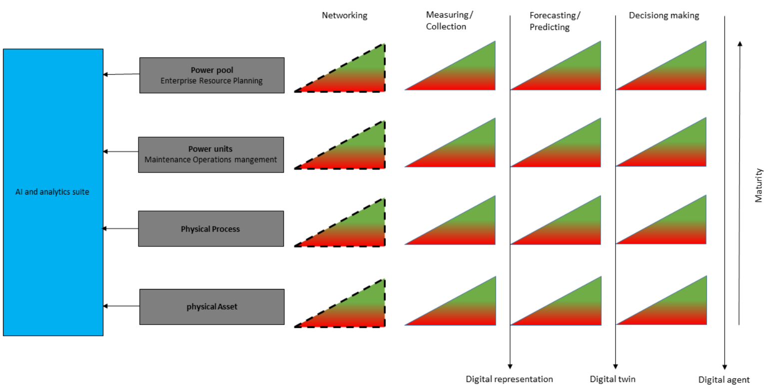


Figure 21. Maturity model over four domains, across AI and analytic suite dimensions..

This can be reorganized into four distinctive domains for more clarity:

- Enterprise Resource Planning (Power pool)
  - AI and analytics - Measuring/Collection.
  - AI and analytics - Forecasting/Predicting.

  - AI and analytics - Decisiong making.
  - Big data
- Maintenance Operations Management (Power unit)
  - AI and analytics - Measuring/Collection.
  - AI and analytics - Forecasting/Predicting.
  - AI and analytics - Decisiong making.
  - Big data
- Physical Process
  - AI and analytics - Measuring/Collection.
  - AI and analytics - Forecasting/Predicting.
  - AI and analytics - Decisiong making.
  - Networking and hosting hardware
- Physical Assets
  - AI and analytics - Measuring/Collection.
  - AI and analytics - Forecasting/Predicting.
  - AI and analytics - Decisiong making.
  - Networking and hosting hardware

Over which also the granularity and quality is questioned, as they can provide some insight into how observable the environment is and if the data is utilized in the right domain.

3.3 Maturity levels, granularity and quality

The class of entities are assessed over different data processes, such as Measuring/Collection, Forecasting/Predicting and decision making. Since these processes are variable in how they are executed, they need to be self-assessed for a level of maturity by the participant. The content of the maturity model is created to reference the conceptual framework with respect to the baseline for a digital representation.

Measuring/Collection - Forecasting/Predicting - Decision making - Networking - The maturity model is derived from the Capability Maturity Model (CMM) as it is a robust model used for processes. It assumes that the risk increase with lower maturity and predictability with a higher maturity. CMM, is a framework, originally intended for software processes, which have inherent generic levels that are good to start from. The final levels that represent the data process best were chosen as:

Ad hoc- reactive- managed- proactive- automated

The rough definition is that ad-hoc level of data processes involve the stochastic measuring, predicting or decision making without any structure, cause or result in a isolated environment. The automated level should represent automatic measuring (digital representation), predicting (digital twin) or decision making (digital agent), according to a clearly defined structure and data sharing. More details can be found in Appendice 5. Networking was produce using same rhetoric applied on network data and network capability. Persse (2001)

Big data Big data maturity using the principles of the CMM, applied using the work done by Halper (n.d.), a maturity model could be created for the use of big data over the levels of:

Basic- Opportunistic- Systematic- Differentiating- Transformational

to understand how the company is leveraging big data which can provide a transformational approach to pre-existing data use where each process is further evaluated for granularity and quality to reveal what extent of items in the activity does the process apply, and how well the result of the process represent reality.Iafrate (2018) This is to better understand how observable the environment is and how well the data represent the right domain. Granularity is evaluated over 4 percentage steps , below 25% , above 25% , above 50% and above 75%, and quality in ”Poor, Fair, Good and Excellent”.

This should reveal the data process utilization and its maturity in the different domain of the conceptual framework. The whole maturity survey on data utilization can be found in the appendix of this thesis.

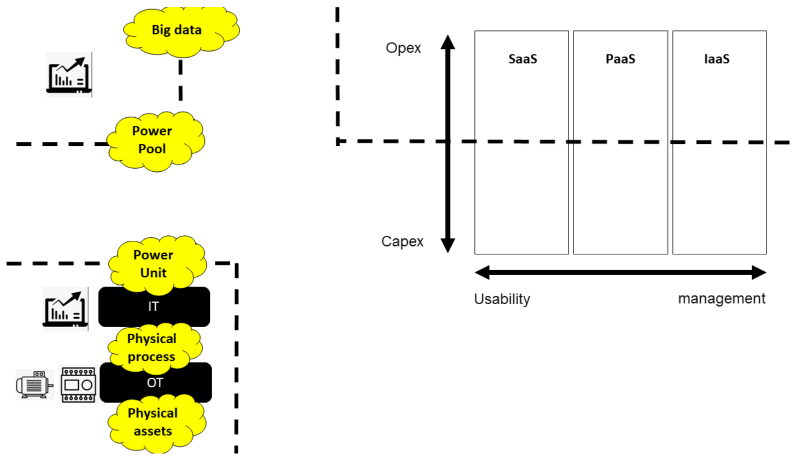


Figure 22. The domains in which the data utilization is assessed.

The survey design is comprise of the class entities and constructs applied over all domains:

Power Pool (Enterprise resource planning) • Purchasing and inventory, *Access and availability status of resources and materials related to the power pool operation and maintenance or other activities:* • Collecting and maintaining goods, machinery and equipment, *Performance of the power pool as a combination of its assets.* • Production scheduling, *Scheduled and measured electricity generation for the power pool to produce what over defined states.* • Requirement management, *Requirements related to operating the power pool.* • Production profit, *Profit from energy generation over defined production schedules.*

Power Unit (Maintenance and Operations Management) • Production schedule, *Sched-*

uled and measured electricity generation for the power unit over defined states. • Resource management, *Availability of resources to handle operation and maintenance, both planned and unplanned.* • Collecting and maintaining goods, machinery and equipment, *Performance of the power unit as a combination of its assets with related machinery and equipment.* • Managing maintenance and storage (Unit), *Scheduled and measured maintenance activities for the power unit in defined states and or reasons* • Overall equipment Effectivenes (Unit), *The overall combined effectiveness of all assets coupled to the power unit at fixed or different levels of capacity* • Operational cost, *Cost for power unit over and in defined states.*

Physical Processes - DCS, ICS, SCADA • Process Values, • Process Commands, • Process specification, *The process is operating according to design specifications* • Process performance, *The process operates in its defined state of production* • Process quality, *Deviations from the requested operational conditions.*

Physical Assets • Design specification, *The asset is operating according to design specifications* • Design requirements, *Requirements related to operating a power generation facility- for example:Efficiency, Safety, Environment* • Lifecycle, *Current and non-current lifecycle states of the asset, and its status relative to the lifecycle of the plant.* • Performance, *Example: Asset running while not needed or not providing value* • Energy Efficiency, *The energy efficiency of the assets in its different states* • Condition, *The condition of the assets.*

4 ABB Case study

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

This page has been intentionally left blank due to confidentiality

5 Result

The result of this thesis is a conceptual framework which identify the value propositions based o the cost and revenue structure of power plants on which the basis for integration into an AI and analytic suite can be made. Physical process and physical assets were identified as domains in which PAEN have domain knowledge, and novel value proposition channels were developed for new and existing projects to answer research question 1.

This would allow PAEN to continue operate mostly in their current knowledge domain, as the ABB Genix AI and analytic suite inherently is focused on offering end customer, asset performance management, which is outside the scope of PAEN. The survey will be followed up with a written assessment of the maturity according to this framework, where the feedback of the customer will be noted and thus also provide validation for research question 2 on level of maturity.

5.1 Conceptual framework

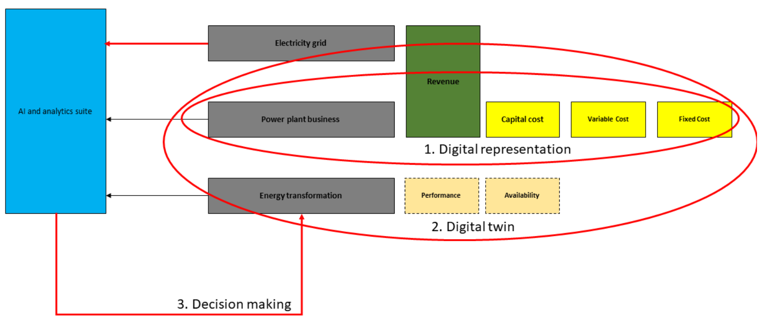


Figure 23. A summary of the conceptual framework in which the data-domains and their content is visible. The circled domains represent the maturity levels according to the AI and analytic suite.

5.1.1 New projects

It was suggested that new projects should incorporate the performance and availability measurements or calculations already in the design phase. This would allow the customer to have access to this information and potentially extend it for further benefit. The Genix-ready design is a suitable option, and is based on the conceptual framework. It should provide all the factors to realize the value proposition to value delivery.

5.1.2 Existing projects

Existing projects were surveyed and at the time of writing, two responses have been returned. the mean value for each domain is shown, all three data processes. From the data it can be seen that the there is a generally lower maturity toward decision making compared to measuring, which is in line with the fact that you cannot make decision without data (1 = ad hoc, 5 = automated). There is a overall poor granularity indicating low effort to measure, but what is measure is considered with good quality (1 = Poor, 4 = Excellent). There is a clear indication that both granularity and the data utilization can be improved. A similar measurement device as in new projects should be provided to existing projects to measure performance and availability. The survey will be followed up with the participants later in December of 2021, where details can be given to strengthen or weaken the validity of the survey. Some feedback has arrived which pointed to the size, being overwhelming. The full surveys can be found in the appendices.

		Measured/ Collected			Forecasted/ Predicted			Decision making/ Optimization		
		Process	Granularity	Quality	Process	Granularity	Quality	Process	Granularity	Quality
Case 1	Power Pool (Business Management)	3,60	75 %	3,20	3,40	75 %	3,20	3,20	75 %	3,00
	Power Unit (Maintenance and Operations Management)	3,17	58 %	2,50	3,00	58 %	2,50	1,83	46 %	2,33
	Processes - DCS, ICS, SCADA	5,00	63 %	3,00	4,75	69 %	3,00	4,00	69 %	3,00
	Physical Assets (prime mover, generator, pumps, motors, fans)	3,33	63 %	3,00	3,50	67 %	3,00	3,50	58 %	3,00
Case2	Power Pool (Business Management)	3,20	60 %	3,00	3,40	60 %	2,80	3,40	60 %	2,80
	Power Unit (Maintenance and Operations Management)	3,17	54 %	3,00	3,00	54 %	2,67	2,83	46 %	2,83
	Processes - DCS, ICS, SCADA	3,50	44 %	2,50	2,50	38 %	2,50	3,25	50 %	2,50
	Physical Assets (prime mover, generator, pumps, motors, fans)	2,33	38 %	2,67	2,33	38 %	2,17	3,00	54 %	2,67

Figure 24. Survey result as a mean value over each domain..

6 Discussion

The Freedom in the domains and the flexibility in the AI and analytics suite was difficult to maintain, since as pointed out before, the value proposition need to derived from a problem statement. The electricity market is also bound to high complexity and a market specific target was not chosen. It is however shown through the literature, that the cost structure and revenue structure is decoupled from business plan, but can have high variety in term of renewable and non-renewable generation units. The ISA95 is a straightforward information model, that create a clear foundation for the function and activities used, even if it originally is directed toward manufacturing, rather than electricity generation. The industry 4.0 concepts are difficult to combine in the power plant environment, since the technology there is more conventional, and the leap between the two create a problematic area to discuss were technologies don't overlap. It can also be argued that it can be difficult to discuss technologies such as AI and machine learning in the conventional space, as the space is underpinned with strict requirements, specifications and regulations. This can easily create the same gap, where the performance of the model need to be shown, before the data is chosen, which is the inverse approach to machine learning. It can of course be understood that investments need to have basis in what can be expected and what should be delivered. The thesis helped me understand the multi-domain environment, while also observing that the difficulty is getting a commitment, as the data for value creation, can involve multiple domain, and it is not necessarily attain-able through the same channels. The effort to build the case to even test a hypothesis is extensive and require an initial investment. The conceptual framework, combined with the conceptual design for the Genix-ready modular design allow for focused investment thought these concept to build a proof of concept and attain the necessary business partners, as well as necessary in-house knowledge. Further work need to be done in creating in-house knowledge in white-box and black-box modelling. A market study into cost and revenue structures in the modern grid. The creation of a modular asset pool that can be used to create specific Genix model. This work should help create a more narrow approach to the market while having a proof of concept.

Bibliography

ABB Ability™ Genix Industrial Analytics and AI Suite. (n.d.). https://new.abb.com/cpm/industrial-software-solutions/genix.

Abolhassan, F. (Ed.). (2017). *The Drivers of Digital Transformation*. Cham: Springer International Publishing.

Alberti, A. M., Santos, M. A. S., Souza, R., Da Silva, H. D. L., Carneiro, J. R., Figueiredo, V. A. C., & Rodrigues, J. J. P. C. (2019). Platforms for Smart Environments and Future Internet Design: A Survey. *IEEE Access*, *7*, 165748–165778.

ANSI/ISA-95.00.01-2010 (IEC 62264-1 Mod) Enterprise-Control System Integration - Part 1: Models and Terminology. (n.d.). https://www.isa.org/products/ansi-isa-95-00-01-2010-iec-62264-1-mod-enterprise.

ANSI/ISA-95.00.02-2018, Enterprise-Control System Integration - Part 2: Objects and Attributes for Enterprise-Control System Integration. (n.d.). https://www.isa.org/products/ansi-isa-95-00-02-2018-enterprise-control-system-i.

ANSI/ISA-95.00.03-2013 Enterprise-Control System Integration - Part 3: Activity Models of Manufacturing Operations Management. (n.d.). https://www.isa.org/products/ansi-isa-95-00-03-2013-enterprise-control-system-i.

Armendia, M., Ghassempouri, M., Ozturk, E., & Peysson, F. (Eds.). (2019). *Twin-Control: A Digital Twin Approach to Improve Machine Tools Lifecycle*. Cham: Springer International Publishing.

Balusamy, B., R, N. A., & Gandomi, A. H. (2021). *Big data: Concepts, technology and architecture* (First edition ed.). Hoboken, NJ: John Wiley and Sons, Inc.

Barricelli, B. R., Casiraghi, E., & Fogli, D. (2019). A Survey on Digital Twin: Definitions, Characteristics, Applications, and Design Implications. *IEEE Access*, *7*, 167653–167671.

Branicky, M. S. (2005). Introduction to Hybrid Systems. In D. Hristu-Varsakelis & W. S. Levine (Eds.), *Handbook of Networked and Embedded Control Systems* (pp. 91–116). Boston, MA: Birkhäuser Boston.

CEER. (2021). *Status review of renewable support schemes in europe for 2018 and 2019 ceer report renewables work stream of electricity working group* [Report].

Colbert, E. J. M., & Kott, A. (Eds.). (2016). *Cyber-security of SCADA and Other Industrial Control Systems* (Vol. 66). Cham: Springer International Publishing.

CWNA Certified Wireless Network Administrator Study Guide. (n.d.). , 1091.

Dorf, R. C., & B., Robert H. (2022). *Modern control systems*.

Dzung, D., Naedele, M., Von Hoff, T., & Crevatin, M. (2005, June). Security for Industrial Communication Systems. *Proceedings of the IEEE*, *93*(6), 1152–1177.

EC. (2020). *Communication from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions powering a climate-neutral economy: An eu strategy for energy system integration* [Report].

EC. (2021). *Annex to the report from the commission to the european parliament, the council, the european economic and social committee and the committee of the regions - 2021 report on the state of the energy union - contribution to the european green deal and the union's recovery* [ANNEX].

EcoStruxure: IoT – Internet of Things. (n.d.). https://www.se.com/ww/en/work/campaign/innovation/overview.jsp.

EEA. (2014). *Why did greenhouse gas emissions decrease in the eu between 1990 and 2012?* [Report].

Erl, T., Puttini, R., & Mahmood, Z. (2013). *Cloud computing: Concepts, technology, & architecture*. Upper Saddle River, NJ: Prentice Hall.

European University Institute. (2017). *Design the electricity market(s) of the future*. LU: Publications Office.

for Human-Centered Artificial Intelligence, A. I. I. S. I. (n.d.). *Ground the conversation about ai in data*. Retrieved from https://aiindex.stanford.edu/

Gartner. (n.d.). *Industrial IoT Platforms Reviews 2021 — Gartner Peer Insights*. https://www.gartner.com/reviews/market/industrial-iot-platforms.

Group, I. . E. W. (n.d.). *Ieee 802.3 industry connections bandwidth assessment part ii* [report].

Halper, F. (n.d.). TDWI Big Data Maturity Model Guide: Interpreting Your Assessment

Score. , 20.

Hussain, H. M., Narayanan, A., Nardelli, P. H. J., & Yang, Y. (2020). What is Energy Internet? concepts, Technologies, and Future Directions. *IEEE Access*, *8*, 183127–183145.

Iafrate, F. (2018). *Artificial Intelligence and Big Data: The Birth of a New Intelligence*. Hoboken, NJ, USA: John Wiley & Sons, Inc.

ISO55000. (n.d.). *Iso 55000:2014 asset management — overview, principles and terminology.*

Kalabin, S. (2018). *Machine learning solutions for maintenance of power plants* (Master's thesis, Aalto University. School of Electrical Engineering). Retrieved from http://urn.fi/URN:NBN:fi:aalto-201809034766

Kare, L. (2019). *Benefits of digitalized asset management for steam turbines* (Master's thesis, Aalto University. School of Engineering). Retrieved from http://urn.fi/URN:NBN:fi:aalto-201908254971

Katsaprakakis, D. A. (2020). *Power plant synthesis* (First ed.). Boca Raton: Taylor and Francis.

Kirschen, D., & Strbac, G. (2004). *Fundamentals of Power System Economics: Kirschen/Power System Economics*. Chichester, UK: John Wiley & Sons, Ltd.

Landeta Echeberria, A. (2020). *A Digital Framework for Industry 4.0: Managing Strategy*. Cham: Springer International Publishing.

Lappi, J. (2019). *Asset Performance Management application for power system condition monitoring in an Internet of Things platform* (Master's thesis, Aalto University. School of Electrical Engineering). Retrieved from http://urn.fi/URN:NBN:fi:aalto-201905122987

Last, M., Kandel, A., & Bunke, H. (Eds.). (2004). *Data mining in time series databases* (No. v.57). New Jersey ; London: World Scientific.

Lei, W., Soong, A. C., Jianghua, L., Yong, W., Classon, B., Xiao, W., ... Saboorian, T. (2021). *5G System Design: An End to End Perspective*. Cham: Springer International Publishing.

Mandler, B., et al. (Eds.). (2016). *Internet of Things. IoT Infrastructures: Second International Summit, IoT 360° 2015, Rome, Italy, October 27-29, 2015, Revised Selected*

Papers, Part II (Vol. 170). Cham: Springer International Publishing.

Mäntysaari, P. (2015). *EU Electricity Trade Law*. Cham: Springer International Publishing.

McCabe, J. D., & McCabe, J. D. (2003). *Network analysis, architecture, and design* (2nd ed ed.). San Francisco, CA: MK/Morgan Kaufmann Publishers.

Murphy, D. J., & Hall, C. A. S. (2011, February). Energy return on investment, peak oil, and the end of economic growth: EROI, peak oil, and the end of economic growth. *Annals of the New York Academy of Sciences*, *1219*(1), 52–72.

Pã, J. (2011). WHAT MAKES A USEFUL MATURITY MODEL? a FRAMEWORK OF GENERAL DESIGN PRINCIPLES FOR MATURITY MODELS AND ITS DEMONSTRATION IN BUSINESS PROCESS MANAGEMENT. , 13.

Paulk, M., Curtis, B., Chrissis, M., & Weber, C. (1993, July). Capability maturity model, version 1.1. *IEEE Software*, *10*(4), 18–27.

Persse, J. R. (2001). *Implementing the capability maturity model*. New York: John Wiley & Sons.

Predix Platform — Industrial IoT Application Platform — GE Digital. (n.d.). https://www.ge.com/digital/iiot-platform.

Ratner, B., Day, S., & Davies, C. (2011). *Statistical and Machine-Learning Data Mining* (Zeroth ed.). CRC Press.

Russell, S. J., & Norvig, P. (2022). *Artificial intelligence: A modern approach* (Fourth edition, global edition ed.). Harlow: Pearson.

Saldanha, T. (2019). *Why digital transformations fail: The surprising disciplines of how to take off and stay ahead* (First edition ed.). Oakland, CA: Berrett-Koehler Publishers, a BK Business Book.

Serpanos, D. N., & Wolf, T. (2011). *Architecture of network systems*. Burlington, MA: Morgan Kaufmann.

Siebel, T. M. (2019). *Digital transformation: Survive and thrive in an era of mass extinction*. New York: RosettaBooks.

Siemens — MindSphere. (n.d.). https://siemens.mindsphere.io/en.

Singh, M., Fuenmayor, E., Hinchy, E. P., Qiao, Y., Murray, N., & Devine, D. (2021, May). Digital Twin: Origin to Future. *Applied System Innovation*, *4*(2), 36.

Smil, V. (n.d.). Energy and Civilization: A History. , 564.

Smil, V. (2010). *Energy transitions: History, requirements, prospects*. Santa Barbara, Calif: Praeger.

Tan, L., Hou, H., & Zhang, Q. (2016). An extensible software platform for cloud-based decision support and automation in precision agriculture. In *2016 ieee 17th international conference on information reuse and integration (iri)* (p. 218-225).

Tian, S., & Hu, Y. (2019). The role of opc ua tsn in it and ot convergence. In *2019 chinese automation congress (cac)* (p. 2272-2276).

Tinga, T. (2013). *Principles of Loads and Failure Mechanisms*. London: Springer London.

Tsoukalas, L. H., & Gao, R. (2008, April). From smart grids to an energy internet: Assumptions, architectures and requirements. In *2008 Third International Conference on Electric Utility Deregulation and Restructuring and Power Technologies* (pp. 94–98). Nanjing, China: IEEE.

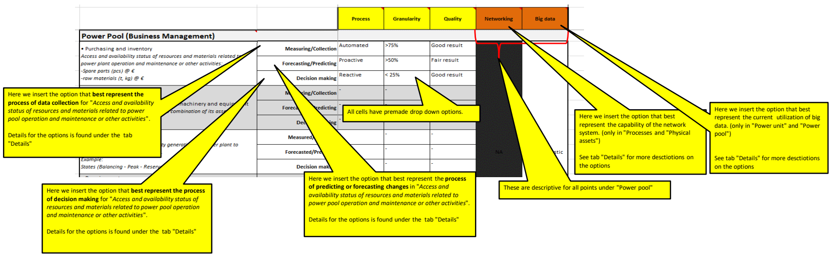
Xu, L. D. (2015). *Enterprise integration and information architecture: A systems perspective on industrial information integration* (No. 2). Boca Raton: Auerbach Publications.

Appendices

Appendix 1. Data survey

		Process	Granularity	Quality	Networking	Big data
Power Pool (Business Management)						
• Purchasing and inventory *Access and availability status of resources and materials related to the power pool operation and maintenance or other activities:* *-Spare parts (pcs) @ €* *-raw materials (t, kg) @ €*	**Measuring/Collection**	-	-	-	NA	-
	Forecasting/Predicting	-	-	-		
	Decision making	-	-	-		
• Collecting and maintaining goods, machinery and equipment *Performance of the power pool as a combination of its assets.*	**Measuring/Collection**	-	-	-		
	Forecasting/Predicting	-	-	-		
	Decision making	-	-	-		
• Production scheduling *Scheduled and measured electricity generation for the power pool to produce what over defined states.* *Example:* *States (Balancing - Peak - Reserve) @ kW or %*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Requirement management *Requirements related to operating the power pool.* *Example:* *-Electricity market requirements* *-Emission requiremetns* *-Support scheme requiremetns*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Production profit *Profit from energy generation over defined production schedules* *Example:* *Estimated return in € from stages in production schedule*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
Power Unit (Maintenance and Operations Management)						
• Production schedule *Scheduled and measured electricity generation for the power unit over defined states.* *Example:* *states: (Grid-Island-backup) @ kW or %*	**Measured/Collected**	-	-	-	NA	-
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Resource management *Availability of resources to handle operation and maintenance, both planned and unplanned.* *Example: Workforce-, Materials-, Tools- availabilityand utilization*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Collecting and maintaining goods, machinery and equipment *Performance of the power unit as a combination of its assets with related machinery and equipment.*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Managing maintenance and storage (Unit) *Scheduled and measured maintenance activities for the power unit in defined states and or reasons* *Example:* *states: Spare_part(x) of Asset(x) @ (planned, unplanned) for €*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Overall equipment Effectivenes (Unit) *The overall combined effectiveness of all assets coupled to the power unit at fixed or different levels of capacity*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Operational cost *Cost for power unit over and in defined states* *Example:* *(Variable-load, start-up, NO-load, planned-unplanned maintenance)*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		

Processes - DCS, ICS, SCADA						
• Process Values • Process Commands	**Measured/Collected**	-	-	-	-	NA
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Process specification *The process is operating according to design specifications* *Example - Rated or Nominal values vs real values*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Process performance *The process operates in its defined state of production* *Example: No-load operations is an example of non-performing state*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Process quality *Deviations from the requested operational conditions* *Example: derating is an example of a low quality state*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
Physical Assets (prime mover, generator, pumps, motors, fans)						
• Design specification *The asset is operating according to design specifications* *Example - Rated or Nominal values vs real values*	**Measured/Collected**	-	-	-	-	NA
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Design requirements *Requirements related to operating a power generation facility- for example:Efficiency, Safety, Environment*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Lifecycle *Current and non-current lifecycle states of the asset, and its status relative to the lifecycle of the plant.*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Performance *Asset contribution* *Example: Asset running while not needed or not providing value*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Energy Efficiency *The energy efficiency of the assets in its different states*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		
• Condition *The condition of the assets*	**Measured/Collected**	-	-	-		
	Forecasted/Predicted	-	-	-		
	Decision making	-	-	-		



Process
Granularity
Quality
Networking
Big data
Power Pool (Business Management)
• Purchasing and inventory
Access and availability status of resources and materials related to power plant operation and maintenance or other activities:
-Spare parts (pcs) @ €
-raw materials (t, kg) @ €
Measuring/Collection
Forecasting/Predicting
Decision making
Automated
Proactive
Reactive
>75%
>50%
< 25%
Good result
Fair result
Good result
Here we insert the option that best represent the process of data collection for "Access and availability status of resources and materials related to power pool operation and maintenance or other activities". Details for the options is found under the tab "Details"
Here we insert the option that best represent the process of decision making for "Access and availability status of resources and materials related to power pool operation and maintenance or other activities". Details for the options is found under the tab "Details"
All cells have premade drop down options.
Here we insert the option that best represent the process of predicting or forecasting changes in "Access and availability status of resources and materials related to power pool operation and maintenance or other activities". Details for the options is found under the tab "Details"
These are descriptive for all points under "Power pool"
Here we insert the option that best represent the capability of the network system. (only in "Processes and "Physical assets") See tab "Details" for more desctiotions on the options
Here we insert the option that best represent the current utilization of big data. (only in "Power unit" and "Power pool") See tab "Details" for more desctiotions on the options

Measuring/Collection

ad hoc	Data is measured or collected manually on ad hoc basis, without any structure or definition and is not shared. Data storage, logging, resolution or other factors are not considered. Each measurement is individual and hard to compare against other measurements since the method can be undefined. *Example: manual observation is done and noted at a local computer station or on paper without further action.*
reactive	Data is measured or collected manually as a result of other events, with some structure or definition to certain groups or organizations. Data storage, loggin, resolution or other factors are not considered. Each measurement is individual and hard to compare against other measurements since the method can be undefined. *Example: a manual observation is done following an event and noted at a local computer station or through a paper process for further action, such as noting number of spare parts after breakdown and sharing the information to necessary groups and organizations .*
managed	Data is measured or collected manually according to a defined structure and definition to necessary groups or organizations. Data storage, logging, data resolution are considered when handling data. Measurements are done according to known methods and observations can be compared against historical observations. *Example: There is a structured and defined way of measuring, such as scheduled walks where spare part inventory is noted or measured and shared to necessary groups and organizations.*
proactive	Data is measured or collected continuously but has to be processed manually for necessary groups or organizations. Data storage, logging, data resolution are considered important. Measurements are done according to known methods and observations can be compared against historical observations. *Example: The measuring of spare parts inventory is automated and collected to a local station from where it manually can be shared further to necessary groups and organizations.*
automated	Data is measured or collected continuously and processed automatically to desired form for necessary groups or organizations. There are efforts to have real-time data measurements and the measurements can be used by all groups and organizations simultaneously. Measurements are done according to known methods and observations can be compared against historical observations. *Example: The measuring of spare parts inventory is automatically collected, processed and shared further to necessary groups and organizations.*

Forecasting/Predicting

ad hoc	Data is forecasted or predicted manually on ad hoc basis, without any structure or definition and is not shared. There is a lack of datasets and machine learning skills. Some estimates and predictions are done using conventional analytics, and if machine learning is used, there is poor understanding of the different types of machine learning methods or how to handel large datasets. Each outcome is individual and not comparable to historical observations. *Example: An individual or group can estimate the average costs of shutdown based on historical data, but with no connection to the next shutdown.*
reactive	Data is forecasted or predicted manually as a result of other events, without structure or definition to certain groups or organizations. Analytics or machine learning is used after events have occurred to estimate or forecast based on historical data. The data or Machine learning method is not considered, but applied as is and each outcome is individual and not comparable to historical outcomes. *Example: The cost of a shutdowns are estimated based on historical data as the shutdown has happened.*
managed	Data is forecasted or predicted manually according to a defined structure and definition to necessary groups or organizations. Datasets structure and characteristics is understood and machine learning efforts are carried out as a part of business activities. There is an understanding of the relation between different methods and types of machine learning. Models are tuned and datasets are evaluated on a regular basis and outcomes are comparable to historical outcomes. *Example: A group or individual has a role to regularly estimate when spare parts should be ordered and is continuously tuning the model as new data is available.*
proactive	Data is forecasted or predicted continuously but has to be processed manually for necessary groups or organizations. Artificial intelligence is applied to structured or unstructured data, such as video, image, speech on a regular basis and the results are valid and accurate. Predictions and estimates are locally integrated in business and other processes of groups and organizations. Outcomes are comparable to historical outcomes and deviations are addressed and analyzed. *Example: An operator is given a probability percentage of an incoming shutdown event at a local HMI screen and can act on this information.*
automated	Data is forecasted or predicted continuously and processed automatically to desired form for necessary groups or organizations. There is an understanding of the environment and events and changes in operation can be anticipated from most activities. Outcomes are comparable to historical outcomes and deviations are addressed and analyzed. *Example: Inventory manager can observe running assets and is given estimates on changes in, asset condition, current stock, logistics, maintenance schedule and the actual production schedule.*

Decisiong making

ad hoc	Decision making is done on ad hoc basis, using minimal data, without any structure, definition or objective and can have unpredictable outcomes. Each decision is individually made and the outcomes and decisions criteria's cannot be compared. *Example: A small dataset or data points is used to trigger an ad-hoc decision based largely on own interpretation*
reactive	Decision making is done as a result of other events, without any structure, definition or objective and can have unpredictable outcomes. Each decision is individually made and the outcomes and decisions criteria's cannot be compared. *Example: A small dataset or data points is used to trigger an reactive event, such as a decision to replace an asset as the data shows a fault. inventory or logistics is not considered.*
managed	Decision making is done according to a defined structure and with more clear objectives. Data collection and predictions are used to drive decisions with higher probability. Decisions criteria's are known and the outcomes can be compared to historical outcomes. *Example: A maintenance manager is collecting daily data on assets and given a decision for conditional maintenance at a estimated threshold.*
proactive	Decision making is automated, optimized and provide support for manual high probability outcome decisions according to goals and objectives. Decisions criterias are well known and the outcomes can be compared to historical outcomes. Deviation in outcomes are addressed and analyzed. *Example: A maintenance manager can observe the real time data and condition of the asset and is given options for maintenance scheduling that are optimized and maximized for the goals and objectives.*
automated	Decision making is automated and optimized according to goals and objectives. Decisions criteria's are well known and the outcomes can be comapred to historical outcomes. Deviation in outcomes are addressed and analyzed. *Example: A maintenance manager is given real-time scheduling of maintenance and operations that are optmized for the business goals and objectives. OR a process is optimized for efficiency or cost reduction and is capable of adjusting according to changes in the environment.*

Networking

Ad hoc	Issues and needs are handled on a ad-hoc basis, often by non-specialists. This involve Information security, computing and networking tasks. These tasks are carried out and followed up by paper-based processes. There is a lack of IT-capability, practices and services. There is no, or poor external communication possibilities
Reactive	Basic IT processes are defined and followed, such that minor needs and issues can be quickly resolved. Occasional backup and monitoring is done on IT and OT assets. Reactive patching and updates are done on OT and IT assets as vulnerabilities are found or informed. There are some basic intrusion prevention and anti-virus measures installed. There are some basic access possibilities that can be activated on demand, such as VPN, intended only for specific activities.
Managed	There are asset management processes for IT and OT assets, and site personnel is aware of needs and operational requirements as well as antivirus measures .Different networks are segregated to separate between different networks (productions / Office). and IT and OT systems can be patched within a defined number of hours (less than 48) if necessary. Workstations and Industrial control systems are applied to comprehensive security programs (e.g. NIST). Access points are managed to handle data transfer and remote access over internet. Both VPN and DMZ is established and configured for more extensive use.
Proactive	IT and OT are more seamlessly integrated and allow for more internal and external data sharing between networks (internal and external) over a secured infrastructure. There are more advanced access points in the network system to allow for smart devices, both wired and wireless. Automatic validation allow patching to be done quicker as vulnerabilities are detected in the IT or OT assets. Both IT and OT systems are able to recover in the event of company wide security incidents though implemented recovery procedures.
Automated	Mature IT network systems keep the infrastructure and services adaptive, automated and secure. Seamless integration of IOT plug-and-play modules can be integrated and the network system scales automatically. The network system allow for secure real-time communication enabling remote control of processes. Security is based on continuous intrusion and threat monitoring , and the response is automatic. Patching does not require any downtime.

Big data

Basic	There is no clear management of data, and data in various formats are stored in data warehouses and minimally defined formats and structures. The use for the data is not defined, but collected on a what-if and ad-hoc basis. The data is often used as-is without preprocessing or analysis of its content or without any overall plan or strategy. The infrastructure is not structured to handle data volume, user and data security, access to information, processing requirements, and information storage. Analytics utilization from big data is specific to certain groups and organization and not shared. There is a difference in structured and unstructured data and mostly structured data is utilized.
Opportunistic	Experimental usage of big data infrastructure technologies such as Hadoop is explored to provide proof of concept to further integrations. There is an understanding of the specifications, requirements as well as the operation, maintenance and configuration aspects, but only proof of concept is carried out. There is a wider understanding on data usage, but data is collected without auditability. Some more advanced analytics is carried out on the collected data and more unstructured data is involved in analytics.
Systematic	Data has more mature and defined metadata structure and more descriptive and predictive applications is utilized in specific business activities were there is clear problems and solutions. There is still no clear data management strategy but there is an understanding on how new, internal or external, data streams and data sources can be integrated into analytic platforms. Specific security, condition monitoring, backup and recovery is now more implemented, as well as performance monitoring, often completely migrated to a cloud environment as the data is more available and utilized throughout the whole organization. There is a concept between existing assets and new data.
Differentiating	The big data infrastructure is in production and multiple groups and organization can simultaneously utilize and leverage data and analytics. The infrastructure is complient with enterprise standards and have mature procedures for security, condition monitoring, backup and recovery is now more implemented, as well as performance monitoring. There is a defined lifecycle management and auditability of data realized in enterprise NoSQL databases, Hadoop and data warehouses.
Transformational	The analytics and predictive applications is in production and fully operational in most organization, even in mission-critical aspects of business such as real-time decision making. These infrastructure is also treated as mission critical and has extensive staff to support to maintain its operability. There is a continuous process to improve on analytics and most data sources are complex data types can be understood and utilized as well as their connections to assets in the business.

Appendix 2. Data survey result, Case 1

Case 1

		Process	Granularity	Quality	Networking	Big data
Power Pool (Business Management)						
• Purchasing and inventory *Access and availability status of resources and materials related to the power pool operation and maintenance or other activities:* *-Spare parts (pcs) @ €* *-raw materials (t, kg) @ €*	**Measuring/Collection**	Managed	>75%	Good result	Managed	Opportunistic
	Forecasting/Predicting	Managed	>75%	Good result		
	Decision making/Optimization	Managed	>75%	Good result		
• Collecting and maintaining goods, machinery and equipment *Performance of the power pool as a combination of its assets.*	**Measuring/Collection**	Managed	>75%	Good result		
	Forecasting/Predicting	Managed	>75%	Good result		
	Decision making/Optimization	Managed	>75%	Good result		
• Production scheduling *Scheduled and measured electricity generation for the power pool to produce what over defined states.* *Example:* *States (Balancing - Peak - Reserve) @ kW or %*	**Measured/Collected**	Proactive	>75%	Good result		
	Forecasted/Predicted	Proactive	>75%	Good result		
	Decision making/Optimization	Proactive	>75%	Good result		
• Requirement management *Requirements related to operating the power pool.* *Example:* *-Electricity market requirements* *-Emission requiremetns* *-Support scheme requiremetns*	**Measured/Collected**	Proactive	>75%	Excellent		
	Forecasted/Predicted	Proactive	>75%	Excellent		
	Decision making/Optimization	Managed	>75%	Good result		
• Production profit *Profit from energy generation over defined production schedules* *Example:* *Estimated return in € from stages in production schedule*	**Measured/Collected**	Proactive	>75%	Good result		
	Forecasted/Predicted	Managed	>75%	Good result		
	Decision making/Optimization	Managed	>75%	Good result		
Power Unit (Maintenance and Operations Management)						
• Production schedule *Scheduled and measured electricity generation for the power unit over defined states.* *Example:* *states: (Grid-Island-backup) @ kW or %*	**Measured/Collected**	Proactive	>75%	Good result	NA	Basic
	Forecasted/Predicted	Proactive	>50%	Good result		
	Decision making/Optimization	Reactive	> 25%	Good result		
• Resource management *Availability of resources to handle operation and maintenance, both planned and unplanned.* *Example: Workforce-, Materials-, Tools- availabilityand utilization*	**Measured/Collected**	Managed	>50%	Fair result		
	Forecasted/Predicted	Managed	>50%	Fair result		
	Decision making/Optimization	Reactive	>50%	Fair result		
• Collecting and maintaining goods, machinery and equipment *Performance of the power unit as a combination of its assets with related machinery and equipment.*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Managed	>50%	Good result		
	Decision making/Optimization	Reactive	>50%	Good result		
• Managing maintenance and storage (Unit) *Scheduled and measured maintenance activities for the power unit in defined states and or reasons* *Example:* *states: Spare_part(x) of Asset(x) @ (planned, unplanned) for €*	**Measured/Collected**	Managed	>75%	Fair result		
	Forecasted/Predicted	Reactive	>75%	Fair result		
	Decision making/Optimization	Reactive	>50%	Fair result		
• Overall equipment Effectivenes (Unit) *The overall combined effectiveness of all assets coupled to the power unit at fixed or different levels of capacity*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Managed	>75%	Good result		
	Decision making/Optimization	Reactive	>50%	Fair result		
• Operational cost *Cost for power unit over and in defined states* *Example:* *(Variable-load, start-up, NO-load, planned-unplanned maintenance)*	**Measured/Collected**	Managed	>50%	Fair result		
	Forecasted/Predicted	Managed	>50%	Fair result		
	Decision making/Optimization	Ad hoc	>50%	Fair result		

Case 1

		Process	Granularity	Quality	Networking	Big data
Processes - DCS, ICS, SCADA						
• Process Values • Process Commands	**Measured/Collected**	Automated	>75%	Good result	Proactive	NA
	Forecasted/Predicted	Automated	>75%	Good result		
	Decision making/Optimization	Proactive	>75%	Good result		
• Process specification *The process is operating according to design specifications* *Example - Rated or Nominal values vs real values*	**Measured/Collected**	Automated	>50%	Good result		
	Forecasted/Predicted	Automated	>75%	Good result		
	Decision making/Optimization	Proactive	>50%	Good result		
• Process performance *The process operates in its defined state of production* *Example: No-load operations is an example of non-performing state*	**Measured/Collected**	Automated	>50%	Good result		
	Forecasted/Predicted	Automated	>50%	Good result		
	Decision making/Optimization	Proactive	>75%	Good result		
• Process quality *Deviations from the requested operational conditions* *Example: derating is an example of a low quality state*	**Measured/Collected**	Automated	>75%	Good result		
	Forecasted/Predicted	Proactive	>75%	Good result		
	Decision making/Optimization	Proactive	>75%	Good result		
Physical Assets (prime mover, generator, pumps, motors, fans)						
• Design specification *The asset is operating according to design specifications* *Example - Rated or Nominal values vs real values*	**Measured/Collected**	Managed	>75%	Good result	Managed	NA
	Forecasted/Predicted	Managed	>75%	Good result		
	Decision making/Optimization	Managed	>50%	Good result		
• Design requirements *Requirements related to operating a power generation facility- for example:Efficiency, Safety, Environment*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Managed	>75%	Good result		
	Decision making/Optimization	Managed	>50%	Good result		
• Lifecycle *Current and non-current lifecycle states of the asset, and its status relative to the lifecycle of the plant.*	**Measured/Collected**	Managed	>75%	Good result		
	Forecasted/Predicted	Managed	>75%	Good result		
	Decision making/Optimization	Managed	>75%	Good result		
• Performance *Asset contribution* *Example: Asset running while not needed or not providing value*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Proactive	>50%	Good result		
	Decision making/Optimization	Proactive	>50%	Good result		
• Energy Efficiency *The energy efficiency of the assets in its different states*	**Measured/Collected**	Proactive	>75%	Good result		
	Forecasted/Predicted	Proactive	>75%	Good result		
	Decision making/Optimization	Proactive	>50%	Good result		
• Condition *The condition of the assets*	**Measured/Collected**	Proactive	>50%	Good result		
	Forecasted/Predicted	Proactive	>50%	Good result		
	Decision making/Optimization	Proactive	>75%	Good result		

Appendix 3. Data survey result, Case 2

Case 2

		Process	Granularity	Quality	Networking	Big data
Power Pool (Business Management)						
• Purchasing and inventory *Access and availability status of resources and materials related to the power pool operation and maintenance or other activities:* *-Spare parts (pcs) @ €* *-raw materials (t, kg) @ €*	**Measuring/Collection**	Reactive	>50%	Fair result		
	Forecasting/Predicting	Managed	>50%	Fair result		
	Decision making/Optimization	Managed	>75%	Good result		
• Collecting and maintaining goods, machinery and equipment *Performance of the power pool as a combination of its assets.*	**Measuring/Collection**	Managed	>50%	Good result		
	Forecasting/Predicting	Managed	>75%	Good result		
	Decision making/Optimization	Managed	>75%	Good result		
• Production scheduling *Scheduled and measured electricity generation for the power pool to produce what over defined states.* *Example:* *States (Balancing - Peak - Reserve) @ kW or %*	**Measured/Collected**	Proactive	>75%	Excellent		
	Forecasted/Predicted	Proactive	>75%	Good result	NA	Opportunistic
	Decision making/Optimization	Proactive	>50%	Good result		
• Requirement management *Requirements related to operating the power pool.* *Example:* *-Electricity market requirements* *-Emission requiremetns* *-Support scheme requiremetns*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Managed	>50%	Good result		
	Decision making/Optimization	Managed	>50%	Good result		
• Production profit *Profit from energy generation over defined production schedules* *Example:* *Estimated return in € from stages in production schedule*	**Measured/Collected**	Proactive	>75%	Good result		
	Forecasted/Predicted	Managed	>50%	Good result		
	Decision making/Optimization	Managed	>50%	Fair result		
Power Unit (Maintenance and Operations Management)						
• Production schedule *Scheduled and measured electricity generation for the power unit over defined states.* *Example:* *states: (Grid-Island-backup) @ kW or %*	**Measured/Collected**	Proactive	>75%	Good result		
	Forecasted/Predicted	Proactive	>75%	Good result		
	Decision making/Optimization	Managed	>50%	Good result		
• Resource management *Availability of resources to handle operation and maintenance, both planned and unplanned.* *Example: Workforce-, Materials-, Tools- availabilityand utilization*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Managed	>50%	Fair result		
	Decision making/Optimization	Reactive	< 25%	Fair result		
• Collecting and maintaining goods, machinery and equipment *Performance of the power unit as a combination of its assets with related machinery and equipment.*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Managed	>50%	Good result		
	Decision making/Optimization	Managed	>50%	Good result		
• Managing maintenance and storage (Unit) *Scheduled and measured maintenance activities for the power unit in defined states and or reasons* *Example:* *states: Spare_part(x) of Asset(x) @ (planned, unplanned) for €*	**Measured/Collected**	Reactive	>50%	Good result	NA	Opportunistic
	Forecasted/Predicted	Reactive	>50%	Good result		
	Decision making/Optimization	Managed	>50%	Good result		
• Overall equipment Effectivenes (Unit) *The overall combined effectiveness of all assets coupled to the power unit at fixed or different levels of capacity*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Managed	>50%	Fair result		
	Decision making/Optimization	Managed	>50%	Good result		
• Operational cost *Cost for power unit over and in defined states* *Example:* *(Variable-load, start-up, NO-load, planned-unplanned maintenance)*	**Measured/Collected**	Proactive	>50%	Good result		
	Forecasted/Predicted	Managed	>50%	Good result		
	Decision making/Optimization	Managed	>50%	Good result		

Case 2

		Process	Granularity	Quality	Networking	Big data
Processes - DCS, ICS, SCADA						
• Process Values • Process Commands	**Measured/Collected**	Proactive	>75%	Good result		
	Forecasted/Predicted	Managed	>50%	Good result		
	Decision making/Optimization	Managed	>50%	Good result		
• Process specification *The process is operating according to design specifications* *Example - Rated or Nominal values vs real values*	**Measured/Collected**	Proactive	> 25%	Good result		
	Forecasted/Predicted	Managed	>50%	Good result		
	Decision making/Optimization	Proactive	>50%	Good result		
• Process performance *The process operates in its defined state of production* *Example: No-load operations is an example of non-performing state*	**Measured/Collected**	Proactive	>50%	Fair result	Managed	NA
	Forecasted/Predicted	Reactive	< 25%	Fair result		
	Decision making/Optimization	Managed	>50%	Fair result		
• Process quality *Deviations from the requested operational conditions* *Example: derating is an example of a low quality state*	**Measured/Collected**	Reactive	> 25%	Fair result		
	Forecasted/Predicted	Reactive	> 25%	Fair result		
	Decision making/Optimization	Managed	>50%	Fair result		
Physical Assets (prime mover, generator, pumps, motors, fans)						
• Design specification *The asset is operating according to design specifications* *Example - Rated or Nominal values vs real values*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Managed	>50%	Good result		
	Decision making/Optimization	Managed	>50%	Good result		
• Design requirements *Requirements related to operating a power generation facility- for example:Efficiency, Safety, Environment*	**Measured/Collected**	Reactive	> 25%	Fair result		
	Forecasted/Predicted	Reactive	> 25%	Fair result		
	Decision making/Optimization	Managed	>50%	Fair result		
• Lifecycle *Current and non-current lifecycle states of the asset, and its status relative to the lifecycle of the plant.*	**Measured/Collected**	Reactive	>50%	Good result		
	Forecasted/Predicted	Reactive	>50%	Fair result		
	Decision making/Optimization	Managed	>75%	Good result	Managed	NA
• Performance *Asset contribution* *Example: Asset running while not needed or not providing value*	**Measured/Collected**	Reactive	> 25%	Fair result		
	Forecasted/Predicted	Reactive	> 25%	Fair result		
	Decision making/Optimization	Managed	>75%	Fair result		
• Energy Efficiency *The energy efficiency of the assets in its different states*	**Measured/Collected**	Reactive	> 25%	Good result		
	Forecasted/Predicted	Reactive	< 25%	Fair result		
	Decision making/Optimization	Managed	> 25%	Good result		
• Condition *The condition of the assets*	**Measured/Collected**	Managed	>50%	Good result		
	Forecasted/Predicted	Managed	>50%	Fair result		
	Decision making/Optimization	Managed	>50%	Good result		